Welcoming Address

Welcome to Salt Lake City! It is a great pleasure to have you attend the first edition of The International Conference on Image Formation in X-ray Computed Tomography, also nicknamed as The CT meeting.

The idea for this meeting came two years ago, from a wish I formulated to have a venue where CT scientists could meet together to discuss in depth all aspects that impact the image formation process in CT. These aspects include dose evaluation and dose reduction strategies, non-linearity effects and compensation schemes for these effects, image reconstruction algorithms, spectral decomposition, dynamic effects, geometrical calibration, phase-contrast physics, and image quality assessment. Both medical imaging and non-destructive testing applications are of interest, along with emerging breast cone-beam tomography techniques, and the attractive micro- and nano-CT technology. To allow for in-depth discussion between the attendees, oral presentations have been allocated a significant time. Also, comfortable poster sessions have been planned, and shared meals have been organized in a quiet environment.

Such a meeting could not be a success without scientific submissions. I am grateful to all attendees who submitted their work for presentation at this conference. The 4-page format for abstract submission is demanding, but also a great way to guarantee high quality. Thank you for complying with this format. I was delighted by the quality and merit of the submissions.

The meeting could not be a success either without the support of a great scientific committee. I am very thankful to the following scientists who worked hard to help building the scientific program, and whose reputation definitely helped set the conference on firm grounds:

Samit Basu	GE Homeland Protection
John Boone	UC Davis, CA, USA
Carl R Crawford	Csuptwo, LLC, WI, USA
Michel Defrise	Vrije Universiteit of Brussels, Belgium
Bruno DeMan	GE Global Research Center, NY, USA
Frank Dennerlein	Siemens Healthcare Sector, Germany
Rebecca Fahrig	Stanford University, CA, USA
Thomas Flohr	Siemens Healthcare Sector, Germany
Stephen Glick	UMass Medical School, MA, USA
Michael Grass	Philips Research Laboratories, Germany
Joachim Hornegger	University of Erlangen, Germany
Jiang Hsieh	GE Healthcare, WI, USA
Marc Kachelrieß	University of Erlangen, Germany
Andrew Karellas	UMass Medical School, MA, USA
Alexander Katsevich	University of Central Florida, FL, USA
Thomas Köhler	Philips Research Laboratories, Germany
Hioryuki Kudo	University of Tsukuba, Japan
Patrick La Rivière	University of Chicago, IL, USA
Günter Lauritsch	Siemens Healthcare Sector, Germany
Cynthia McCollough	Mayo Clinic, MN, USA
Michael McNitt-Gray	David Geffen School of Medicine of UCLA, CA, USA

Kyle Myers	Office of Science and Engineering Labs, CDRH, FDA, MD, USA
Johan Nuyts	Katholieke Universiteit, Belgium
Xiaochuan Pan	University of Chicago, IL, USA
Roland Proksa	Philips Research Laboratories, Germany
Jeff Siewerdsen	Johns Hopkins University, MD, USA
Mike Silver	Toshiba Medical Research Institute, IL, USA
Karl Stierstorfer	Siemens Healthcare Sector, Germany
Katsuyuki Taguchi	Johns Hopkins University, MD, USA
Ge Wang	Virginia Tech University, VA, USA
Alex Zamyatin	Toshiba Medical Research Institute, IL, USA

We were fortunate to receive generous support from

- Siemens Medical Solutions
- GE Healthcare
- Toshiba Medical Systems Corporation

This support significantly helped in creating a nice environment for presentations and discussions, and also in reducing the registration cost to a bare minimum for students.

Last, I am grateful to Professor Larry Zeng for his help regarding all local arrangements, to Adam Wunderlich for his scientific assistance, and to Dominic Heuscher and Zhicong Yu for providing a strong hand for the numerous tasks that awaited us in the last days before the conference started.

Frédéric Noo, Ph.D.
General Chair
Utah Center for Advanced Imaging Research (UCAIR)
Department of Radiology
University of Utah
E-mail: noo@ucair.med.utah.edu

TABLE OF CONTENTS

Session M.1: Spectral Imaging
Moderated by G. H. Chen and K. Stierstorfer

Session M.2: Patient specific dose evaluation
Moderated by A. Karellas and M. Kachelrieß

Session M.3: Breast imaging
Moderated by A. Alessio

Plenary Talk
Moderated by F. Noo

Session T.1: Few view tomography
Moderated by H. Kudo and S. Leng

Session T.2: Dose, noise and image quality
Moderated by M. McNitt-Gray and S. Glick

Plenary Talk
Moderated by F. Noo

Session T.3: Classical posters
Moderated by A. Wunderlich

Session W.4: Dynamic imaging
Moderated by K. Taguchi and H. Yu

Modeling the performance of a photon counting x-ray detector: Energy response and pulse pileup effects

Katsuyuki Taguchi[1], Mengxi Zhang[1], Eric C. Frey[1], Xiaolan Wang[1], Einar Nygard[2], Jan S. Iwanczyk[2], Neal E. Hartsough[2], Benjamin M. W. Tsui[1], and William C. Barber[2]

***Abstract*—Recently, novel cadmium telluride (CdTe) photon counting x-ray detectors (PCXDs) with energy discrimination capabilities have been developed. The aim of this study was to evaluate the agreement between a PCXD and analytical models regarding the energy response, the deadtime losses, and the distorted, recorded spectrum due to pulse pileup effects. It was found that the performances characteristics measured by the PCXD agree exceptionally well with those predicted by the analytical models.**

I. Introduction

Currentl x-ray computed tomography (CT) images provide accurate anatomical information. Two major problems are that the images are not tissue-type specific and the contrast-to-noise ratio of images is not sufficient for some applications. These problems are due to the use of energy integrating detectors (EIDs). The EIDs reduce tissue contrast by applying larger weights to x-ray photons with higher energies, where the tissue contrast is less than for lower energies. Moreover, electric and Swank noise are added to data.

Photon counting x-ray detectors (PCXDs) with energy discrimination capabilities could eliminate the above discussed inherent problems of EIDs—PCXDs count photons with no additional weight and noise. Multiple energy windows of PCXD would allow for tissue-specific images through accurate material decomposition and improve contrast-to-noise ratio of images.

Major problems with PCXDs are a non-linear energy response and a limited detector speed (i.e., small operational count rates). If a "slow" PCXD is used under very intense x-rays, quasi-coincident photons are recorded as one event with a higher energy (Fig. 1). This phenomenon is called pulse pileup, which results both in a loss of counts, which is called "deadtime losses," and in the distortion of the recorded spectrum. The guideline was to operate a PCXD with a less than 5% of count rate loss in order to maintain the quality of measurements (Fig. 1), which limit the operational count rate at less than 10% of the maximum output count rate of PCXDs.

Thus, while the detector speed and the linearity of the energy response may need to be improved, it is also desirable to develop algorithmic methods to compensate for the energy response, the count rate loss, and the pulse pileup effects, in order to take full advantage of the energy discrimination capabilities of PCXDs. The key to successful compensation methods is to have accurate models of the corresponding properties of PCXDs.

These models have been investigated elsewhere. The energy response of counting detectors has been outlined [1]. Thus the aim of this study was to evaluate the agreement between the data measured by a PCXD [2, 3] and those estimated by analytical models in terms of the energy response, the deadtime losses, and the distorted, recorded spectrum due to the pulse pileup effects.

II. PCXD and Analytical Models

In the following, we outline some of the feature of a PCXD, two simplified detection models (*non-paralyzable* and *paralyzable* detection models), and analytical models of the energy response, deadtime losses, and the spectral distortion of the recorded spectrum with pulse pileup effects.

A. DXMCT-1 PCXD

The PCXD we used in this study is model DXMCT-1 (DxRay, Inc.; Northridge, CA) [2, 3]. One cadmium telluride (CdTe) PCXD block consisted of 16 × 16 pixels with a pitch of ~1 mm and a thickness of ~3 mm. Two such blocks were assembled and connected to four application specific integrated circuits (ASICs). Each ASIC contained 128 parallel channels, each consisting of an amplifier, a pulse shaper, two digital-to-analog converters (DAC), two comparators (energy thresholds), and two counters per detector pixel (see Fig. 1, top). Thus there were 1024 thresholds in total in the DXMCT-1.

The energy deposited by each x-ray photon creates a pulse. The height of the pulse is compared with a given energy threshold. A count is registered in the counter associated with the given threshold if the pulse height exceeds the threshold. Subtracting counts in counters from adjacent energy thresholds yields the counts in the energy window defined by the two thresholds. The DAC allowed for fine tuning of the thresholds per channel in order to compensate for channel-to-channel variations. The details of the DXMCT-1 architecture have been described previously [2, 3].

B. Non-paralyzable and paralyzable detection model

The detection mechanism for most of PCXDs can be well modeled by either the *non-paralyzable* (NP) or *paralyzable* (P) detector model [1], although most detectors behave somewhere

Manuscript received February 4, 2010. This work was supported in part by National Institutes of Health under grants R43 EB008612 and R44 EB008612.

K.T., E.C.F., X.W., and B.M.W.T. are with The Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins University School of Medicine, Baltimore MD21287 USA (phone: 443-287-2425; fax: 410-610-1060; e-mail: ktaguchi@ jhmi.edu).

M.Z. was with Johns Hopkins University when the work was performed.

E.N., J.S.I., N.E.H., and W.C.B. are with DxRay, Inc., Northridge, CA, USA.

in between. When a detector is in the active state, the first photon incident on the detector will put the detector into the detection (inactive) state for a finite period of time called the deadtime (or resolving time), τ. All photons incident on the detector during the deadtime (potentially) contribute to the pulse shape of the recorded count. For non-paralyzable detectors, such photons will *not* reset the time clock for the deadtime and will return to the active state after the deadtime τ. By contrast, in paralyzable detectors a detected photon resets the time clock; and the detector returns to the active state at a time τ after an event if and only if there is no additional incident photon during the deadtime interval. For low true count rates ($a\tau << 1$, where a is the true count rate), the two methods provide identical results [1].

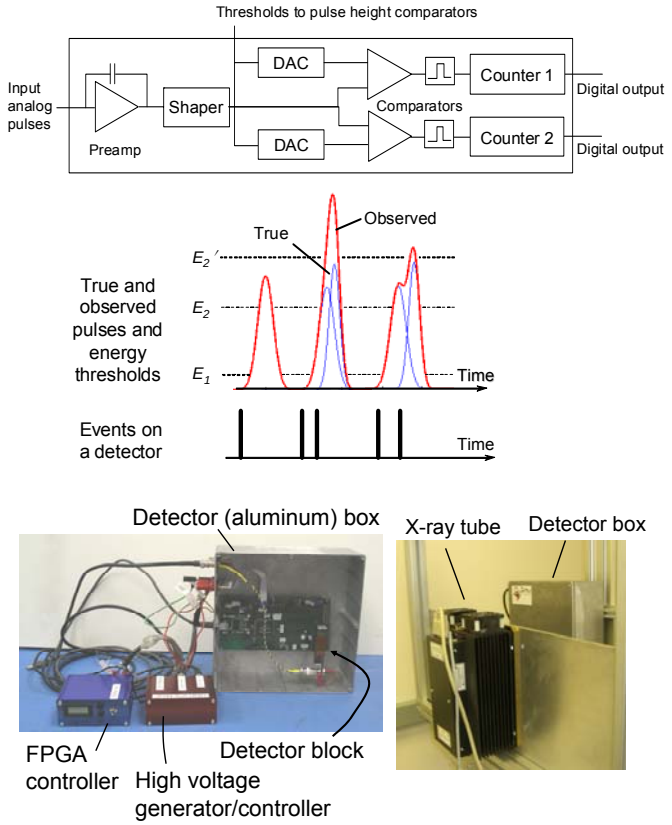


Fig. 1 (Top) The basic architecture of an individual channel in the ASIC. (Middle) When the pulse height exceeds a given energy threshold, a count will be added to an associated counter. Coincident photons will be recorded as one event with a higher energy than the original energies. (Bottom left) The DxRay's PCXD. (Bottom right) The experimental setting.

C. Energy response

The energy response of PCXDs involves the following two factors: a non-linear photon energy-pulse height response curve and a shift-variant finite energy resolution. The energy of the photon is measured by the height of the pulse (in millivolts, *mV*) generated by the photon. But their relationship may not be linear. Thus we model the non-linear relationship by

$$mV(E) = c_1 - c_2 \exp(-c_3 E), \quad (1)$$

where E is the photon energy, mV is the measured pulse height, and c_1, c_2, c_3 are three parameters we will obtain for each threshold/comparator through calibrations.

A method to model the shift-variant energy resolution was successfully developed but was not used in this study, as preliminary results showed the effect was not significant under the conditions used this study.

D. Deadtime losses

The recorded count rate a_R with a given true count rate a and deadtime τ can be expressed as

$$a_R = a \times \Pr(rec \mid a\tau), \quad (2)$$

where $\Pr(rec \mid a\tau)$ is the probability of events being recorded, which has been discussed in Ref.[1] and is:

$$\Pr(rec \mid a\tau) = \begin{cases} 1/(1+a\tau) & \text{NP detector} \\ \exp(-a\tau) & \text{P detector} \end{cases}. \quad (3)$$

E. Distortion of the recorded spectrum

The pulse pileup effects on the recorded spectrum are the most complex phenomenon of PCXD, for which we have successfully developed analytical models in Ref. [4]. Briefly, the expected counts recorded at energy E , $N_r(E)$, can be calculated by a product of probabilities [4, 5]:

$$\begin{aligned} N_r(E) &= (\text{True count rate}) \times \begin{pmatrix} \text{Probability of} \\ \text{events being recorded} \end{pmatrix} \\ &\times \sum_m \left[\begin{pmatrix} \text{Probability of} \\ \text{the } m\text{th order pulse pileup} \\ \text{provided the events are recorded} \end{pmatrix} \right. \\ &\left. \times \begin{pmatrix} \text{Probability of} \\ \text{counts recorded at an energy } E \\ \text{with } m\text{th order pulse pileup} \end{pmatrix} \right] \quad (4) \\ &= a \times \Pr(rec \mid a\tau) \times \sum_{m=0}^{\infty} \Pr(m \mid rec) \Pr(E \mid m) \end{aligned}$$

where mth order pileup means m+1 photons are recorded as one count. The first probability, $\Pr(rec \mid a\tau)$, is the same as Eq. (3), and the second probability, $\Pr(m \mid rec)$, has been discussed in Ref. [1] and is:

$$\Pr(m \mid rec) = \begin{cases} (a\tau)^m \exp(-a\tau)/m! & \text{NP detector} \\ \left[1-\exp(-a\tau)\right]^m \exp(-a\tau) & \text{P detector} \end{cases}, \quad (5)$$

The third probability of Eq. (4), $\Pr(E \mid m)$, will be discussed in detail in Ref. [4], which takes the following factors into account: the shape of pulses, the probability distribution of time intervals between photons, and the probability density function of the original photon energies.

III. Evaluation Methods

In this section we outline the experimental settings to acquire data and data analysis methods to assess the performance characteristics of the DXMCT-1in terms of the energy response, the deadtime losses, and the distortion of the recorded spectrum due to pulse pileup effects. Assuming the true count rate a is linearly related to the tube current I as $a = k \times I$, we estimated five parameters (c_1, c_2, and c_3 in Eq. (1), and τ and k) and one data (the incident x-ray spectrum, $S_0(E)$) for each threshold; we then evaluated the agreement of the analytical models with the detector outputs. We used neither empirical parameters nor fudge factors in the models.

A. *Energy response*

We estimated the three parameters of the photon energy-pulse height (mV) curve for each energy threshold/comparator, c_1, c_2, and c_3 of Eq. (1), using ^{99m}Tc (140 keV), ^{57}Co (122 keV), and four x-ray tube voltages (35, 50, 65, and 80 keV). The DXMCT-1 was placed 150 mm from an x-ray focus (see Fig. 2, right). There was no attenuator except for the detector box (an aluminum plate with a thickness of 1.2 mm). The DXMCT-1 was operated in energy sweeping mode, where energy thresholds of the ASICs were decreased with an increment of -2 mV after every acquisition. Counts above the corresponding threshold were acquired at each acquisition. The measurements were repeated 5 times for each x-ray tube voltage settings, 29 times for ^{99m}Tc, 24 times for ^{57}Co, respectively. The mean counts of multiple measurements at the same threshold mV value were calculated. A count between two thresholds was calculated by subtracting the mean counts acquired by adjacent threshold settings. The resulting energy spectrum is called an uncalibrated spectrum.

The mV values that correspond to the energies of the tube voltage or the radioisotopes were found with the uncalibrated spectrum for each threshold/comparator. From the five mV values, three parameters, c_1, c_2, and c_3 of Eq. (1) were estimated by the least squares method.

Equation (1) was then evaluated at five energies and the coefficients of variation (COV) (i.e., the ratio of the root mean square difference to the mean of five mV values presented in a percentage) were calculated for each threshold.

B. *Deadtime losses*

We used the same setting as outlined in Sec. III.A. The DXMCT-1 was placed 150 mm from an x-ray focus (see Fig. 1, right). The x-ray tube voltage and the time period per acquisition were fixed at 80 keV and 10 ms, respectively. The tube current values used were 10, 25, 50, 75, 100, 200, 300, 400, 500, 800, and 1000 μA. The DXMCT-1 was operated in energy sweeping mode. The measurements were repeated 20-28 times in each tube current setting. Notice that considering a geometrical magnification factor, the x-ray intensity with the tube current of 1000 μA corresponded to that with 54 mA with no bowtie filter for clinical CT scanners.

For each threshold/comparator, the mean of counts over multiple measurements, a_R, was obtained at each tube current setting. Then, the deadtime τ and the conversion coefficient k for both NP and P detection models were estimated from the counts obtained by nine tube current values ≤ 500 μA using the weighted least squares method.

The mean and the standard deviation of each of τ and k over all of the thresholds were calculated. The count rate model, Eq. (3), was evaluated at nine tube current settings and COV was calculated for both NP and P detection models.

C. *Distorted, recorded spectrum with pulse pileup effects*

The same data outlined in Sec. III.B were used. The mean counts of multiple measurements at the same threshold mV value were calculated for each tube current setting. A count between two thresholds was calculated by subtracting the mean counts acquired by adjacent threshold settings. This uncalibrated spectrum was then scaled and converted to the recorded spectrum, $N_{PCXD}(E)$, in the unit of counts per keV with an increment of 1 keV using the result of energy response calibration performed in Sec. III.A.

The true incident spectrum $S_0(E)$ was estimated for each threshold from the recorded spectrum at 10 μA. Then, for each threshold, the recorded spectra $N_{PPE}(E)$ for both NP and P detectors were estimated for each tube current setting using the pulse pileup model described in Sec. II.E, the deadtime τ and the conversion coefficient k estimated in Sec. III.B, and the incident spectrum $S_0(E)$ obtained above.

As a reference, a scaled incident spectrum with the recorded count rate, $N_{SCL}(E)$,

$$N_{SCL}(E) = a_R \times S_0(E), \quad (6)$$

was also obtained for each tube current setting.

At each tube current setting the root mean square difference (RMSD) between the spectrum estimated by the model, $N_{PPE}(E)$, and the mean spectrum measured by DXMCT-1, $N_{PCXD}(E)$, was calculated over the energy range between 30 keV and 150 keV. The COV was then calculated by dividing RMSD by the mean of the measured counts per keV between 30 keV and 150 keV. The RMSD and COV were also calculated between $N_{SCL}(E)$ and $N_{PXCD}(E)$.

IV. Evaluation Results

We identified 14 bad pixels among 512 pixels (thus 28 out of 1024 thresholds) which had untypical outputs, e.g., extremely large count rates, convex energy-mV curves, etc. and excluded them from this study. Thus, 996 thresholds/comparators in total were examined.

A. *Energy response*

The mean and the standard deviation of the three parameters were c_1 = 848.3 ± 5.3 [mV/keV], c_2 = 688.4 ± 18.7 [mV/keV], and c_3 = 0.0310 ± 0.0014 [/keV], respectively. Figure 3 shows the mean energy response over all thresholds. The standard deviation calculated over all thresholds was small. The mean COV value of all thresholds was as small as 0.2 %, which demonstrated that Eq. (1) modeled the energy response very well.

B. *Deadtime losses*

The mean and the standard deviation of the deadtime τ and conversion coefficient k over all thresholds were τ = 146.9±6.3 [ns] and k = 30.93±3.17 [× 10^3 cps/μA] for the non-paralyzable

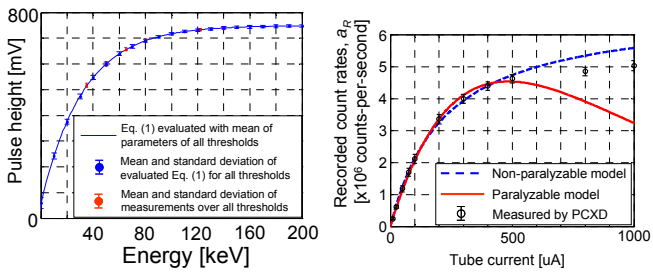


Fig. 3. (Left) The mean energy response curve, i.e., the photon energy-pulse height curve, of all thresholds. (Right) Estimated and measured count rate curves of mean of all of the thresholds.

detector model and τ = 81.2±2.2 [ns] and k = 25.35±1.88 [× 10^3

cps/μA] for the paralyzable detection model, respectively. The standard deviation of deadtime τ over all thresholds was quite small: 4.3% for the NP detection model and 2.7% for the P detector model, respectively.

The mean recorded count rates over all thresholds were estimated using Eqs. (2)-(3) and the parameters obtained above, for both NP and P detection models, and were plotted against tube currents in Fig. 3. The estimated count rates showed excellent agreement with those measured by DXMCT-1 (see Fig. 3).

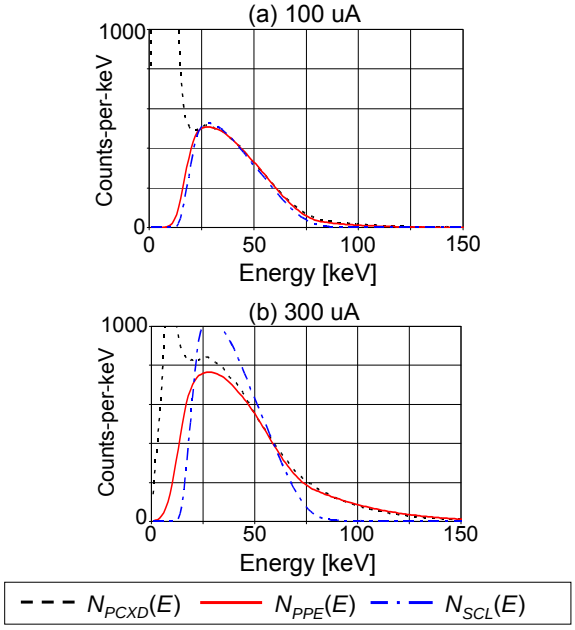


Fig. 4. The following three energy spectra for a tube setting of 80 kVp are shown: the mean energy spectrum measured by all of the thresholds of DXMCT-1 (labeled $N_{PCXD}(E)$ in the figure); the energy spectrum predicted by the model of the spectral distortion due to pulse pileup effects with the **paralyzable** detection model (labeled $N_{PPE}(E)$); and the scaled incident spectrum, $N_{SCL}(E)$, shown in Eq. (6) (labeled $N_{SCL}(E)$). The estimated deadtime loss ratio under the four tube current settings was 19% loss at 100 μA and 46% loss at 300 μA, respectively.

C. *Distorted, recorded spectrum with pulse pileup effects*

Figure 4 shows the mean recorded spectra over all thresholds with the P detection model. The estimated true count rate and count rate loss ratio at each of the four tube current settings were 2.0×10^6 cps and 19% loss at 100 μA and 7.6×10^6 cps and 46% loss at 300 μA, respectively. The pulse pileup effects can be observed with the measured spectrum: there were more counts recorded at energies above 80 keV with larger tube currents than with smaller tube currents. Notice that despite the challenging large amount of count rate loss, the spectrum predicted by the pileup model agreed reasonably well with the measured spectrum especially at energies higher than 50 keV. In contrast, deviations between the scaled incident spectrum and the measured spectrum increased as increasing the tube currents, especially with energies higher than 50 keV. Counts recorded below 15 keV in the measured spectrum seemed to contain a lot of electric noise.

Table I summarizes the quantitative analysis of the results. The COVs between the spectra measured and those predicted by a pulse pileup model with P detector were in the agreement of 3.7-7.2 % with tube currents of 100-300 μA where 19-46 % of counts were lost.

V. Conclusion

It has been demonstrated that the performances of the PCXD agreed exceptionally well with analytical models, regarding the energy response, the count rate loss, and the pulse pileup effects. We believe the models are sufficiently accurate so that one can develop the corresponding compensation schemes based on these models.

Table I. The RMSD and COV calculated against the mean of the recorded spectra measured by all of the thresholds of the PCXD.

		Tube current, μA [mean counts]							
	Scheme	25 [30.9]	50 [63.9]	75 [97.4]	100 [119.6]	200 [192.9]	300 [234.7]	400 [261.6]	500 [277.5]
RMSD [cnts]	NP*, $N_{PPE}(E)$§	10.8	7.8	11.1	12.3	53.0	73.9	85.3	87.5
	NP*, $N_{SCL}(E)$†	6.6	17.3	30.8	37.2	62.7	80.5	104.4	128.8
	P#, $N_{PPE}(E)$§	3.7	1.7	5.8	4.5	10.0	17.1	41.3	71.8
	P#, $N_{SCL}(E)$†	3.9	7.5	11.6	16.5	49.7	95.2	130.5	153.3
COV [%]	NP*, $N_{PPE}(E)$§	32.7	11.7	11.6	10.2	27.3	31.3	32.4	31.3
	NP*, $N_{SCL}(E)$†	20.2	25.9	32.1	30.7	32.3	34.1	39.6	46.0
	P#, $N_{PPE}(E)$§	11.3	2.6	6.0	3.7	5.2	7.2	15.7	25.7
	P#, $N_{SCL}(E)$†	12.0	11.3	12.1	13.6	25.6	40.3	49.5	54.8

*Non-paralyzable detection model § Estimated by pulse pileup model, Eq. (6)
#Paralyzable detection model † Scaled incident spectrum, Eq. (15)

Acknowledgment

We thank Zhihui Sun, M.Sc., and Hideaki Tashima, Ph.D., for their help with data acquisition and Jochen Cammin, Ph.D., Somesh Srivastava, Ph.D., and Ronald J. Jaszczak, Ph.D., for their helpful discussions and suggestions.

References

[1] G. F. Knoll, *Radiation detection and measurement*, 3rd ed. New York: John F. Wiley and Sons, 2000.

[2] W. C. Barber, E. Nygard, J. S. Iwanczyk, M. Zhang, E. C. Frey, B. M. W. Tsui, J. C. Wessel, N. Malakhov, G. Wawrzyniak, N. E. Hartsough, T. Gandhi, and K. Taguchi, "Characterization of a novel photon counting detector for clinical CT: count rate, energy resolution, and noise performance," in *SPIE Medical Imaging 2009: Physics of Medical Imaging*, 1 ed. vol. 7258 Lake Buena Vista, FL, USA: SPIE, 2009, pp. 725824-9.

[3] J. S. Iwanczyk, E. Nygard, O. Meirav, J. Arenson, W. C. Barber, N. E. Hartsough, N. Malakhov, and J. C. Wessel, "Photon Counting Energy Dispersive Detector Arrays for X-ray Imaging," *Nuclear Science, IEEE Transactions on*, vol. 56, pp. 535-542, 2009.

[4] K. Taguchi, E. C. Frey, X. Wang, J. S. Iwanczyk, and W. C. Barber, "An analytical model of the effects of pulse pileup on the energy spectrum recorded by energy resolved photon counting x-ray detectors," in *SPIE Medical Imaging 2010: Physics of Medical Imaging* San Diego, CA: SPIE, 2010.

[5] E. C. Frey, X. Wang, Y. Du, K. Taguchi, J. Xu, and B. M. W. Tsui, "Investigation of the use of Photon Counting Detectors with Energy Discrimination Capability for Material Decomposition in Micro-computed Tomography," in *SPIE Medical Imaging 2007: Physics of Medical Imaging*, 1 ed. vol. 6510 San Diego, CA, USA: SPIE, 2007, pp. 65100A-1-11.

Sinogram restoration algorithm for photon counting clinical X-ray CT with pulse pileup compensation

Somesh Srivastava[1], *Member, IEEE* , Katsuyuki Taguchi[1], *Member, IEEE*

Abstract—**Photon counting x-ray detectors (PCXD) can potentially boost the usefulness of clinical x-ray computed tomography (CT) scanners by efficiently measuring x-ray spectrum information [4,5]. A major hindrance in the adoption of PCXDs in the clinical setting is their limited speed. But, clinical CT imaging with such PCXDs was shown to be feasible provided four methods were developed [2]. Here, we present the development of one of those four methods: a sinogram restoration algorithm that includes pulse pileup as a part of the forward model (i.e., imaging process). We use simulation studies to show that with the use of the proposed algorithm, useful data can be obtained from PCXDs even when they are operating at a 20% count rate loss. This implies a 4-fold increase in the operational count rate of PCXDs.**

***Index Terms*—Computed tomography, photon counting, sinogram restoration .**

I. Introduction

X-RAY computed tomography (CT) is one of the *foremost* diagnostic imaging modalities in the clinic. It is widely used and is the only imaging modality of choice when a high-resolution image (≈ 1 mm) is required in a short scan duration (< 1 s) and the benefits of such a scan outweigh the risk of side-effects due to x-ray radiation exposure. However, current x-ray CT scanners do not utilize energy spectrum information. X-ray CT imaging methods that use energy information would have many advantages: improved contrast-to-noise ratio in images, reduced radiation dose, improved material decomposition, etc.

Recent research has indicated that photon counting x-ray detectors (PCXDs) with energy discrimination capabilities can provide all these advantages. But, a major factor preventing the incorporation of PCXDs into clinical x-ray CT scanners is their limited speed. When detectors are "slow," two or more x-ray photons arriving at the detector during its dead-time are recorded as a single count with a different energy. The resulting distortion of the x-ray spectrum is called the pulse pileup effect, and the loss of counts is termed count rate loss.

The solution to the problem of pulse pileup and count rate loss may be based on hardware or software [2] or both. Recently developed PCXD technology, based on Cadmium-Telluride (CdTe) and an advanced application-specific integrated-circuit (ASIC), has achieved a maximum output count rate (i.e. speed) of 6 Mcps (million counts per second per square millimeter) with two energy thresholds [1]. Even though this count rate is an achievement in its own right, it is still insufficient to avoid pulse pileup effect and count rate loss if used for clinical x-ray CT scanners, where incident count rates could be as high as 1000 Mcps [2].

While improvements in detector hardware could one day handle such high count rates, a combination with numerical (software) techniques such as Ref. [2] is desirable. Taguchi, et al., proposed to combine the following three techniques [2]: (1) a bowtie filter to reduce the attenuated count rate (which is incident onto the detector) to a manageable level; (2) a pulse pileup correction algorithm using an accurate model of the pulse pileup effect of the PCXD; and (3) the exact interior region-of-interest reconstruction algorithm to obtain exact images in the interior of the scanned object (while avoiding the inaccurate data obtained at higher count rates near the edge or outside the object).

Two key parts of the second technique are the pulse pileup model and the sinogram restoration algorithm. The analytical pulse pileup model has been recently developed [3]; the sinogram restoration algorithm has been developed in this study, using the concept of material decomposition. The pulse pileup model was used in the forward model to estimate the thicknesses (i.e., line integrals) of basis functions of the material decomposition for each ray. Images were then reconstructed from the estimated sinogram (line integrals).

The contribution made in this paper is a new sinogram restoration algorithm whose forward imaging model includes pulse pileup.[1] To our knowledge, the proposed algorithm is the only one that incorporates the sophisticated pulse-pileup model developed in Ref.[3].

This paper is organized as follows. In section II, we present an overview of the solution presented in [2]. In section III, we present the proposed sinogram restoration algorithm. Section IV presents the simulation setup. Section V discusses methods and results, and section VI presents conclusions and future work.

This work was supported in part by National Institutes of Health under grant R44 EB008612.

S. Srivastava and K. Taguchi are with The Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins University School of Medicine, Baltimore MD. 21287. USA. (phone: 410-502-3773; fax: 410-610-1060; e-mail: someshs@jhu.edu)

[1] This algorithm is much more robust and sophisticated than the highly simplified proof-of-concept algorithm first introduced by our group for a much simpler pulse pileup model at the NSS-MIC 2008 conference in *Meng Xi, et al.*

II. Overview of the Proposed Solution

PCXDs can be used in clinical x-ray CT scanners with a combination of four methods proposed in [2]. Here, we summarize each of those methods.

A. Bowtie filter

A bowtie filter (see [2]) reduces the pulse pileup problem by reducing the count rate incident on a PCXD. It is an x-ray attenuator placed just in front of the x-ray source, and is thicker towards the periphery than the center. Peripheral detectors have a much worse pulse pileup problem than the center detectors due to a higher count rate towards the periphery, and are benefited by the bowtie filter as it attenuates some of the count rate.

B. Detector model

The detector model is used to model the characteristics of the PCXD in terms of pulse pileup effect and count rate loss. An x-ray photon incident on the PCXD creates an electrical pulse[2] in the detection circuit (see Fig. 1). If one or more photons arrive within the dead-time period triggered by the first photon, the two electrical pulses are added and observed as a single electrical pulse (see Fig. 1).

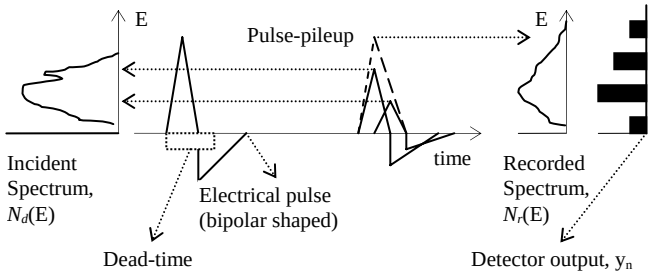


Fig. 1. A schematic diagram showing how the recorded energy differs from the incident spectrum with pulse-pileup effects.

Due to pulse pileup, the recorded spectrum, $N_r(\mathrm{E})$ differs from the incident spectrum, $N_d(\mathrm{E})$. The two are related as follows [3]:

$$N_r(E) = \left(\text{True count rate}\right) \times \left(\begin{array}{l}\text{Probability of}\\ \text{events being recorded}\end{array}\right) \times \sum_m \left[\left(\begin{array}{l}\text{Probability of}\\ \text{the } m\text{th order pulse pileup event}\end{array}\right) \times \left(\begin{array}{l}\text{Probability of}\\ \text{counts recorded at an energy } E\\ \text{with } m\text{th order pulse pileup}\end{array}\right)\right]$$

$$= N_t \times \Pr\left(rec \mid N_t\tau\right) \times \sum_{m=0}^{\infty} \Pr\left(m \mid rec\right) \Pr\left(E \mid m\right)$$

where, N_t is the true count rate, τ is the detector dead-time, and m is the order of the pileup, where m+1 photons are recorded as one count. We assume a non-paralyzable detector model, i.e. the detector always returns to the active state τ seconds after a photon arrival. For this model, $\Pr(rec|\ N_t\tau) = 1/(1+ N_t\tau)$ and $\Pr(m|rec) = (N_t\tau)^m \exp(-N_t\tau)/m!$. $\Pr(E|m)$ is formulated using the incident spectrum, $N_d(\mathrm{E})$, and the parameters of the bipolar pulse-shape, and can be found in [3]. The detector output, y_n, is the number of photons recorded in a set of N_{bin} energy-bins, (E_n, E_{n+1}), $n = 1, \ldots, N_{bin}$, $E_{Nbin+1} = \infty$, that span the recorded energy-range, i.e. $y_n = \int w_n(E) N_r(E) \mathrm{d}E$ over (E_n, E_{n+1}), where $w_n(E)$ is a normalized window function for energy window n.

[2] The shape of the electrical pulse in the PCXD studied here is bipolar, which was modeled by two triangles [3].

C. PCXD compensation scheme

The PCXD compensation scheme proposed here uses a material decomposition approach (see Section III.A) and a penalized-likelihood sinogram restoration algorithm to compensate for the pulse pileup effect. For a given attenuation image of the object, we can compute the transmitted x-ray spectrum from the spectrum exiting the bowtie (obtained during calibration). And, for a given transmitted x-ray spectrum, we can use the pulse pileup detector model (see Section II.B) to compute the recorded counts. Thus, by minimizing (see Section III.C) the mismatch between the measured and computed recorded counts, which is given by the penalized-likelihood cost function (see Section III.B), we can determine the attenuation image that is compensated for pulse pileup.

D. Exact interior ROI reconstruction

The corrected sinogram obtained above might still have inaccuracies for peripheral detector channels where count rates are too high. For such cases, the exact interior ROI reconstruction based on Kudo's algorithm can be used. Kudo's algorithm uses *a priori* knowledge of pixel values in a small region inside the ROI to reconstruct an exact image in the interior of the object.

III. Sinogram Restoration Algorithm

The sinogram restoration algorithm that compensates for the pulse pileup effect in a PCXD is presented here. Below, we outline four major aspects of the algorithm: the object model based on material decomposition, the imaging model for a PCXD, penalized-likelihood cost function, and the minimization algorithm.

A. Object model based on material decomposition

The object being imaged is represented as the distribution in space of the x-ray attenuation coefficient at different energies, $f(\underline{x}, E)$. It is represented as a sum of L basis materials as: $f(\underline{x},E)=\Sigma_{k=1}^{L} a_k(\underline{x})\rho_k \mathrm{m}_k(E)$, where k is the material index, ρ_k is the material density, $m_k(E)$ is the mass attenuation coefficient, and $a_k(\underline{x})$ is the basis coefficient of the kth material at location $\underline{x}$. The goal of the proposed algorithm is to compute the corrected line integral for basis material k and ray i of the sinogram: $l_{i,k} = \int_{ray(i)} a_k(x) d\underline{x}$. Once $l_{i,k}$ is computed, $f(\underline{x}, E)$ can be easily reconstructed using, e.g., an analytical reconstruction method.

B. Imaging model for a PCXD

The imaging model for a PCXD is described based on the object model as follows. The spectrum incident on the detector for ray i is given by:

$$N_{d,i}(E) = N_{b,i}(E)\exp(-\sum_{k=1}^{L} l_{i,k}\rho_k m_k(E)),$$

where, $N_{b,i}(E)$ is the spectrum exiting the bowtie, which is determined during calibration. The incident spectrum, $N_{d,i}(E)$, is distorted by pulse pileup into the recorded spectrum, $N_{r,i}(E)$, as described in Section II.B. The recorded spectrum is binned into energy bins by the PCXD to give us the recorded counts, $y_{n,i}$. It is also known that $y_{n,i}$ are independent, Poisson distributed random variables [3]. The imaging model can be now summarized as: $l_{i,k} \xrightarrow[x-ray\ spectrum]{Polychromatic} N_{d,i}(E) \xrightarrow[pileup]{Pulse} N_{r,i}(E) \xrightarrow[Poisson\ noise]{Binning} y_{n,i}$.

C. Penalized-likelihood cost function

The corrected line integrals, $\underline{l}$, are estimated by minimizing the Penalized likelihood (PL) cost function, $\Phi(\underline{l})=-L(\underline{y};\underline{\bar{y}}(\underline{l}))+R(\underline{l})$, where $-L$ is the negative log-likelihood function and $R(\underline{l})$ is the regularization function.

The first term is the negative log–likelihood, $-L$,which measures the degree of mismatch between the recorded counts measured during the scan, $y_{n,i}$, and the recorded counts computed using the imaging model, $\bar{y}_{n,i}(\underline{l}_i)$, $\underline{l}_i = (l_{i,1},...,l_{i,L})$. (Since $\bar{y}_{n,i}$ is a function of $\underline{l}_i$ through the imaging model, it written as $\bar{y}_{n,i}(\underline{l}_i)$.) Since $y_{n,i}$ are independent and Poisson distributed [3] with mean $\bar{y}_{n,i}(\underline{l}_i)$,

$$-L(\underline{y};\underline{\bar{y}}(\underline{l})) = \Sigma_i\Sigma_n\ \bar{y}_{n,i}(\underline{l}_i) - y_{n,i}\log\bar{y}_{n,i}(\underline{l}_i).$$ Thus, $-L$ incorporates both the object model based on material decomposition and the imaging model that models spectrum distortion due to pulse pileup.

The second term of the cost function, the regularization function, is to improve algorithm stability by including prior information about the estimated parameters, like smoothness. For now, we use a quadratic penalty function that regularizes within a single view only as it allows parallelizing the algorithm. $R(\underline{l})=1/2\Sigma_k\beta_k\Sigma_i\Sigma_j w_{i,j}(l_{i,k}-l_{j,k})^2$, where β_k is the regularization parameter for the kth material, and $w_{i,j}$ is the neighborhood function, which equals 1 when rays i and j are neighbors and equals 0 otherwise.

We assume the following about $\Phi(\underline{l})$: a non-linear, non-convex, continuous function, with continuous gradient, which could possibly have local minima. It is difficult to find an explicit expression for its gradient due to the complex nature of the expression for pulse-pileup.

D. Minimization algorithm

The minimization algorithm estimates the corrected sinograms for L basis materials by minimizing the PL cost function, $\Phi(\underline{l})$. Since, finding a one step solution to the minimization problem is difficult, we use iterative methods (see Fig. 3). We start by computing a good initial guess, which is necessary since $\Phi(\underline{l})$ could have local minima. Next, we note that each view of the sinogram can be updated in parallel, due to the design of $R(\underline{l})$. Within each view, we update for $\underline{l}_i$, considering all other sinogram bins of that view as constants – a grouped coordinate descent (GCD) approach. The update for $\underline{l}_i$ is compute using the conjugate gradient (CG) method [6] and a backtracking line-search (Brent's method [6]). The gradient for the CG method is computed by numerical means as an analytical expression for the gradient is hard to determine. The complete minimization algorithm is summarized in Fig 3.

```
* Compute initial guess of l
* Loop over iterations
    * Loop over views (Parallel)
        * Loop over channels, i, of a particular view (GCD)
            * Update l_i using CG and line-search
              Do {
                    * Compute gradient of Φ(l) numerically
                    * Compute search direction using CG
                * Use backtracking line-search to update l_i }
              While (change in l_i is large)
```

Fig. 3. Summary of the minimization algorithm.

IV. Simulation Setup

We simulate a single-slice fan-beam x-ray CT scanner with an arc-detector. The source-to-isocenter distance is 570 mm, source-to-detector distance is 1140 mm, and fan-angle is ~52°. The number of detector channels is 384. 360 views are collected from a 360° scan of duration 1s. The tube voltage is 120 kV, and the tube current is adjusted so that the count rate exiting the bowtie and incident on the outermost detector (which is in the air path) is 10 Mcps. The bowtie filter is made of Aluminum and its thickness varies quadratically from 1 mm for the central ray to 50 mm for peripheral rays. For the PCXD, 5 energy thresholds were simulated: 20, 50, 70, 90, and 120 keV. The dead-time of the PCXD was 100 nsec.

A water cylindrical phantom of diameter 30cm contained two bone cylindrical inserts of diameter 5cm.

The size of the reconstructed images for the proposed and baseline methods in Section V is 256×256 with pixel size 1.25 mm $\times$ 1.25 mm.

C++ was used to code the sinogram restoration algorithm. 36 cores, each equivalent to one core of Intel Xeon 2.33 GHz, were used for an execution time of 2 hours.

V. Methods and Results

The proposed sinogram restoration algorithm, called the PL method here, is compared with a baseline method, an uncompensated reconstruction approach called the direct

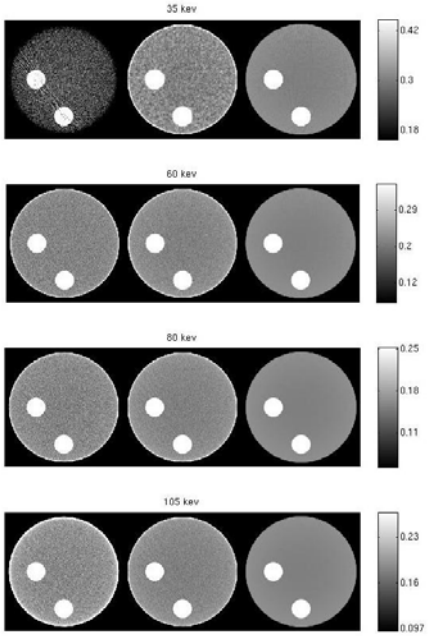


Fig. 4. Images of the x-ray attenuation coefficient (in cm^{-1}) at 4 energies. Left: direct method, center: PL method, right: truth.

method here. Both methods reconstruct images of the x-ray attenuation coefficient at 4 energies: 35, 60, 80 and 105 keV. The PL method was used to obtain line integrals of material density of basis materials (water and bone) for each ray. Then, the images of spatial distributions of densities of basis materials were computed by a standard filtered backprojection and converted to the attenuation images using the object model (see Section III.A). The direct method corrects the recorded counts for count rate loss before directly reconstructing attenuation images using the detector output. The air scan for the direct method avoids pulse pileup by using a low tube current and later scaling the recorded counts to the tube current used for the phantom.

The attenuation images at 4 energies for the PL method, the direct method and the truth are presented in Fig 4. The images from the PL method have much lesser noise, and the bias is also lower, especially for the 35 keV case. To quantify bias and noise, we first segment the image into water and bone regions (while avoiding the edge regions completely) and compute the mean and the standard deviation of attenuation coefficients of water and bone at 4 energies. Figs 5 and 6 and Table 1 show that the PL method outperforms the direct method in both mean and standard deviation for both water and bone. The direct method is especially poor for the 35 keV case where the signal is very noisy due to very few photons, while the PL method performs much better because it uses counts over the entire spectrum.

VI. Conclusions and Future Work

Based on the simulations presented here, we conclude that with the use of the proposed algorithm, PCXDs can yield useful data for as much as 20% count rate loss. This is equivalent to a 4-fold increase in the useful output count rate of a PCXD – from 0.5 Mcps to 2 Mcps, for a dead-time of 100 nsec. Thus, the proposed algorithm brings us one step closer to using PCXDs in clinical x-ray CT scanners.

In future, the proposed algorithm shall be expanded to include a third basis material, e.g. a contrast agent. Also, the regularization will be changed to take differences across views, and the minimization algorithm shall also change accordingly.

Acknowledgement

We thank Mengxi Zhang, M.Sc., and Jingyan Xu, Ph.D., for insightful discussions and help with the preliminary version of the algorithm proposed here. We also thank William C. Barber, Ph.D., Benjamin M. W. Tsui, Ph.D., and Eric C. Frey, Ph.D., for their support.

References

[1] W. C. Barber, E. Nygard, J. S. Iwanczyk, M. Zhang, E. C. Frey, B. M. W. Tsui, J. C. Wessel, N. Malakhov, G. Wawrzyniak, N. E. Hartsough, T. Gandhi, and K. Taguchi, "Characterization of a novel photon counting detector for clinical CT: count rate, energy resolution, and noise performance," in *SPIE Medical Imaging 2009: Physics of Medical Imaging*, Lake Buena Vista, FL, USA, 2009, pp. 725824-9.

[2] K. Taguchi, S. Srivastava, H. Kudo, and W. C. Barber, "Enabling photon counting clinical X-ray CT," in *IEEE Nuclear Science Symposium and Medical Imaging Conference*, (IEEE, Orlando, USA, 2009).

[3] K. Taguchi, E. C. Frey, X. Wang, J. S. Iwanczyk, and W. C. Barber, "An analytical model of the effects of pulse pileup on the energy spectrum recorded by energy resolved photon counting x-ray detectors," in *SPIE Medical Imaging 2010: Physics of Medical Imaging*, San Diego, CA, 2010.

[4] E. Roessl, and R. Proksa, "K-edge imaging in x-ray computed tomography using multi-bin photon counting detectors," *Physics in Medicine and Biology*, 52 (2007), pp. 4679-4696.

[5] J.P. Schlomka et al., "Experimental feasibility of multi-energy photon-counting K-edge imaging in pre-clinical computed tomography," *Physics in Medicine and Biology*, 53 (2008), pp. 4031-4047.

[6] W. H. Press, S. A. Teukolsky, W. T. Wetterling, and B. P. Flannery, "Numerical recipes 3rd edition: the art of scientific computing," Cambridge University Press, 2007, New York, NY, USA.

TABLE I
PERCENTAGE REDUCTION IN STANDARD DEVIATION - FROM DIRECT METHOD TO PL METHOD

	Water	Bone
35 keV	57	79
60 keV	26	36
80 keV	38	45
105 keV	44	44

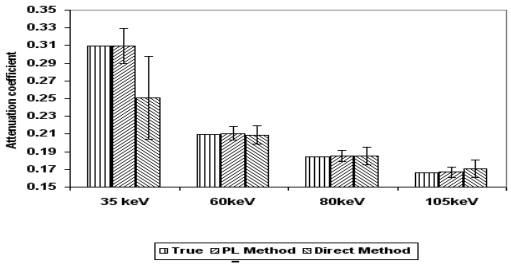


Fig. 5. Mean value of attenuation coefficient (in cm^{-1}) of water at 4 energies. (Error bars are +/- 1 standard deviation.)

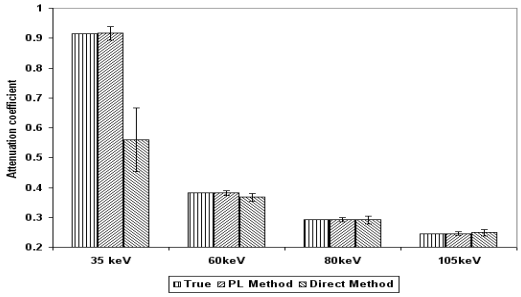


Fig. 6. Mean value of attenuation coefficient (in cm^{-1}) of bone at 4 energies. (Error bars are +/- 1 standard deviation.)

A Reconstruction Technique for Dual-Energy X-ray Computed Tomography

J. Hsieh, *Sr. Member, IEEE*, J. Fan, N. Chandra, P. Crandall, M. Kulpins

***Abstract*—Beam hardening artifact is caused by the polychromatic nature of the x-ray source and the energy-dependent characteristics of the scanned object. Despite the significant research activities in this area, beam-hardening artifact remains one of the major issues in x-ray computed tomography (CT). With the recent technology advancements, dual-energy CT has become commercially available and clinically feasible. In this paper, we present a projection-space beam-hardening correction technique to effectively overcome such artifact. The approach is based on the differential signal from the water-calibrated dual-energy projections, and its efficacy is demonstrated by theoretical analysis and phantom experiments.**

***Index Terms*—Dual-energy CT, Beam-hardening Artifact, Material Decomposition, Monochromatic Imaging**

I. Introduction

Dual energy CT (DECT) has been one of the most active areas of research for x-ray computed tomography (CT) in recent years. Although the concept of the DECT was proposed in the late 1970s, not long after the introduction of the first commercial CT scanner [1-4], technological limitations prevented this feature from becoming a useful clinical tool. The limitations include the slow data acquisition speed to control the patient motion and low x-ray tube power to overcome the increased noise. With the introduction of helical and multi-slice CT in recent years, the technological barrier of DECT has been significantly reduced or eliminated. Many research activities and advancements have been reported, and new clinical applications have been explored [5-14].

In most DECT data acquisition, two sets of projections are acquired, one at a higher tube voltage setting (e.g., 140kVp) and the other at a lower tube voltage setting (e.g., 80kVp). The energy-dependent attenuation characteristic of a material enables the two datasets to provide additional information about the material than a single kVp projection. The extra information can be used to effectively reduce or eliminate the beam-hardening effect and provide improved characterization of the scanned object.

There are different approaches in the data collection of DECT. The two datasets can be collected sequentially or simultaneously. In the sequential data collection, one set of projection is collected first at high kVp (or low kVp) and the generator and tube are switched quickly to collect the low kVp projections. The advantage of this approach is its simplicity since any conventional CT scanner can be programmed to accomplish such a data acquisition. The disadvantage is its inability to handle patient motion. Note that the time difference between the high-kVp and low-kVp is on an order of a second and significant patient motion can take place during that time interval. Since DECT relies on a good registration between the two datasets, patient motion can lead to artifacts and a significant compromise in dual-energy performance.

In the simultaneous data collection scheme, one can use two sets of tube-detector pairs that are offset by at least 90^o in projection angle, and use one pair to produce the high kVp data and the other the low kVp data [15]. The two datasets are collected at the same time, but at different orientations. Alternatively, a conventional CT scanner can be redesigned in which the tube and generator switch quickly between high- and low-kVp settings in adjacent projections, as shown in Fig. 1. Since the time interval between two successive projections is a small fraction of a millisecond, and the angular difference between them is significantly smaller than a degree, the high- and low-kVp projections are acquired nearly simultaneously at nearly the same orientation. This method is called fast kVp-switching.

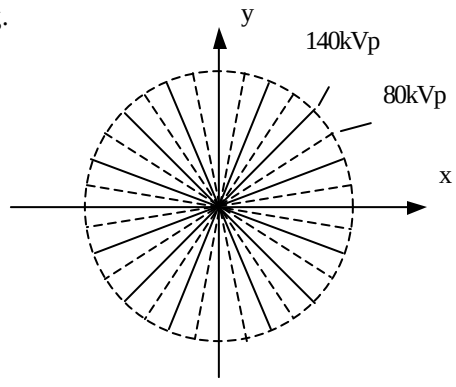


Fig. 1 Schematic diagram of fast-kVp switching

One of the key advantages of a dual-energy approach is the potential to significantly reduce metal and beam-hardening artifact which has been present in CT images since its first

J. Hsieh, J. Fan, N. Chandra, P. Crandall, and M. Kulpins are with GE Healthcare Technologies, Waukesha, Wisconsin (e-mail for the corresponding author: jiang.hsieh@med.ge.com).

J. Hsieh is an adjunct professor in the Medical Physics Department, University of Wisconsin, Madison.

invention in the early 70s. In this paper, we describe an approach that significantly suppresses the metal and beam-hardening artifacts by leveraging the different information provided in the dual-energy projections.

II. METHOD AND MATERIAL

Before a detailed discussion on the proposed approach, a general description of the current reconstruction process is in order. In the conventional CT reconstruction, the collected scan data first undergoes a series of calibration steps to remove potential non-ideal performance of the CT subsystems and inherent nonlinear behavior of the x-ray physics. One of the calibration steps is the so-called water-calibration. In this step, the original projection is mapped to a new set of projections based on the attenuation characteristics of water. This step enables the removal of the beam-hardening effect due to water or soft-tissue.

Our approach is based on the observation that "water" beam-hardening correction is quite accurate if the scanned object is made of water, soft-tissue, or similar material, and the amount of beam-hardening artifact present in the reconstructed CT images is negligible. The water beam-hardening correction can be carried out either in the projection domain or image domain. In the projection domain, the measured projection after the minus logarithm operation is mapped by a high order polynomial to a new set of projections prior to the filtered backprojection (FBP) process. The polynomial coefficients are determined based on phantom measurements and theoretical calculations. As a result, the line integral of the attenuation coefficients accurately represents the "water" portion of the scanned object. Since in x-ray CT, the CT number of water is defined to be zero regardless of the acquisition x-ray tube voltage setting, the portion of the projection contributed by water (or the soft-tissue) in either the high- or the low-kVp settings are identical.

The residual beam-hardening and metal artifacts should then be solely due to the presence of materials that are significantly different from the water in terms of their attenuation characteristics. As a result, to remove such artifacts, we should focus our attention on the portions of the projection that are contributed by non-water materials. If we examine the differential signals between two projections of different kVps collected at the same projection angle, the signals should contain, in theory, the contributions of the non-water material only. For illustration, Fig. 2(a) shows the projection profiles of a water-calibrated phantom scan collected at 140kVp and 80kVp respectively, and Fig. 2(b) shows the difference projection. Note that the "water" portion of the projection is effectively removed.

In the "water" region of the projection, the noise of the difference projection is larger than either the 80kVp or the 140kVp projections. Based on the previous analysis, the signal in the water portion of the projection should be essentially zero and contain no useful information. The increased noise in the water region only increases the noise in the corrected projection, and the contribution of these signals should be suppressed. One way of accomplishing this is by scaling the difference projection prior to its use in the correction:

$$D(g,b,z) = \begin{cases} p_L(g,b,z) - p_H(g,b,z), & p_L(g,b,z) - p_H(g,b,z) \geq t \\ x\left[p_L(g,b,z) - p_H(g,b,z)\right], & p_L(g,b,z) - p_H(g,b,z) < t \end{cases} \quad (1)$$

where γ, β, and z denote the channel angle, view angle, and detector row, respectively. $D(\gamma,\beta,z)$ represents the processed difference projection, $P_L(\gamma,\beta,z)$ is the water-calibrated projection sample collected with the low kVp, $P_H(\gamma,\beta,z)$ is the water-calibrated projection of the high kVp, and $\xi(x)$ is a non-increasing function of x. Alternatively, adaptive filtering can be employed to suppress the noise in the difference projection. The amount of smoothing depends on the magnitude of the difference signal.

(a)

(b)

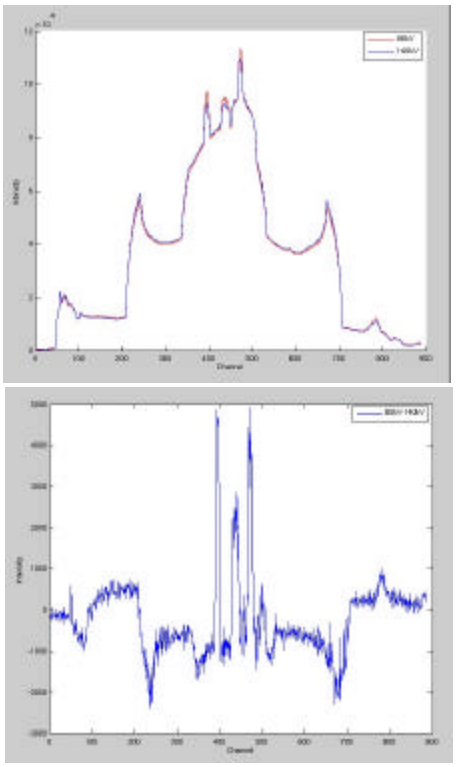


Fig. 2(a) Water-calibrated phantom projection profiles of 140kVp (blue) and 80kVp (red) scans
(b) Difference projection

Given the nature of metal artifact and beam-hardening, we know that the error is caused by the inaccurate measurement of the projection in the non-water object. Therefore, we want to estimate the error, $E_H(\gamma,\beta,z)$ and $E_L(\gamma,\beta,z)$, based on $D(\gamma,\beta,z)$. We develop a mapping function, *f*, between *E* and *D* based on theoretical calculation and phantom experiments of known materials. In the implementation, we use polynomials to establish the relationship between the two:

$$E_H(g,b,z) = \sum_{k=0}^{K} a_k \cdot D(g,b,z)^k \quad (2)$$

$$E_L(g,b,z) = \sum_{k=0}^{K} b_k \cdot D(g,b,z)^k$$

where α_k and β_k are the coefficients of polynomials, and are determined experimentally. Although the coefficients appear to be constants in the equation, they change based on the intensity of the difference projections. That is, α_k and β_k change as functions of $D(\gamma,\beta,z)$. Once the error projection is obtained, we can subtract the error from the original projection to obtain the projections without metal and beam-hardening effects:

$$p'_H(g,b,z) = p_H(g,b,z) - E_H(g,b,z) \qquad (3)$$
$$p'_L(g,b,z) = p_L(g,b,z) - E_L(g,b,z)$$

The corrected projection can then be used to reconstruct the high-kVp and low-kVp images. Note that these images are similar to the conventional CT images with the exception of the removal of beam-hardening artifacts.

Often, it is desirable to generate material-decomposed images. That is, the scanned object is depicted in a pair of density images of a pair of basis materials. For example, if soft-tissue and bone are selected as the basis materials, the bone will be automatically removed from the soft-tissue equivalent density image, and the soft-tissue will be removed form the bone density image. Other materials, such as iodine, will be distributed into both density images. Mathematically, this can be accomplished by applying another mapping prior to the reconstruction:

$$q_1(g,b,z) = h_1\left[p'_H(g,b,z), p'_L(g,b,z)\right] \qquad (4)$$
$$q_2(g,b,z) = h_2\left[p'_H(g,b,z), p'_L(g,b,z)\right]$$

where η_1 and η_2 are the functions to map from the high- and low-kVp projections to the material basis pair, m_1 and m_2. q_1 and q_2 are the material density projections of the selected material basis pair. The process to produce the mapping functions, η_1 and η_2, is well known and will not be discussed here. The FBP reconstruction algorithm can then be used on the density projections to generate equivalent density images for the basis materials m_1 and m_2.

Using the material density projections, one can easily generate a synthesized monochromatic projections (a set of projections mimicking the data acquisition if a monochromatic x-ray source is used) based on the following formulation:

$$p(g,b,z,E) = m_1(E)q_1(g,b,z) + m_2(E)q_2(g,b,z) \qquad (5)$$

where μ_1 and μ_2 are the mass attenuation coefficients for materials m_1 and m_2, respectively. E represents a pre-defined x-ray energy. The flow diagram of the reconstruction process is shown in Fig. 3.

It should be pointed out that since the beam-hardening correction is performed in the projection space, the additional computation to the reconstruction process is negligible. In addition, since the correction is performed on a channel-by-channel basis, little impact on spatial resolution is expected.

As is shown in the reconstruction flowchart in Fig. 3, the beam-hardening corrected low- and high-kVp projections can be used directly to produce conventional single-kVp images, or can be used to produce material decomposed density projections to generate material basis images or monochromatic images.

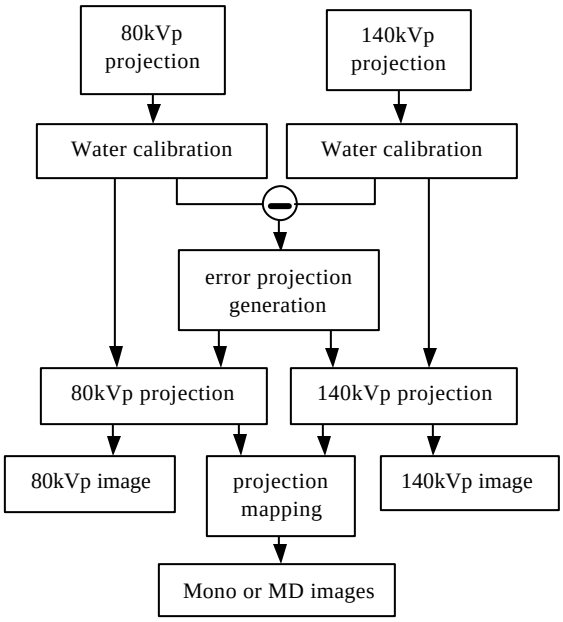


Fig. 3 Flowchart for image generation

III. EXPERIMENTAL RESULTS

To determine the efficacy of the proposed approach, several phantom experiments were conducted. The phantoms were scanned on a Discovery 750 HD scanner (GE Healthcare, Waukesha, Wisconsin) under GSI (Gemstone Spectral Imaging) acquisition mode. The scan data is regrouped into even and odd views to separate the high-kVp and low-kVp scan data. Both datasets undergo their respective water-calibration steps similar to the single-energy data acquisition and calibration. To enhance the appearance of the beam-hardening artifacts, a cardiac phantom was scanned with the inserts filled with diluted iodine contrast with concentration levels higher than typical clinical cardiac studies. A metal pin was placed on top of the cardiac phantom to produce additional artifacts. To establish the baseline, the water-calibrated 80kVp and 140kVp projections were used to produce two reconstructed images as shown in Figs. 4 (a) and (b). It is clear from the images that the beam-hardening artifacts are quite pronounced. The artifacts are shown as the dark shadings or streaks connecting the dense objects in the image. Note in particular the dark streaks connecting the metal pin and its nearby iodine inserts.

The projection dataset was then corrected with the proposed method and FBP algorithm was used to produce the reconstructed image of the cardiac phantom as shown in Fig. 4(c). Notice that the dark streaks between the iodine inserts

are nearly eliminated. Even the streaking artifact connecting the pin and the iodine inserts is substantially reduced. This experiment clearly demonstrates the effectiveness of the proposed correction algorithm in terms of beam-hardening artifact correction.

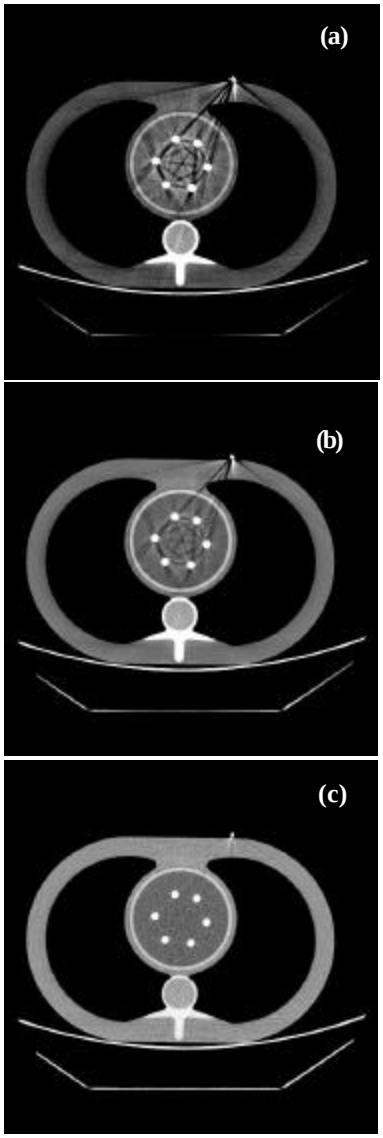


Fig. 4 Reconstructed images of a phantom experiment (a) 80kVp with water-correction, (b) 140kVp with water correction, (c) the proposed correction

IV. CONCLUSION

In this paper, we present a dual-energy reconstruction algorithm to combat beam-hardening and metal artifacts. The algorithm is based on the observation that the beam-hardening associated with the water or soft-tissue material is properly corrected with the water-calibration. The residual beam-hardening effect due to other material types can be corrected based on the difference of the dual-energy projections.

Phantom studies were conducted to investigate the efficacy of the proposed approach. Phantom results clearly demonstrate the effectiveness of the proposed algorithm, even under conditions worse than typical clinical environment, and with the presence of multiple materials. Furthermore, since the proposed correction is performed in the projection space, the computational impact to the overall reconstruction is kept to a minimum.

REFERENCES

[1] R. E. Alvarez and A. Macovski, "Energy-selective reconstructions in x-ray computerized tomography," *Phys. Med. Biol.* **21**(5), 733–744 (1976).

[2] F. A. Rutherford, B. R. Pullan, and I. Isherwood, "Measurement of effective atomic number and electron density using an EMI scanner," *Neuroradiology* **11**(1), 15–21 (1976).

[3] G. Christ, "Exact treatment of the dual-energy method in CT using polyenergetic x-ray spectra," *Phys. Med. Biol.* **29**(12), 1501–1510 (1984).

[4] W. A. Kalender, W. H. Perman, J. R. Vetter, and E. Klotz, "Evaluation of a prototype dual-energy computed tomographic apparatus, I. Phantom studies," *Med. Phys.* **13**(3), 340–343 (1986).

[5] F. L. Roder, "Principles, history, and status of dual-energy computerized tomographic explosives detection," *J. Test. Eval.* **13**(3), 211–216 (1985).

[6] J. Hsieh, "Dual energy imaging enhancement with fuzzy logic," *Proc. SPIE* **2434**, 290–299 (1995).

[7] Z. Chen, R. Ning, D. Conover, Y. Yu, and X. Lu, "Dual-basis-material decomposition for dual-kVp cone-beam CT breast imaging," *Proc. SPIE* **5745**, 1322–1333 (2005).

[8] D. Walter, X. Wu, Y. Du, E. Tkaczyk, and W. Ross, "Dual kVp material decomposition using flat-panel detectors," *Proc. SPIE* **5368**, 29–39 (2004).

[9] N. Pelc, "Dual energy CT: physics principles," *Med. Phys.* **35**(6), 2861 (2008).

[10] E. T. D. Hoey, D. Gopalan, and N. J. Screaton, "Dual-energy CT pulmonary angiography: a new horizon in the imaging of acute pulmonary thromboembolism," *Am. J. Roentgen.* **192**, 341–342 (2009).

[11] K. Deng, C. Liu, R. Ma, C. Sun, X. Wang, Z. Ma, and X. Sun, "Clinical evaluation of dual-energy bone removal in CT angiography of the head and neck: comparison with conventional bone-subtraction CT angiography," *Clin. Radiol.* **64**(5), 534–541 (2009).

[12] D. T. Boll, N. A. Patil, E. K. Paulson, E. M. Merkle, W. N. Simmons, S. A. Pierre, and G. M. Preminger, "Renal stone assessment with dual-energy multidetector CT and advanced postprocessing techniques: improved characterization of renal stone composition-pilot study," *Radiol.* **250**(3), 813–820 (2009).

[13] C. L. Brown, R. P. Hartman, O. P. Dzyubak, N. Takahashi, A. Kawashima, C. H. McCollough, M. R. Bruesewitz, A. M. Primak, and J. G. Fletcher, "Dual-energy CT iodine overlay technique for characterization of renal masses as cyst or solid: a phantom feasibility study," *Euro. Radiol.* **19**(5), 1289–1295 (2009).

[14] F. Schwarz, B. Ruzsics, U. J. Schoepf, G. Bastarrika, S. A. Chiaramida, J. A. Abro, R. L. Brothers, S. Vogt, B. Schmidt, P. Costello, and P. L. Zwerner, "Dual-energy CT of the heart-principles and protocols," *Euro. J. Radiol.* **68**(3), 423–433 (2008).

[15] C. H. McCollough, A. N. Primak, O. Saba, H. Bruder, K. Stierstorfer, R. Raupach, C. Suess, B. Schmidt, B. M. Ohnesorge, and T. G. Flohr, "Dose performance of a 64-channel dual-source CT scanner," Radiology **243**, 775-784 (2007).

Spectral CT Imaging using a Slow kVp switching Technique and PICCS Image Reconstruction

Timothy P. Szczykutowicz and Guang-Hong Chen

***Abstract*—Dual energy CT using a fast kVp switching technique and the standard filtered backprojection (FBP) image reconstruction method has recently been studied. With conventional FBP methods, high slew rates are required for acceptable image reconstruction with high image quality. However, high slew rates also require hardware changes to enable data acquisition. In this work, we aim at studying the necessary slew rate for dual energy CT imaging provided that the PICCS algorithm is used for image reconstruction. Two important results from this study are as follows: (1) a slew rate better than 7.5 kV/view is sufficient for dual energy imaging; (2) the slow slew rate also allows for spectral CT imaging in which more than two different kVp images are reconstructed.**

***Index Terms*—Slew Rate, kV switching, Dual Energy, triple energy.**

I. Introduction

HOUNSFIELD'S original paper on CT mentioned a method to determine a materials atomic number using a simple image subtraction technique using two different energy spectra[1]. Later, the work of Alvarez and Mocovski confirmed the need for two energies for material discrimination by showing how the attenuation coefficient can be decomposed into two parts over the diagnostic energy range[2]. The two parts represent the dominant interactions of x-rays with matter over the diagnostic energy range in CT imaging, the photoelectric and Compton effects. Using basis functions representing these two interactions, one can create electron density, atomic number, and material specific density images.

In order to realize dual energy CT in a clinical setting, a CT system must be capable of acquiring CT projection data at two or more different energies. Collecting this much data within a single gantry rotation is not a trivial problem. Many solutions have been proposed and are currently being used clinically and investigated experimentally [3], [4], [5], [6], [7]. In this study, we show how a dual energy data acquisition can be realized using a single gantry rotation that collects a complete set of mixed energy projection data using a slow kVp modulation scheme. The slow kVp modulation causes the collected data at the low and high energies to be undersampled. Low and high energy data are required in order to reconstruct low and high kVp images from which, an image space dual energy decomposition can be performed. Conventional reconstruction algorithms cannot provide adequate image quality as streaking artifacts would be present due to the undersampled data acquisition. Using the prior image constrained compressed sensing (PICCS) algorithm, we have demonstrated that low and high kVp images can be reconstructed without streaks and that these images can then be used to perform dual energy decomposition. We have also shown the ability to reconstruct more than two different kVp images. This effectively means there are more measurements than are required for a standard two material basis decomposition. Therefore, a material decomposition into more than two basis components is possible.

T. P. Szczykutowicz is with the Department of Medical Physics , University of Wisconsin Madison, Madison, WI 53705-2275. Phone: +1 608 263–0211, e-mail: szczykutowic@wisc.edu

G-H. Chen is with the Departments of Medical Physics, Radiology and Human Oncology, University of Wisconsin Madison, Madison, WI 53705-2275. Phone: +1 608 263–0089, e-mail: gchen7@wisc.edu

II. Methods and Materials

(a) (b) (c)

Fig. 1. The three phantom objects used. (a) Phantom I: The outside ring is muscle and the inside is water. The circles, clockwise from the top, are blood, fat, muscle, and bone respectively. (b) Phantom II: The outside sheath is muscle. The white ellipses and ring are cortical bone and the large cavities are blood surrounded by water. (c) Phantom III: The outside sheath is muscle. The inside is water and the circular ROI's are Calcium, Iodine and Blood clockwise from 9 o'clock.

A. kVp Modulation Simulation

The kVp slew rate was simulated for a range of values yielding x-ray acquisition patterns in which the kVp varied angularly. The total number of views acquired per rotation was kept constant and the kVp was varied from 80-140 kVp at different slew rates. A slew rate of 60 kVp per view angle (kV/view) corresponds to "fast kVp switching" as the kVp is capable of making the entire transition from low to high or high to low in a single view angle increment. A total of 2,000 views were simulated per rotation. A slew rate of 60 kV/view with 2,000 total views yields 1,000 views at 80 and 140 kVp. The total number of view angles at 80 or 140 kVp depends on the slew rate, total number of acquired angles (T), and the kVp separation according to equation 1.

This model assumes a piecewise linear transition from kVp to kVp, hence the kVp modulation scheme takes on the appearance of a triangle waveform. Figure 2 plots kVp as a function of angle for a complete 360 degree scan for

$$\text{Total \# of low or high kVp projections} = \frac{T \cdot \text{SlewR}}{2 \cdot (\text{kVp}_{high} - \text{kVp}_{low})} \quad (1)$$

different slew rates. One important observation can be made from this plot: The total number of view angles for 80 and 140 kVp decreases as the slew rate decreases. As a result, undersampling artifacts appear when the FBP image reconstruction algorithm is used.

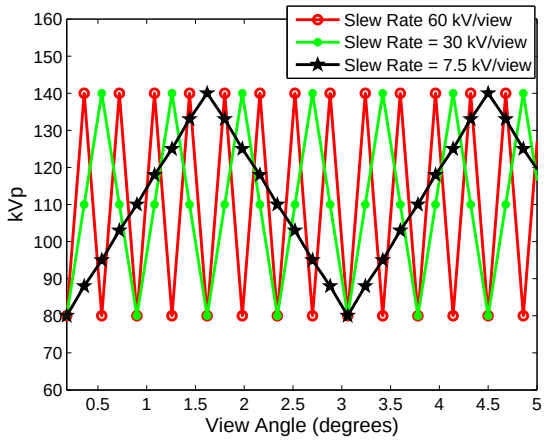


Fig. 2. Plots of the three different kVp modulations. All modulations were simulated for a full 360°s, a limited angular range is depicted here to better visualize the modulation.

B. Filtered Back Projection and Iterative Reconstructions

In order to compare PICCS with other reconstruction methods, reconstructions were performed using the standard filtered back projection method[8] using a Shepp-Logan filtering kernel and a slightly modified version of the simultaneous algebraic reconstruction technique SART[8]. An interleaved projection update strategy was used with a ray based forward projection and a pixel based backprojection in which, the image was updated view by view.

C. Reconstructing undersampled kVp data sets using PICCS

In this paper, an image (the prior image) is reconstructed using all of the data collected over a complete rotation and used to constrain the reconstructions of the low and high kVp data to produce accurate and streak free low and high kVp images (target images). Figure 3 depicts the data usage in this application of PICCS. The numerical implementation of the PICCS algorithm used in this study is the same as that used in references [9], [10].

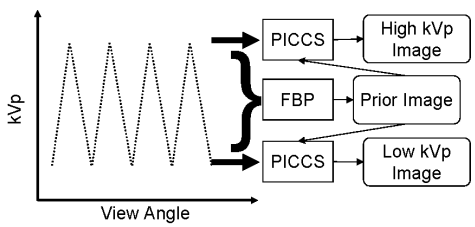


Fig. 3. The data flow and usage in this application of the PICCS algorithm.

D. Triple energy image reconstruction in the PICCS framework

The previously described framework uses all of the view angles to reconstruct the prior image and then uses the lowest and highest energy projections to obtain streak free low and high energy images. As the kVp is modulated from the lowest energy to the highest energy, a significant number of intermediate view angles are acquired. Taking advantage of this fact, one can utilize these intermediate energy projections and reconstruct an intermediate energy image. In this study, an example of this procedure for the following energy triplet was performed: 80, 110, and a 140 kVp. Projection data was acquired using a slew rate of 7.5 kV/view, which provided 125 views for the 80 and 140 kVp images and 250 view angles for the 110 kVp image. The prior image was a fully sampled image acquired during the same gantry rotation as the 80, 110, and 140 kVp projections.

E. Material Density Image Decomposition

An image based material density scheme was used to process the resulting low and high kVp images after the PICCS reconstruction step. The basis materials chosen were water and cortical bone for the two material basis decomposition and blood, Calcium and Iodine for the three material decomposition. For the two material basis decomposition, the average pixel values were determined for water and bone using fully sampled filtered back projection images at 80 and 140 kVp. These values were used to create basis material vectors onto which a pair of low and high kVp images could be decomposed onto. The same procedure was performed for the three material basis images using 80, 110 and 140 kVp images.

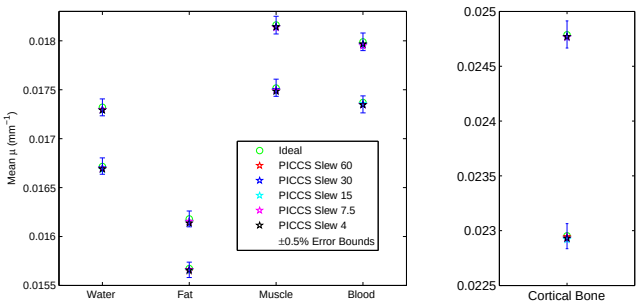
III. Results

The results section consists of three parts. The first part includes the results from Phantom I shown in Figure 1.a. This phantom was chosen such that a quantitative evaluation of the PICCS results could be made. Phantom II (Figure 1.b) was chosen to test the reconstruction fidelity of the PICCS algorithm as it provided areas of small detail and highly attenuating bony structures. Phantom III (Figure 1.c) was chosen for the evaluation of a three material basis decomposition and regions of blood, Calcium and Iodine were present in order to evaluate if one could distinguish between the Calcium and Iodine regions after a three basis material decomposition.

A. Phantom I

The results for Phantom I are shown in Figures 4, 5, and 6. Quantitative results will not be shown for the rest of the phantoms as a detailed quantitative comparison is shown for Phantom I. The relative root mean square error (RRMSE) plot shows how the PICCS results show a relatively constant error while the FBP and SART images diverge at low slew rate as seen in Figure 5. A quantitative analysis of the reconstruction accuracy of the PICCS

images given in Figures 4 and 6 show how accurate the PICCS images are in both the attenuation image and material basis image spaces. The PICCS images are within ± 0.5% of the fully sampled FBP result for the attenuation image results and within ± 10% of the standard deviation in the background noise in the material basis images. As these errors are small, the plots in figures 4 and 6 are not meant to show the differences between different slew rate results for PICCS, but to show the accuracy of all the PICCS results. In effect, the PICCS mean values all lie very close to the fully sampled FBP mean values.



(a) (b)

Fig. 4. Mean ROI values from the PICCS Images for Phantom I (the actual images are not shown in this abstract). Mean ROI values from a fully sampled FBP image are also included and the ± 0.5% bounds on the FBP value. The scale is not sufficient to distinguish between individual PICCS values, but one can appreciate how all of the PICCS values are within ± 0.5% of the fully sampled FBP result. For each material, the lower cluster of points are for 140 kVp, and the higher for 80 kVp.

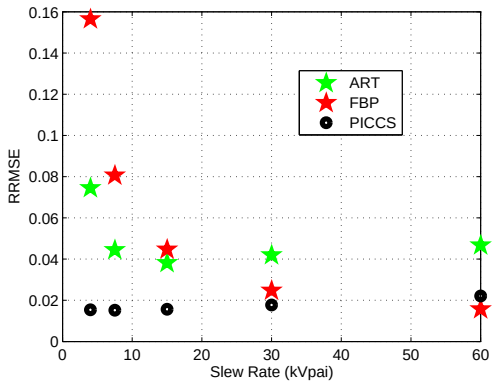


Fig. 5. Relative Root Mean Square Error between FBP, SART, and the PICCS algorithm for Phantom I.

B. Phantom II

The same reconstruction procedure was carried out as in the last section for Phantom II. The results are shown in Figures 7 and 8. In this case, streaking artifacts become evident in the FBP and SART images due to the small high contrast bony structures. For this phantom, the onset of streaking artifacts in the attenuation images for FBP and SART can be seen at a slew rate of 15 kV/view.

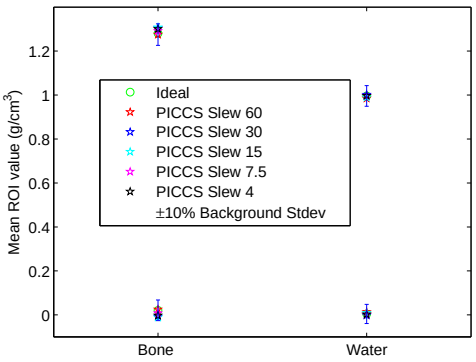


Fig. 6. Mean ROI values taken from the bone and water material density images. The clusters centered around zero for bone and water represent the mean ROI value of the opposite basis for bone and water (the bone ROI on the water basis image is zero and the water ROI on the bone image is zero). For comparison, fully sampled FBP basis material decompositions were compared and the PICCS results are all within ±10% of the noise background to the FBP results.

The PICCS material basis results show streak free reconstruction for all slew rates studied but at a slew rate of 4 kV/view a degradation of the the smaller structures is clearly visible. At this slew rate, the FBP and SART images are severely affected by streaking artifacts. The stopping point or limiting slew rate for the PICCS reconstruction appears to be somewhere around 7.5 kV/view as at this slew rate the smaller structures are still well visualized especially around the spine (compare the PICCS slew rate 7.5 and 4 kV/view images on Figure 8).

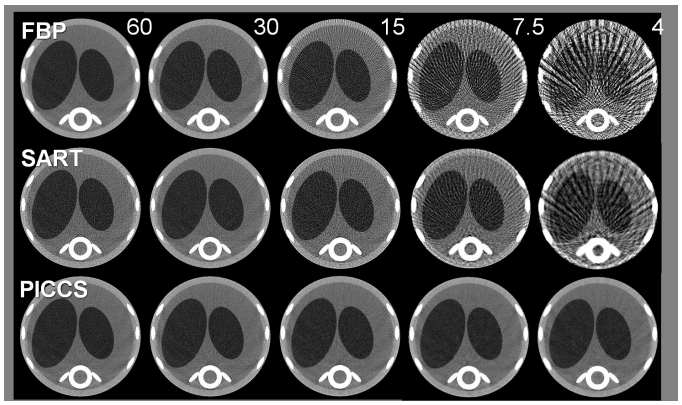


Fig. 7. Results of Phantom II. The numbers refer to the kV/view slew rate used to acquire the data for each image.

C. Phantom III

As explained in the materials and methods section, the phantom for this section was specially chosen such that the differences in the Calcium and Iodine regions in normal attenuation images would not be discernible. This represents the clinical situation in which the ability to differentiate calcifications and an Iodine contrast agent is desirable. The slew rate of 7.5 kV/view was chosen providing 125 projections for the 80 and 40 kVp images and 250 projections for the 110 kVp image. Figure 9 is a scatter plot

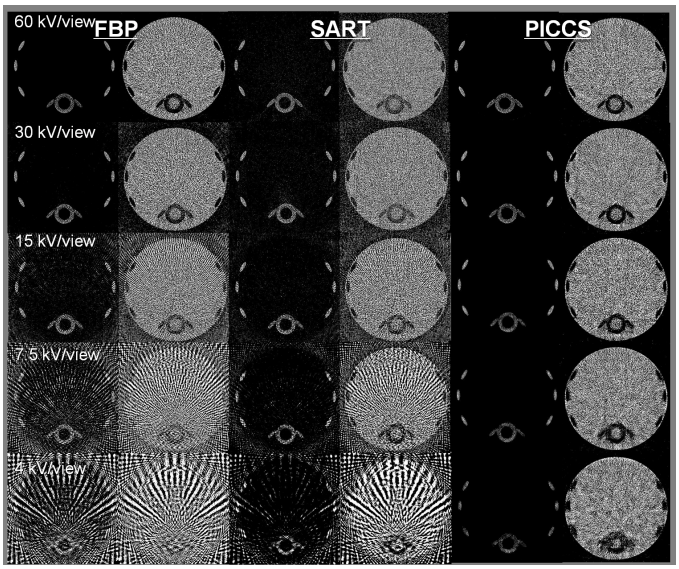


Fig. 8. Results of complex phantom two basis material decomposition. Note that at a slew rate of 4 kV/view the PICCS result seems to loose the ability to accurately reconstruct the small bony structures, while streaking artifacts are present in the FBP and SART images at a slew rate of 15 kV/view.

of the values within a Calcium, blood and Iodine region of the reconstructed images for 80, 110, and 140 kVp energies. A clear separation in energy is visible between the energies in the attenuation values. Using these three images, a three basis material decomposition was performed and the results are given in Figure 10. The Calcium and Iodine images were filtered with a mean filter of radius = 2 pixels to reduce the image noise. No such filtering was done to the blood image.

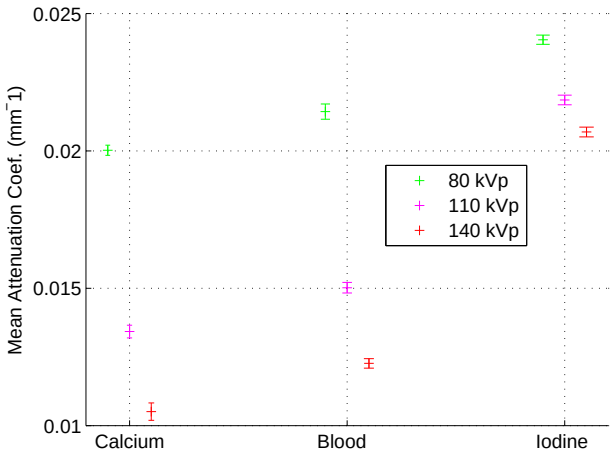
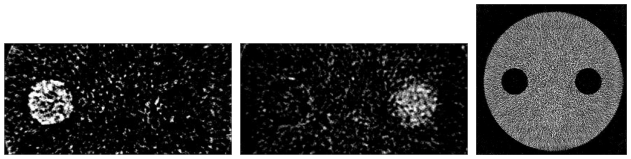


Fig. 9. Mean ROI values for Phantom III for the blood, Calcium and Iodine regions. The error bars are the standard deviation of the pixel values inside the ROI regions.

IV. Conclusions

It was demonstrated that the PICCS algorithm allows for dual and triple energy image reconstruction from a single gantry rotation using a kVp modulated scanning protocol. For the dual energy application, the PICCS algorithm



(a) (b) (c)

Fig. 10. (a) Calcium Basis Image. (b) Iodine basis image. (c) Blood basis image. The Calcium insert is at the 9 o'clock position and the Iodine insert is at the 3 o'clock position. Excellent differentiation is observed between these two inserts.

was compared to standard FBP and SART and was shown to outperform them in terms of the presence of streaking artifacts and the image quality of the basis material images. A three material basis decomposition was also shown to be made possible by reconstructing an additional energy within the lower and upper kVp values. This allowed for the differentiation between a Calcium and an Iodine region on a phantom where their respective attenuation values were roughly the same.

In summary, the central results of this study are as follows:

- Dual Energy Imaging was realized using a single gantry rotation with a slow kVp slew rate enabled by the PICCS algorithm.
- The possibility of reconstructing images at more than two energies from a single gantry rotation was demonstrated, and a three material decomposition was performed.

References

[1] G. N. Hounsfield, "Computerized transverse axial scanning (tomography): Part i. description of system," *British Journal of Radiology* **46**, pp. 1016–1022, 1973.

[2] R. Alvarez and A. Macovski, "Energy-selective reconstruction in x-ray comptued tomography," *Phys. Med. Biol.* **21**, pp. 733–744, 1976.

[3] F. Kelcz, P. M. Joseph, and S. Hilal, "Noise considerations in dual energy ct scanning," *Med. Phys.* **6(5)**, 1979.

[4] T. Johnson, B. Kraub, M. Sedlmair, M. Grasruck, H. Bruder, D. Morhard, C. Fink, S. Weckbach, M. Lenhard, B. Schmidt, T. Flohr, M.F.reiser, and C. Becker, "Material differentiation by dual energy ct: initial experience," *Eur. Radiol* **17**, 2007.

[5] D. Xu, D. Langan, X. Wu, J. Pack, T.M., Benson, J. Tkaczky, and A. Schmitz, "Dual energy ct via fast kvp switching spectrum estimation," *Proc. of SPIE* **7258**, p. 72583T, 2009.

[6] K. Taguchi, M. Zhang, E. Frey, and J. Xu, "Image-domain material decomposition using photon-counting ct," *Proc. of SPIE* **6510**, p. 651008, 2007.

[7] Y. Zou and M. D. Silver, "Analysis of fast kv–switching in dual energy ct using a pre–reconstruction decomposition technique," *Proc. of SPIE* **6913**, p. 691313, 2008.

[8] A. C. Kak and M. Slaney, "Principles of computerized tomographic imaging," *IEEE Press New York* , 1988.

[9] G. H. Chen, J. Tang, and S. Leng, "Prior image constrained compressed sensing (piccs): A method to accurately reconstruct dynamic ct images from highly undersampled projection data sets," *Med. Phys.* **35**, pp. 660–663, 2008.

[10] G. H. Chen, J. Tang, and J. Hsieh, "Temporal resolution improvement using piccs in mdct cardiac imaging," *Med. Phys.* **36**, pp. 2130–2135, 2009.

Monte Carlo Simulation Approach to Estimating Patient Radiation Dose from MDCT Exams

Michael McNitt-Gray, John DeMarco, Chris Cagnon, Adam Turner, Di Zhang
David Geffen School of Medicine at UCLA
Los Angeles, CA 90024 USA

Abstract- : X-ray CT now accounts for 50% of the medical radiation exposure to the population in the United States. Methods to estimate radiation dose and especially dose to radiosensitive organs are becoming even more important. Monte Carlo simulation methods have been used over the past few decades to estimate radiation dose to patients undergoing CT exams. The purpose of this work is to describe recent developments in Monte Carlo simulation methods. These methods include: (a) advances in the representations for MDCT source models, (b) patient models that represent anatomy more realistically as well as different size, age and gender models and (c) advances in modeling MDCT capabilities such as tube current modulation, dynamic collimation, etc. All of these have advanced our ability to accurately estimate radiation dose to specific organs and assess the efficacy of various radiation dose reduction methods on actual patient radiation dose.

I. INTRODUCTION

Recent studies report that CT contributes 50% of the medical radiation dose to the United States population [1]. It has been suggested that the most appropriate quantity for assessing the risk due to diagnostic imaging procedures is the radiation dose to individual organs [2]. However, the standard dosimetry metric for CT is the Computed Tomography Dose Index (CTDI) which is a calculated based on measurements made in cylindrical polymethyl methacrylate (PMMA) phantoms [3]. CTDI values are widely used for quality assurance and accreditation purposes, however, they are not intended to represent dose to any particular patient or, more importantly, to any particular organ.

To estimate organ doses, Monte Carlo simulation methods have been utilized in many previous efforts. This approach was used in dosimetry studies performed by both the NRPB (Chilton, UK) [4] and the GSF (Oberschleissheim, Germany)[5], the results of which have been incorporated into software packages such as the ImPACT CT Patient Dosimetry Calculator (ImPACT, London, England)[6] and CT-Expo (Medizinische Hochschule, Hannover, Germany)[7]. These original studies were based on single detector row, non-helical scanners; however methods have been developed to extend the results to current, commercially available helical CT scanners. For example, a method to match new scanners to those originally simulated based on physical measurements (such as CTDI)[6] was developed.

These original efforts also used geometric descriptions of patients and their radiosensitive organs, using either the MIRD Phantom [8] or modifications to make both male and female versions [9].

The purpose of this work is to describe recent developments in Monte Carlo simulation methods. These methods include: (a) advances in the representations for MDCT source models, (b) patient models that represent anatomy more realistically as well as different size, age and gender models and (c) advances in modeling MDCT capabilities such as tube current modulation, dynamic collimation, etc. All of these have advanced our ability to accurately estimate radiation dose to specific organs and assess the efficacy of various radiation dose reduction methods on actual patient radiation dose.

II. DESCRIPTION OF MONTE CARLO MODELS

A. Monte Carlo Simulation Software

There are many Monte Carlo transport simulation packages available. Most of the software packages used for CT dosimetry are either general-purpose Monte Carlo codes (such as MCNP or EGS) that can be used for neutron, photon, electron or coupled neutron/photon/electron transport, or they are modifications of these transport codes. The simulation of photon transport includes incoherent and coherent scattering, photoelectric absorption with the creation of K and L shell fluorescent photons or Auger electrons and bremsstrahlung. These packages have detailed descriptions of photon cross sections and electron binding effects associated with Compton and Rayleigh scatter The photon transport model may create electrons but may also assumes that they travel in the direction of the primary photon and that the electron energy is deposited at the photon interaction site, creating a condition of charged particle equilibrium (CPE). Under conditions of CPE, the assumption that collision kerma is equal to absorbed dose is valid and a collision-kerma tally can be used for the absorbed dose calculations and can be based upon an energy fluence tally.. Energy fluence is converted to dose rate using the mass energy-absorption coefficients obtained by Seltzer and Hubbell (Hubbell and Seltzer 1995).

B. Modeling the MDCT Scanner

There are several aspects of the MDCT scanner that need to be modeled in some detail in order to obtain an accurate model of the CT source. This includes: (a) scanner geometry, (b) x-ray spectra, (c) filtration, (d) collimation, (e) source movement and (f) any radiation dose reduction method such as tube current modulation, etc. Scanner geometry includes

descriptions of source to isocenter distance and fan angle. X-ray spectra includes description of the photon fluence as a function of energy. These are sometimes obtained from the manufacturer or from some estimation software such as TASMIP [11]. The spectrum is typically represented as a cumulative density function because part of the simulation will be to select the starting energy of a photon using a random number generator.

Filtration includes representation of physical filtration, both the internal and external to the x-ray source. This includes any added filtration and the bowtie filtration used by each scanner. This is often proprietary information and is not readily available from manufacturers; however recent advances include the development of methods to estimate the effects of the bowtie filtration [12,13]. Collimation effects represent the pre-patient collimation provided by the scanner. This is represented by the actual measured radiation dose profile in air at isocenter (rather than nominal beam collimation). For MDCT and cone beam CT, this can range up to 160 mm nominal beam width [14].

MDCT scanners are operated in sequential axial, helical and cine (axial scan with no table movement) modes; therefore Monte Carlo simulations need to be able to accommodate each of these source motion schemes, including helical scans with pitch ranging from 0.2 up to 3.2. Finally, recent technical developments specially targeted towards radiation dose reduction such as tube current modulation [15] and dynamic collimation [16].

C. Object Models

This section describes several different objects that are typically used in Monte Carlo simulations. The first set of objects are the cylindrical CTDI phantoms that are used in conventional dosimetry. These are often the basis for benchmarking Monte Carlo models. The next level of model complexity is the. MIRD V anthropomorphic model, which , as described above, describes a standard patient model in terms of geometric descriptions of 27 radiosensitive organs. This model typically uses three basic materials; tissue, lung and bone.

Finally, voxelized models of patient anatomy are typically based on actual patient anatomy for a series of different patients. For example, the GSF family [17] contains eight models ranging from an 8 week old female ("Baby") to an adult male ("Visible Human"). For each model, radiosensitive organs are identified in voxelized models. There are many other examples including the fetal models described in [18].

D. Simulating the Scan

The detailed models described above allow researchers to estimate radiation dose to a specific patient from a specific MDCT scanner using a specific scan protocol. The last step requires the user/researcher to describe the actual scan being simulated. This includes selection of parameters such as:

1. *Make and model of MDCT scanner*
2. *Patient size, gender*
3. *Anatomic region (e.g. head, neck, thorax, etc.), which in turn determines:*
    a. *beam on and beam off locations with respect to the patient model*
    b. *any anatomic region-specific filtration (e.g. head bowtie filter, body bowtie filter, etc.)*
4. *Beam energy (kVp)*
5. *Collimation(nominal and actual)*
6. *Table movement (if any) including helical pitch*
7. *Tube Current and rotation time*
8. *Any other dose reduction method (tube current modulation, dynamic collimation, etc.)*

When detailed models of the scanner and patient are provided and detailed descriptions of the desired scans are provided, then estimates of doses to radiosensitive organs can be provided. Examples of these kinds of investigations include: (a) dose to specific organs for thoracic scans as a function of patient size [19], (b) dose to fetus from abdominal scans as a function of patient size [18] and (c) dose to breast and lung in female patients undergoing thoracic scans with tube current modulation [15].

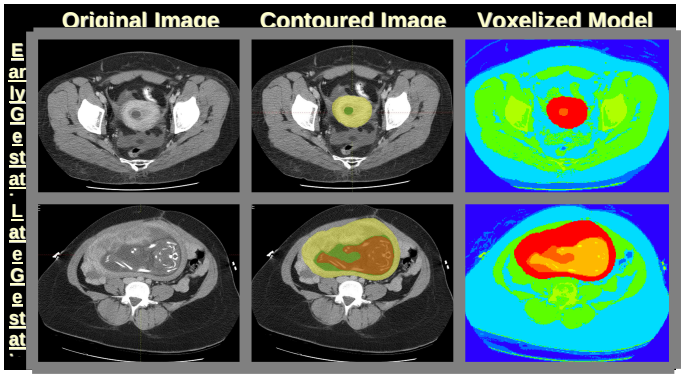


Figure 1 – Examples of voxelized models for fetal anatomy as described in [18]

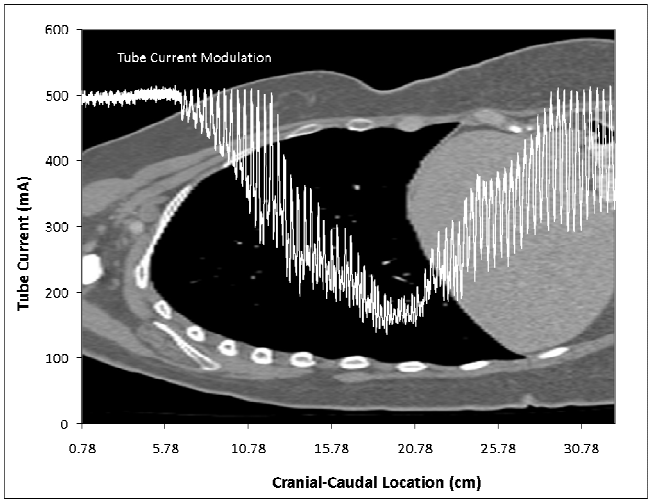


Figure 2 – Examples of tube current modulation results for a patient described in [15]

III. Discussion

The purpose of this work was to describe recent developments in the use of Monte Carlo simulations as a method to accurately estimate individual organ doses from MDCT exams. These methods can be applied for any scanner, patient, and protocol combination.

This has allowed important investigations into the estimation of dose to fetus for pregnant patients undergoing abdominal CT scans, the effects of patient size of dose to specific radiosensitive organs due to thoracic scans as well as investigations into the impact on dose reduction to radiosensitive organs from tube current modulation.

This approach and these investigations allow us to increase our understanding of the impact of MDCT scanning and radiation dose reduction technologies on dose to radiosensitive organs. This allows us to begin to make meaningful decisions about the use of CT, the efficacy of different radiation dose reduction technologies and to lay the foundation for accurate estimations of risk in the future.

Ackowledgments

This work was supported by NIBIB Grant R01-EB004898.

References

[1] F.A. Mettler Jr., B.R. Thomadsen, M. Bhargavan, D.B. Gilley, J.E. Gray, J.A. Lipoti, J. McCrohan, T.T. Yoshizumi, and M. Mahesh, "Medical radiation exposure in the U.S. in 2006: preliminary results," Health Phys. 95(5), 502-507 (2008).

[2] National Research Council. Health Risks from Exposure to Low Levels of Ionizing Radiation: BEIR VII Phase 2. The National Academies Press, Washington, DC. (2005).

[3] American Association of Physicists in Medicine. The measurement, reporting and management of radiation dose in CT. New York: AAPM; Report No 96; 2008.

[4] D.G. Jones and P.C. Shrimpton, "Survey of CT practice in UK. Part 3: Normalized organ doses calculated using Monte Carlo techniques," National Radiological Protection Board, NRPB-R250, 1991

[5] M. Zankl, W. Panzer, and G. Drexler, "The calculation of dose from external photon exposures using reference human phantoms and Monte Carlo methods. VI. Organ doses from computed tomographic examinations. GSF report 30/91. Oberschleissheim, Germany: GSF-Forschungszentrum, (1991).

[6] Imaging Performance Assessment of CT scanners group. "ImPACT CT patient dosimetry calculator," 2006, available at http://www.impactscan.org.

[7] G. Stamm and H.D. Nagel, "CT-Expo – a novel program for dose evaluation in CT [in German]," Rofo. 174, 1570-1576 (2002).

[8] Snyder W S, Fisher H L, Ford M R and Warnet G C 1969 Estimates of absorbed fractions for monoenergetic photon sources uniformly distributed in various organs of a heterogeneous phantom J. Nucl. Med. 10 (Suppl. 3)

[9] Kramer R, Zankl M, Williams G and Drexler G 1982 The calculation of dose from external photon exposures using reference human phantoms and Monte Carlo methods. Part I. The male (Adam) and female (Eva) adult mathematical phantoms GSF-Bericht S-885

[10] Hubbell JHand Seltzer SM1995 Tables of x-raymass attenuation coefficients andmass energy-absorption coefficients 1 keV to 20 MeV for elements Z = 1 to 92 and 48 additional substances of dosimetric interest NISTIR 5632

[11] Boone J M and Seibert J An accurate method for computer-generating tungsten anode x-ray spectra from 30 to 140 kV Med. Phys. 24 1661–70

[12] A.C. Turner, D. Zhang, H.J. Kim, J.J. DeMarco, C.H. Cagnon, E. Angel, D.D. Cody, D.M. Stevens, A.N. Primak, C.H. McCollough, and M.F. McNitt-Gray. "A method to generate equivalent energy spectra and filtration models based on measurement for multidetector CT Monte Carlo dosimetry simulations," Med. Phys. 36(6), 2154-2164 (2009).

[13] JM Boone, Method for evaluating bow tie filter angle-dependent attenuation in CT: Theory and simulation results. Medical Physics -- January 2010, 37(1): pp. 40-48

[14] Geleijns J, Salvadó Artells M, de Bruin PW, Matter R, Muramatsu Y, McNitt-Gray MF Computed tomography dose assessment for a 160 mm wide, 320 detector row, cone beam CT scanner. Phys Med Biol. 2009 May 21;54(10):3141-59.

[15] Angel E, Yaghmai N, Jude CM, DeMarco JJ, Cagnon CH, Goldin JG, McCollough CH, Primak AN, Cody DD, Stevens DM, McNitt-Gray MF Dose to radiosensitive organs during routine chest CT: effects of tube current modulation. AJR Am J Roentgenol. 2009 Nov;193(5):1340-5

[16] Deak PD, Langner O, Lell M, Kalender WA.Effects of adaptive section collimation on patient radiation dose in multisection spiral CT. Radiology. 2009 Jul;252(1):140-7.

[17] N. Petoussi-Henss, M. Zankl, U. Fill, and D. Regulla, "The GSF family of voxel phantoms," Phys. Med. Biol. 47, 89–106 (2002).

[18] Angel E, Wellnitz CV, Goodsitt MM, Yaghmai N, DeMarco JJ, Cagnon CH, Sayre JW, Cody DD, Stevens DM, Primak AN, McCollough CH, McNitt-Gray MF. Radiation dose to the fetus for pregnant patients undergoing multidetector CT imaging: Monte Carlo simulations estimating fetal dose for a range of gestational age and patient size. Radiology. 2008 Oct;249(1):220-7.

[19] DeMarco JJ, Cagnon CH, Cody DD, Stevens DM, McCollough CH, Zankl M, Angel E, McNitt-Gray MF. Estimating radiation doses from multidetector CT using Monte Carlo simulations: effects of different size voxelized patient models on magnitudes of organ and effective dose.Phys Med Biol. 2007 May 7;52(9):2583-97.

Dose from CT Brain Perfusion examinations: A Monte Carlo study to look into deterministic effects

Di Zhang, Adam Turner, Chris Cagnon, John DeMarco, Michael McNitt-Gray
David Geffen School of Medicine at UCLA
Los Angeles, CA 90024 USA

Abstract- In brain perfusion studies, the patients are scanned repeatedly at one location of the head over a short period of time to monitor contrast to wash in and wash out. This can result in very high radiation dose and it may cause deterministic effects, such as skin burn, or even cataract, if the eye lenses are covered in the x-ray beam. Skin dose and eye lens dose were estimated using Monte Carlo method with detailed patient model and accurate CT source models to simulate brain perfusion single axial scans using the widest collimation of 24x1.2mm and all the available kVps (including 80kVp, 100kVp, 120kVp, and 140 kVp) for a Siemens Sensation 64 Multi-detector CT (MDCT) scanner at various scan locations. Meanwhile, CTDIvol was obtained to investigate how well it predicts eye lens dose and skin dose. The results showed that depending on the scan protocol, there is a high possibility that deterministic effects may happen when a high enough mAs/rotation and/or enough rotations are used. For example, for a 300 mAs/rotation scan at 120kVp, 75 rotations would result in eye lens dose of 2Gy, and 62 rotations would result in a maximum skin dose of 2Gy. CTDIvol overestimates eye lens dose by 20% to 30% and it overestimates skin dose by 40% to 60%. This study provides detailed information about radiation dose to eye lens and skin from CT brain perfusion examinations and it could help to improve the design of scan protocols.

I. INTRODUCTION

Radiation dose to patients continues to be a important concern to medical physicists and to the broader medical community as well. Radiation dose from CT exams has been identified as the largsest source of medical radiation exposure [1]. There has been greatly increased attention on the radiation dose from CT in the past few years.

The radiation dose received by patients from CT examination is usually relatively low (on the order of mGy) comparing to that from radiotherapy because of the lower energy of the x-ray beam and lower current-time-product (mAs). Therefore, stochastic effect has been regarded as the primary concern in terms of the biological consequences from radiation dose introduced by CT examinations. However, the nature of some scanning protocols requires exposing an area for a considerable amount of time. For example, in brain perfusion studies, the patients are scanned repeatedly at one location of the head to monitor the contrast to wash in and wash out. This can result in very high radiation dose and it may cause deterministic effects, such as skin burn, or even cataract, if the eye lenses are covered in the x-ray beam.

Currently CTDI is the most widely used dose metric for the estimation of radiation dose from CT. But it represents the radiation dose to a homogeneous cylindrical phantom and does not necessarily predict the radiation dose received by patients. Monte Carlo based methods using realistic voxelized patient models is a much more accurate method for the estimation of radiation dose to individual organs. The purpose of this work is to accurately estimate the radiation dose to eye lenses and local skin tissues from CT brain perfusion studies for a patient, and investigate how well can CTDIvol predict these doses.

II. Material and Methods

A. Monte Carlo Method Based CT Scanner Model

A Monte Carlo particle transport package, MCNP eXtended v2.6 (MCNPX), which was developed at Los Alamos National Laboratory, was utilized to perform the simulations to estimate radiation dose deposition from CT. The simulations were operated in photon transport mode with a low-energy cutoff of 1keV. Charged-particle equilibrium (CPE) was assumed so that all the secondary electrons deposit their energy at the photon interaction sites.

In this study, a 64-slice CT scanner system (Sensation 64, Siemens Medical Solutions, Forcheim, Germany) was modeled for all simulations using Monte Carlo-based methods. The models were based on previous work [2] and take into account the x-ray source spectra, beam filtration (including bowtie filter) and scanner geometry (focal spot to isocenter distance, fan angle, etc.) as provided by the manufacturer. For this scanner, the widest available beam collimation is 24x1.2mm (nominal beam width of 28.8mm). The actual radiation profile was measured using Optically Stimulated Luminescences (OSLs, CT Dosimeter, Landauer, Inc. Glenwood, Illinois) that were exposed in air at isocenter during a single axial scan using the 24x1.2mm nominal collimation. The OSL dosimeter was then sent to Landauer for reading. From the normalized radiation dose profile that resulted (a table of relative dose values as a function of z-axis location), the full width at half maximum (FWHM) of the dose profile was calculated to be 32.2 mm. This value was used as the measured beam width in the CT scanner model.

The source file of MCNPX was modified to model a Siemens Sensation 64 scanner such that axial scans with any clinical relevant kVp, mAs and collimation settings could be simulated. The pathway of the CT source in x, y, z was

explicitly defined for any scan given scan location and nominal collimation. The location of each emitting photon was then sampled from the pathway specified. The photon spectra for each selected kVp and descriptions of any beam filtration, including the bowtie filter, as well as the geometry were provided by the manufacturer. As in previous work [2], this model was validated and benchmarked using comparisons based on standard dosimetry (CTDI) measurements and corresponding simulations, which agreed to within 5%.

B. Patient Model

The GSF (now: Helmholtz Zentrum München) phantoms are a series of voxelized patient models with segmented individual organs [3]. Irene model was chosen from the GSF series of models to represent an average sized adult female. She has the age of 32 years, the weight of 51kg, and the height of 163cm. In order to incorporate each phantom into the Monte Carlo code for particle transportation calculation, the element composition and mass density of each organ is required. These values were derived from the ICRU 44 organ composition tables [4]. Of specific interest for this project was that both skin and lens of the eye were explicitly represented in this patient model and so radiation dose could be tallied in these voxels

C. Dose estimation for Eye Lenses

Under charged particle equilibrium condition the absorbed dose to each voxel is equivalent to the collision kerma in this voxel, which was calculated by multiplying the mass energy-absorption coefficients with the energy fluence evaluated by MCNPX based on track-length estimation. Then the mean dose to eye lenses was estimated by averaging the dose to each voxel across all the voxels belonging to the eye lenses. The dose results were converted from MCNPX raw output (MeV/g/particle) to absolute dose normalized to tube current (in mGy/mAs) using normalization factors calculated from scan measurements in air and corresponding simulations in air, described in a previous publication [5].

D. Maximum skin dose estimation

By defining the tally points at various locations, the radiation dose can be assessed anywhere in the model using the Monte Carlo method. In order to get the maximum skin dose, mesh tally feature was used to get 3D dose distribution in the patient model. Mesh tallies are composed of a 3D array of voxels in a high-resolution Cartesian-coordinate mesh structure. A MATLAB subroutine was created to map the matrix which represent the original voxelized patient model with the matrix of the 3D dose distribution to localize skin tissues. The maximum skin dose value in the 3D dose distribution matrix was identified as maximum dose of those voxels identified as belonging to the skin. The dose results were first divided by the density of the skin to convert the unit from MeV/cm^3/particle to MeV/g/particle, then it was multiplied by the normalization factors to get absolute dose as described in II.C.

E. Simulation Experiments

For the experiments in this study, simulated brain perfusion single axial scans were performed using the widest collimation of 24x1.2mm and all the available kVps, including 80kVp, 100kVp, 120kVp, and 140 kVp. It should be noted that although 140kVp is not usually used for brain perfusion scans in clinical protocols, the results could still serve as a reference to indicate the resulting radiation dose if 140kVp were used.

To investigate the effect of scan location on the dose to eye lenses, simulations were performed separately at 4 different locations as illustrated in Figure 1. The center of location 1, 2 and 3 are 3.5 cm apart from one to the next, and they are all outside the region of eye lenses. The center of location 4 is 2.5 cm apart from that of location 3, and the x-ray primary beam completely covers the eye lenses at this location.

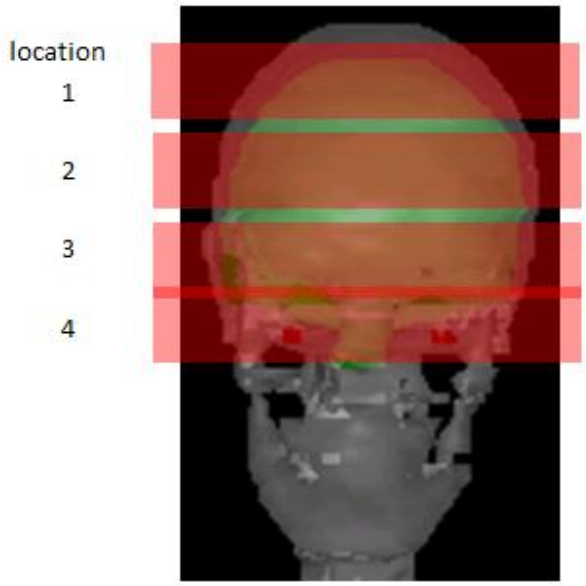


Fig. 1. Illustration of all 4 scan locations for brain perfusion single axial scans. The nominal beam width is 2.88cm. The center of location 1, 2 and 3 are 3.5cm from one to the next. Location 4 is partially overlapped with location 3 to completely cover the eye lenses.

After the eye lens dose and maximum skin dose were estimated on a per mAs basis for all the combinations of locations and kVps, total mAs were calculated under which the total dose would reach 2Gy using (1). Deterministic effects may occur at this level of radiation dose.

$$\textit{Total mAs} = 2000(\textit{mGy}) / \textit{eye lens dose (mGy/mAs)} \quad (1)$$

Furthermore, to help the development and design of scan protocols, numbers of rotations to reach 2Gy were calculated for a typical scan protocol of 170 mAs/rotation, using (2).

$$\textit{Number of rotations} = \textit{Total mAs} / (170\ \textit{mAs/rotation}) \quad (2)$$

Finally, to investigate how well CTDIvol as reported on the scanner console can predict dose to eye lenses or skin, CTDIvol (mGy/mAs) were recorded for different kVps and multiplied by the total mAs calculated above that would result in dose of 2Gy to eye lenses or skin. The dose predicted by

CTDIvol was then compared to 2Gy to see if it overestimates or underestimates the real dose estimated by Monte Carlo simulations.

III. RESULTS

A. Dose to eye lenses

The simulated dose to eye lenses at all four locations and all four kVps is shown in Table I. The total mAs that would result in a dose to eye lens of 2 Gy, as calculated by (1), is shown in Table II. The number of rotations for a 170 mAs/mAs scan protocol that would result in 2 Gy of eye lens dose, as calculated by (2), is shown in Table III. CTDIvol calculated from the scanner on a per mAs basis for all four kVps are shown in Table IV, as well as the dose predicted by CTDIvol at location 4 (because the eye lenses are only covered at location 4). This is the multiplication of CTDIvol and total mAs as described in II.E.

TABLE I. Simulated dose to **eye lenses** at all four scan locations and all four kVps on a mGy per mAs basis.

Dose to eye lenses (mGy/mAs)				
	Location1	Location2	Location3	Location4
80kVp	0.0002	0.0007	0.0025	0.0227
100kVp	0.0006	0.0017	0.0058	0.0499
120kVp	0.0011	0.0033	0.0106	0.0893
140kVp	0.0018	0.0052	0.0167	0.1421

TABLE II. Total mAs that would result in the dose to **eye lenses** reaching 2Gy for all scan locations and kVps.

Total mAs to reach eye lens dose of 2Gy				
	Location1	Location2	Location3	Location4
80kVp	9694400	3066500	811600	87900
100kVp	3482500	1187700	345700	40100
120kVp	1791800	615200	189100	22400
140kVp	1097800	382900	120000	14100

TABLE III. Total number of rotations that would result in **2Gy eye lens dose** for all scan locations and kVps for a 170 mAs/rotation scan protocol.

Total number of rotations to reach eye lens dose of 2Gy				
	Location1	Location2	Location3	Location4
80kVp	57026	18038	4774	517
100kVp	20485	6987	2033	236
120kVp	10540	3619	1113	132
140kVp	6458	2252	706	83

TABLE IV. CTDIvol collected from the scanner on a per mAs basis for all four kVps at location 4, as well as the eye lens dose predicted by CTDIvol

	CTDIvol (mGy/mAs)	Predicted dose (Gy)
80kVp	0.0360	3.17
100kVp	0.0750	3.01
120kVp	0.1267	2.84
140kVp	0.1920	2.70

B. Maximum skin dose

The same set of tables for the maximum skin dose at each scan locations for all four kVps are shown from Table V to Table VIII. This includes the simulated maximum skin dose on a mGy/mAs basis (Table V), total mAs to reach maximum skin dose of 2Gy (Table VI), total number of rotations to reach maximum skin dose of 2Gy for a 170 mAs/rotation scan protocol (Table VII), as well as maximum skin dose predicted by CTDIvol (Table VIII).

TABLE V. Simulated **maximum skin dose** at all four scan locations and all four kVps on a mGy per mAs basis.

Maximum skin dose (mGy/mAs)				
	Location1	Location2	Location3	Location4
80kVp	0.029	0.028	0.028	0.029
100kVp	0.061	0.058	0.058	0.060
120kVp	0. 107	0.101	0.101	0.105
140kVp	0.163	0.154	0.155	0.161

TABLE VI. Total mAs that would result in the **maximum skin dose** reaching 2Gy for all scan locations and kVps.

Total mAs to reach maximum skin dose of 2Gy				
	Location1	Location2	Location3	Location4
80kVp	67900	72000	72300	69000
100kVp	32600	34500	34600	33100
120kVp	18600	19800	19800	19000
140kVp	12300	13000	12900	12500

TABLE VII. Total number of rotations that would result in **2Gy maximum skin dose** for all scan locations and kVps for a 170 mAs/rotation scan protocol.

Total number of rotations to reach skin dose of 2Gy				
	Location1	Location2	Location3	Location4
80kVp	399	423	425	406
100kVp	192	203	204	195
120kVp	110	116	117	112
140kVp	72	76	76	73

TABLE VIII. The maximum skin dose predicted by CTDIvol at all four scan locations and four kVps.

Maximum skin dose predicted by CTDIvol (Gy)				
	Location1	Location2	Location3	Location4
80kVp	2.44	2.59	2.60	2.49
100kVp	2.44	2.59	2.60	2.49
120kVp	2.36	2.50	2.51	2.41
140kVp	2.36	2.49	2.48	2.39

IV. DISCUSSION and CONCLUSION

In this work, we have used Monte Carlo simulations to investigate the radiation dose delivered to eye lenses and skin during brain perfusion studies and the results shows the possibility of deterministic effects. The threshold of 2Gy for deterministic effects was chosen quite arbitrary here, however, it has been reported in recent studies that the threshold for cataract may be as low as 1Gy or there is no threshold at all, based on statistical analysis of postoperative cataract cases among atomic bomb survivors [6].

The eye lens dose and maximum skin dose from brain perfusion scans are highly related to the tube potential that was used. Higher kVp results in higher the dose even when the same mAs was used. This is reasonable because the x-ray intensity is proportional to the square of kVp, therefore there is larger amount of photons coming out of the x-ray tube at a higher kVp, even at the same mAs.

The dose to eye lenses for a 170 mAs/rotation scan protocol at 120kVp can be as high as 2Gy when 132 rotations are performed. This number of rotation can be easily achieved in two or three brain perfusion examinations. The dose to eye lenses is highly related to the scan location. When the scan range covers the eye lenses completely, radiation dose is an order of magnitude higher than the cases where the eye lenses are completely out of the primary beams. Therefore the contribution of scatter component to eye lens dose is fairly small. The skin dose for a 170 mAs/rotation scan protocol at 120kVp can be 2Gy when 110 rotations are performed. Since skin is a fairly uniformly distributed organ, the scan location does not make much difference to the skin dose.

Depending on the scan protocol (mAs/rotation) that was used, the number of rotations to reach threshold beyond which deterministic effects may happen can be very different, for example, for a 300 mAs/rotation scan at 120 kVp, the number of rotations to reach 2Gy for eye lenses and skin could be as low as 75 and 62, respectively. This could be sometimes achieved in a single brain perfusion examination. Since the brain perfusion scan protocols are not regulated among different clinic sites, it is very possible that one single brain perfusion examination could cause cataract or skin burn.

Although CTDI is defined as radiation dose to a 100cm pencil ion chamber in homogeneous phantoms and it does not necessarily equal to dose delivered to patients, CTDIvol reported on the scanner console is often used as dose that a patient receive. From this study, it was shown that CTDIvol overestimates skin dose by 20% to 30% (Table VIII), and it overestimates eye lens dose by 40% to 60% (Table IV). It should be noted that in this study only one phantom and one scanner were modeled, therefore future work includes the investigation of other scanner models and other phantoms with different sizes to see if CTDIvol predicts patient dose better. In summary, CTDIvol can be used as a conservative predictor for skin and eye lens dose in brain perfusion studies, but physicists and physicians should be aware that it does not equal to patient dose. This study provides detailed information about radiation dose to eye lens and skin from CT brain perfusion examinations and it could help to improve the design of scan protocols.

REFERENCES

[1] *NCRP*, "Ionizing Radiation Exposure of the Population of the United States," *NCRP Report No. 160, 2009.*

[2] G. Jarry, J. J. DeMarco, U. Beifuss, C. H. Cagnon and M. F. McNitt-Gray, "A Monte Carlo-based method to estimate radiation dose from spiral CT: from phantom testing to patient-specific models," Physics in Medicine and Biology, 2645, 2003.

[3] N. Petoussi-Henss, M. Zankl, U. Fill and D. Regulla, "The GSF family of voxel phantoms," Physics in Medicine and Biology, 89, 2002.

[4] ICRU, "Tissue Substitutes in Radiation Dosimetry and Measurement," ICRU Report No. 44, 1989.

[5] J. J. DeMarco, C. H. Cagnon, D. D. Cody, D. M. Stevens, C. H. McCollough, J. O. Daniel and M. F. McNitt-Gray, "A Monte Carlo based method to estimate radiation dose from multidetector CT (MDCT): cylindrical and anthropomorphic phantoms," Physics in Medicine and Biology, 3989, 2005.

[6] K.Neriishi, E. Nakashima, A. Minamoto, S. Fujiwara, M. Akahoshi, H.K. Mishima, T. Kitaoka, R.E. Shore. "Postoperative cataract cases among atomic bomb survivors: radiation dose response and threshold," Radiation Research, 168(4), 404, 2007.

Scanner- and Patient-Specific Organ Dose Estimates from CTDI and Patient Size Measurements

Adam C. Turner[1] and Michael McNitt-Gray[1]
[1]Departments of Biomedical Physics and Radiology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90024 USA

***Abstract*- : The overall purpose of this work was to develop a method to accurately estimate individual organ doses from MDCT exams that can be applied for any scanner, patient, and protocol combination. Feasibility studies have been performed that demonstrate: (a) that the use of $CTDI_{vol}$ measurements as organ dose normalization factors eliminate scanner-dependent dose differences across different MDCT scanners, and (b) the values of organ doses normalized by $CTDI_{vol}$ have a strong dependence on patient size. The first study implies a single $CTDI_{vol}$ to organ dose conversion coefficient exists for each organ which may be applied for any scanner. The second study illustrates that the values of $CTDI_{vol}$ to organ dose conversion coefficients scale with patient size, specifically the patient perimeter. This work suggests the feasibility of developing patient- and organ-specific $CTDI_{vol}$ to organ dose conversion coefficients that can be applied to any MDCT scanner to quickly and accurately estimate patient dose in the clinic.**

I. Introduction

Recent studies report the radiation dose delivered to the United States population from computed tomography (CT) exams have significantly increased [1]. It has been suggested that the most appropriate quantity for assessing the risk due to diagnostic imaging procedures is the radiation dose to individual organs [2]. These findings suggest that a method to quickly and accurately determine the dose delivered to the individual organs of patients undergoing CT examinations would be extremely useful in a clinical setting.

Patient dose from CT scans are typically evaluated using the CT dose index (CTDI) which is a metric measured in a cylindrical polymethyl methacrylate (PMMA) phantom and is meant to represent the average absorbed dose, along the z-axis, from a series of contiguous irradiations [3]. CTDI values are widely used for quality assurance and accreditation purposes, however, they are not intended to represent dose to any particular patient or, more importantly, to any particular organ.

Monte Carlo radiation transport techniques have made it possible to accurately estimate the radiation dose to radiosensitive organs in patient models from scans performed with modern Multidetector row CT (MDCT) scanners. However, it is not currently feasible to perform Monte Carlo simulations for patients on a routine clinical basis as the development of patient models requires image segmentation to properly define tissue boundaries [4].

The purpose of this work is to use Monte Carlo CT dosimetry simulations to develop a method to quickly and accurately estimate organ dose to any patient from any scanner. We hypothesize that organ dose can be obtained from a universal set of dose coefficients along with knowledge of a scanner's CTDI and a patient's size. This proposition will be explored in two phases: (a) first investigating the variance of scanner-specific simulated organ doses normalized by CTDI measurements (organ dose/CTDI) across a range of 64-slice MDCT scanners for a *single* patient model to investigate the feasibility of a scanner-independent CTDI to organ dose conversion coefficient; and (b) the size dependencies of CTDI to organ dose conversion coefficients will be explored by generating mean organ dose/CTDI values for a range of patient models and obtaining regression equations with patient size metrics as independent variables. Such equations would be useful for calculating patient-specific mean organ dose/CTDI values, which when used with scanner-specific CTDI measurements may be used to obtain scanner- and patient-specific organ dose values.

II. Methods

A. *Organ doses normalized by CTDI in a single patient model*

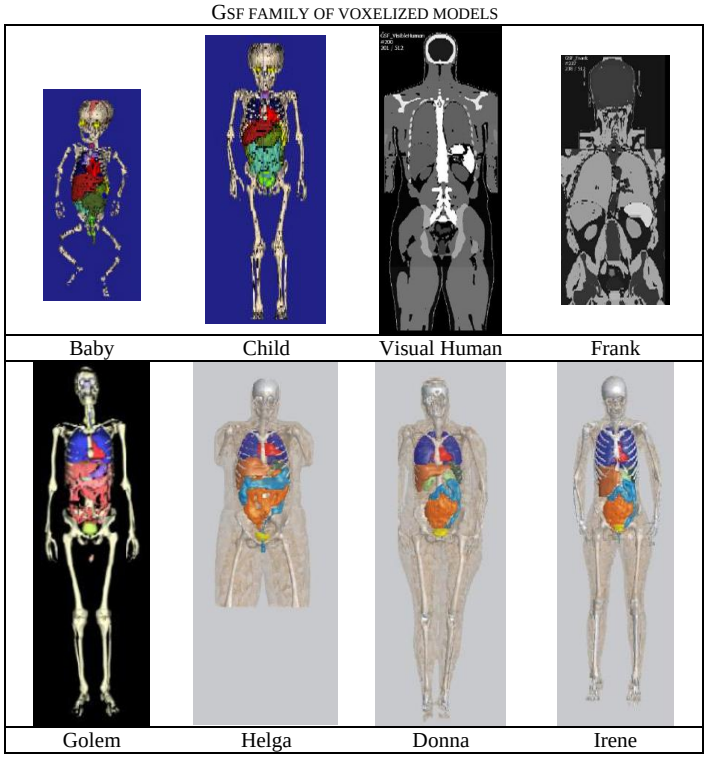
A previously presented Monte Carlo package [5][6] built on the MCNPX radiation transport code and modified to model specific MDCT scanner sources was used to simulate organ doses for this work. For this sub-study, organ dose was obtained using the Irene model from the GSF family of voxelized models, shown in Table I [7]. Helical CT exams were simulated for 64-slice MDCT scanners produced by the four major scanner manufacturers. For each scanner a 120 kVp exam was performed using a pitch of 1, the widest possible collimation, and the largest possible bowtie filter. Scans were performed over the entire body (head to toe) in order to ensure that each organ was fully encompassed in the scan length (fully-irradiated). Organ doses were obtained in units of mGy/mAs, as described in [5].

For each scanner and protocol used in the organ dose simulations the $CTDI_{vol}$ value was measured using the 32 cm (body) CTDI phantom, also in units of mGy/mAs, using the standard techniques outlined in [3].

For each scanner simulation the resulting organ dose values were normalized by the corresponding $CTDI_{vol}$ measurements. The variation across scanners was calculated for both un-normalized dose values (in mGy/mAs) and $CTDI_{vol}$ normalized dose values (unitless). Variance was calculated as the pcoefficient of variation (CoV), defined as the standard deviation divided by the mean of the organ dose and organ

dose normalized by CTDI values across scanners, and expressed as a percentage.

TABLE I
GSF FAMILY OF VOXELIZED MODELS



B. *Patient size dependency of organ doses normalized by CTDI values*

The organ dose simulation methods described in II.A were repeated using seven additional patient models, all from the GSF family of voxelized models shown in Table I. These models span a range of sizes, genders, ages, and body habitus. For this study the scan region was limited to the abdomen. For each model the abdominal scan range was defined from ~1 cm superior to the top of the diaphragm to ~1 cm inferior to the illiosacral joint. Simulations were again performed for four different 64-slice MDCT scanners to obtain dose (in mGy/mAs) to the radiosensitive organs in the abdomen. The resulting organ doses were normalized by the corresponding CTDI measurements and the average dose/CTDI value across scanners was obtained for each abdominal organ.

In order to characterize the size of each patient the perimeter of the central slice in the scan region for each patient was acquired in centimeters. For each abdominal organ, the mean dose/CTDI across scanners was plotted as a function of perimeter. Exponential regression equations were obtained in the form of:

$$\textit{mean organ dose/CTDI} = A\exp(B \times \textit{perimeter}) \quad (1)$$

where unique A and B values exist for each organ. The correlation coefficient (R^2) was also obtained for each organ.

III. RESULTS

A. *CTDI measurements*

The $CTDI_{vol}$ measurements obtained with the 32 cm (body) CTDI phantom using a tube voltage of 120 kVp and the widest possible scanner collimation for each scanner are reported in Table II on a per mAs basis. This table shows that the $CTDI_{vol}$ varies considerably across scanners (including a factor of two difference between scanner 2 and 4) and the CoV across scanners was 34.1%.

TABLE II
CTDI MEASUREMENTS

Scanner	$CTDI_{vol}$ (mGy/mAs)
1	0.063
2	0.062
3	0.089
4	0.123
CoV (%)	34.1%

B. *Absolute organ dose results for one patient model (Irene)*

The organ doses (in mGy/mAs) obtained for Irene, described in section II.A, are plotted in Figure 1 for Scanners 1-4.

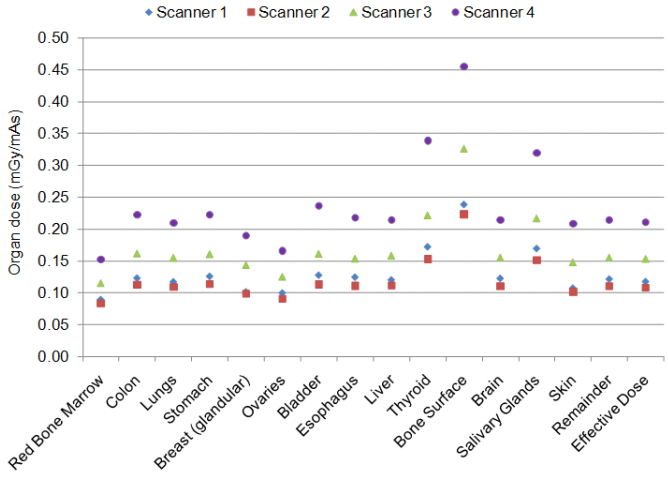


Fig. 1. Un-normalized organ dose values in mGy/mAs to radiosensitive organs from 4 different 64-slice MDCT scanners

It can be seen from the plot that, for most organs, there is a considerable difference in dose values between some of the different scanners. For example, the dose to most organs from Scanner 4 is approximately twice that of Scanner 2. These organ dose variations are very similar to those observed in CTDIvol values and the CoV across scanners ranged from 26.7% (for the adrenals) to 37.7% (for the thyroid), with a mean across all organs of 31.6%.

C. *Organ dose normalized by $CTDI_{vol}$ results for one patient model (Irene)*

The organ doses normalized by $CTDI_{vol}$ measurements (uniteless) obtained for Irene, as described in section II.A, are plotted in Figure 1 for Scanners 1-4.

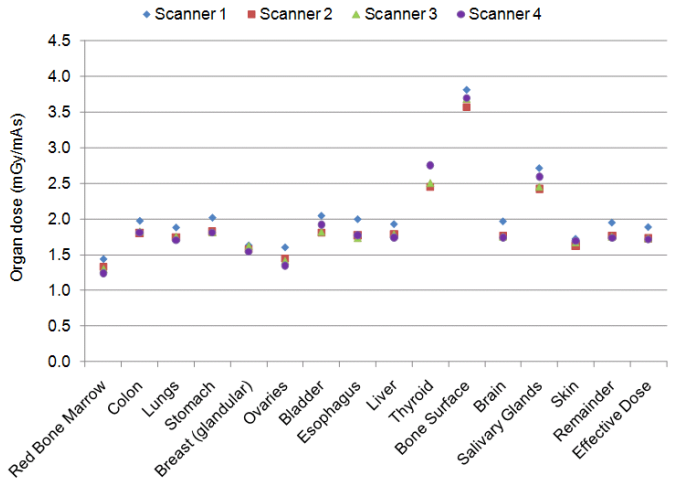


Fig. 2. Organ dose normalized by $CTDI_{vol}$ measurement values to radiosensitive organs from 4 different 64-slice MDCT scanners

Qualitatively, it is clear that the points in Figure 2 converge much more so than the points in Figure 1. The mean CoV across all organs was 5.2%, with a minimum of 2.4% (for skin tissue) and a maximum of 8.5% (for the adrenals).

D. Size dependency results for a range of patients undergoing simulated abdomen scans

For each GSF patient model undergoing a simulated abdomen scan , the mean organ dose/$CTDI_{vol}$ across scanners is plotted as a function of the patient perimeter in figure 3.

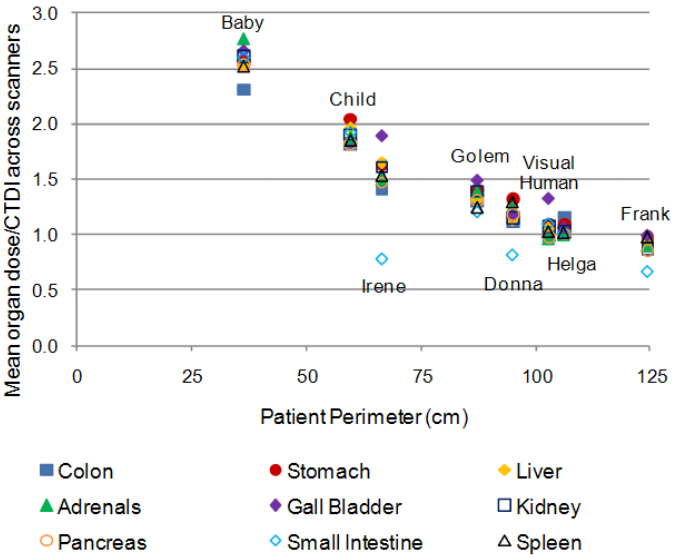


Fig. 3. Mean dose normalized by $CTDI_{vol}$ across scanners for each organ as a function of patient perimeter (in cm)

The plot suggests an exponential relationship between mean dose/$CTDI_{vol}$ across scanners for all organs except the small intestine. It should be noted that the small intestine was the only organ that was not fully encompassed in the scan range (partially-radiated) for each patient. Furthermore, the degree of partial-radiation varied from one GSF model to another (i.e. the percentage of the total small intestine included in the scan range was greater for the baby than it was for Irene).

The A and B coefficients derived from the exponential regression equations discussed in section II.B, which describe how mean dose/$CTDI_{vol}$ depends on patient perimeter, are presented in Table III. The correlation coefficient (R^2), which quantifies the degree that the regression lines approximate the data points, is also presented. These results demonstrate the excellent correlation between organ dose and patient size, especially for fully irradiated organs.

TABLE III
A and B coefficients and R^2 from exponential regression equations

Organ	A	B	R^2
Stomach	3.79	-0.0114	0.97
Liver	3.83	-0.0121	0.98
Adrenals	4.03	-0.0128	0.95
Gall Bladder	4.01	-0.0117	0.95
Kidney	3.97	-0.0125	0.99
Pancreas	3.72	-0.0123	0.97
Spleen	3.51	-0.0112	0.94
Colon	3.24	-0.0106	0.94
Small Intestine	3.36	-0.0127	0.62

IV. Discussion

The overall purpose of this project is to develop a method to accurately estimate individual organ doses from MDCT exams that can be applied for any scanner, patient, and protocol combination. The work presented here was meant to serve as feasibility studies to test the hypothesis that organ dose can be obtained from scanner-independent $CTDI_{vol}$ to dose conversion coefficients and that these coefficients can be derived for patients of varying sizes.

The first sub-study, discussed in section II.A, was carried out in order to investigate the reduction in organ dose variance across different MDCT scanners when the dose values were normalized by $CTDI_{vol}$ measurements. The CoV for un-normalized organ dose values across scanners ranged between 26.7% (for the adrenals) and 37.7% (for the thyroid), with a mean across all organs of 31.6%. It is suspected that the major factors contributing to organ dose differences among scanners is discrepancies in x-ray energy spectra and filtration designs (especially the bowtie filters). The mean CoV value was similar to the 34.1% CoV across scanners of the measured $CTDI_{vol}$ values. This indicates that scanner-specific characteristics leading to organ dose differences may be reflected in $CTDI_{vol}$ measurements and, furthermore, that organ dose normalization using $CTDI_{vol}$ values may remove those differences across scanners.

After normalizing organ dose values with $CTDI_{vol}$ measurements the variation across scanners reduced considerably. The mean CoV across all organs was 5.2%, with a minimum of 2.4% (for skin tissue) and a maximum of 8.5% (for the adrenals). These results suggest that $CTDI_{vol}$ measurements do account for differences among scanners leading to organ dose variations and that, for a given organ, the mean dose/$CTDI_{vol}$ across scanners is within 10% of any

individual dose/$CTDI_{vol}$ value. The implication of this conclusion is that the mean value can be used as a $CTDI_{vol}$ to organ dose coefficient that can be multiplied by any scanner's measured $CTDI_{vol}$ value to obtain organ dose for that scanner. Of course, the data presented in this study is only valid for the particular patient (Irene) and scanner protocol (120 kVp, pitch 1, widest collimation, large bowtie) for which it was generated. Furthermore, these coefficients are only pertinent to organs fully-irradiated in a head-to-toe scan. It would be necessary to generate protocol specific $CTDI_{vol}$ to organ dose coefficients for fully-irradiated organs in particular scans, such as thoracic, abdomen, or pelvic scans.

The second sub-study, discussed in section II.B, was performed in order to extend the proposed $CTDI_{vol}$ to organ dose estimation method to account for patients of varying size. The results indicate that, for fully-irradiated organs in a typical abdomen exam, $CTDI_{vol}$ normalized organ doses correlate well with patient size. Specifically, these values depend exponentially on patient perimeter, as shown in figure 3. Table III presents the A and B values for each organ generated by exponential regression techniques that can be applied, using equation 1, to obtain dose/$CTDI_{vol}$ values as a function of perimeter. The correlation coefficients listed in table III indicate that for all fully-irradiated abdominal organs the exponential functions are excellent approximations to the simulated data.

The small intestine, which was not fully encompassed in the scan regions for any of the patient models, indicated that this dose estimation approach may not be appropriate for partially-irradiated organs. Instead, the method must be extended as radiosensitive organs including red bone marrow, bone surface (endosteal tissue), skin, and muscle will almost always be only partially irradiated. Future work will involve partitioning these tissues into "in-scan" and "out-of-scan" regions and separately calculating dose to these two regions for each organ. It is hypothesized that the "in-scan" portion will behave similar to the fully-irradiated organs in this study and that the "out-of-scan" dose values will be very small.

IV. Conclusions

The work presented in this abstract represent the first feasibility studies in developing an organ dose estimation method based on scanner-specific $CTDI_{vol}$ values and patient-specific size measurements. The finding that the $CTDI_{vol}$ metric reflects scanner characteristics that cause dose differences across scanners suggests that, for a given patient and fully-irradiated organ, the mean dose/$CTDI_{vol}$ across all scanners can be used as a universal (scanner-independent) $CTDI_{vol}$ to organ dose conversion coefficient. Furthermore, for fully-irradiated organs, the mean dose/$CTDI_{vol}$ values correlate well with patient size, specifically the perimeter of the patient taken at the center of the scan region.

In this work A and B coefficients were derived for fully-irradiated organs for an abdomen scan that can be used with equation 1 and a patient perimeter in order to calculate a patient-size corrected mean dose/$CTDI_{vol}$ value. Then, with a knowledge of a particular scanner's measured $CTDI_{vol}$ the organ dose value in mGy/mAs can be obtained. Finally, by multiplying by the mAs used in the scan protocol an organ dose in mGy can be estimated. This process is displayed in figure 4. The results from these studies indicate that these organ dose estimates are accurate to within 10%, on average.

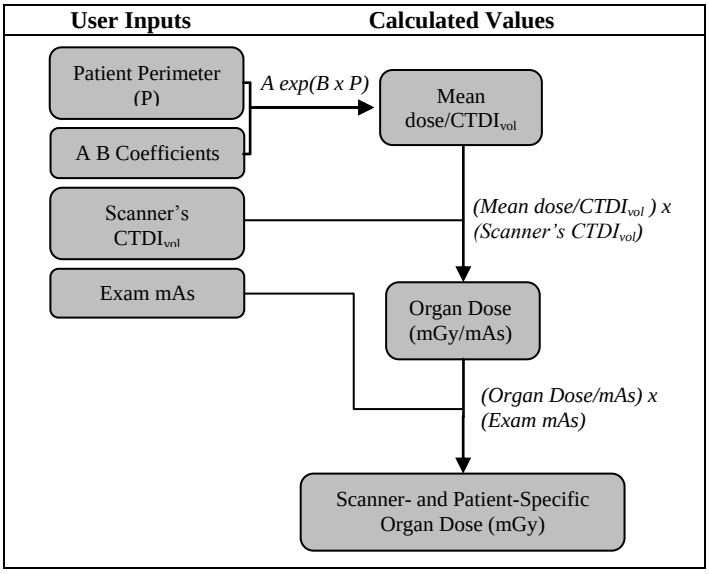


Fig. 4. Proposed Method for estimating scanner- and patient-specific organ dose values in mGy

Ackowledgments

This work was supported by NIBIB Grant R01-EB004898.

References

[1] F.A. Mettler Jr., B.R. Thomadsen, M. Bhargavan, D.B. Gilley, J.E. Gray, J.A. Lipoti, J. McCrohan, T.T. Yoshizumi, and M. Mahesh, "Medical radiation exposure in the U.S. in 2006: preliminary results," Health Phys. 95(5), 502-507 (2008).

[2] National Research Council. Health Risks from Exposure to Low Levels of Ionizing Radiation: BEIR VII Phase 2. The National Academies Press, Washington, DC. (2005).

[3] American Association of Physicists in Medicine. The measurement, reporting and management of radiation dose in CT. New York: AAPM; Report No 96; 2008.

[4] W. Bolch, C. Lee, M. Wayson, and P. Johnson, "Hybrid computational phantoms for medical dose reconstruction," *Radiat. Environ Biophis.* Epub ahead of print., Dec 2009.

[5] J.J. DeMarco, C.H. Cagnon, D.D. Cody, D.M. Stevens, C.H. McCollough, J. O'Daniel, and M.F. McNitt-Gray, "A Monte Carlo based method to estimate radiation dose from multidetector CT (MDCT): cylindrical and anthropomorphic phantoms," Phys. Med. Biol. 50 3989–4004 (2005).

[6] A.C. Turner, D. Zhang, H.J. Kim, J.J. DeMarco, C.H. Cagnon, E. Angel, D.D. Cody, D.M. Stevens, A.N. Primak, C.H. McCollough, and M.F. McNitt-Gray. "A method to generate equivalent energy spectra and filtration models based on measurement for multidetector CT Monte Carlo dosimetry simulations," Med. Phys. 36(6), 2154-2164 (2009).

[7] N. Petoussi-Henss, M. Zankl, U. Fill, and D. Regulla, "The GSF family of voxel phantoms," Phys. Med. Biol. 47, 89–106 (2002).

Digital breast tomosynthesis with a straight-line X-ray source in X-ray CT

Emil Y. Sidky[1], Yuval Duchin[1], Ingrid S. Reiser[1], Xiaochuan Pan[1], and Christer Ullberg[2]

***Abstract*—The adaptive-steepest-descent projection-onto-convex-sets (ASD-POCS) algorithm is applied to a digital breast tomosynthesis (DBT) system employing a straight-line X-ray source trajectory. This configuration along with the detector orientation allows for decomposition of the 3D volume into a set of 2D planes for expensive operations such as projection or back-projection. This decomposition is of particular interest for iterative algorithms, where computational effort is an important consideration. The ASD-POCS algorithm is also tested on data sets with many ray measurements each of which have a relatively high level of noise.**

I. Introduction

The XCounter DBT system presents two interesting features from the point of view of image reconstruction research, especially for iterative algorithms. First, it employs a photon counting detector which has no significant noise other than that due to the counting statistics of the detected photons. The practical advantage of this type of detector is that the flux can be divided among many measurements without paying a penalty of additional electronic noise. Obviously, this capability provides a challenge to iterative image reconstruction to be able to fully utilize the information from the extra measurements while minimizing the impact of the associated high noise levels from the small number of photons detected in each bin. Second, this DBT system employs a linear X-ray source trajectory. Such a trajectory allows for a decomposition of the reconstruction volume into a set of planes, each of which contain the source trajectory and measurements along one set of channels of the detector. Such a volume decomposition can result in a substantial improvement in terms of computation time.

In a previous work [1], we have described the ASD-POCS framework for constructing image reconstruction algorithms based on constrained minimization. The purpose of this framework is to design iterative algorithms with computed tomography (CT) in mind. In general, X-ray tomographic modalities provide data with many more counts than nuclear medicine imaging – currently the main application of iterative image reconstruction. As a result, the usual methods for regularizing images may be too strong, or rough, for X-ray tomography. The ASD-POCS framework is designed to allow finer control. Within this framework, we investigated a practical algorithm that can yield useful images with iteration numbers on the order of ten. The application in Ref. [1] was to a GE prototype DBT scanner that had fewer measurements with lower noise than the present XCounter scanner. We stress here that we make no comparison between the scanners. Instead, one of the purposes of this work is to try out the algorithm of Ref. [1] on a very different data set, which has more measurements at the expense of higher noise.

This abstract is organized as follows: in Sec. II the XCounter configuration is discussed along with the decomposition of the image volume, in Sec. III the application of ASD-POCS to this system is described, and in Sec. IV resulting images are shown from clinical data.

II. System configuration and volume decomposition

A schematic of the XCounter DBT system is shown in Fig. 1. With this configuration, the imaging volume can be conveniently represented in a cylindrical grid with the X-ray source trajectory as its axis as seen in Fig. 2. The advantage of such a grid is twofold: (1) the image reconstruction problem can be decoupled into a stack of 2D problems, and (2) such a geometry potentially allows one to take advantage of magnification inherent in the divergent beam projection. With the present arrangement, resolution of the volume is not uniform. It is higher near the source trajectory.

Aside from geometry, a major difference between the current DBT device and “traditional” DBT is how the X-ray illumination is divided up. DBT scanners generally all illuminate the subject to a dose equivalent to that of a mammogram. Our previous study in Ref. [1]

[1]The University of Chicago, Department of Radiology MC-2026, 5841 S. Maryland Avenue, Chicago IL, 60637; [2]XCounter AB Svärdvägen 11, SE-182 33 Danderyd, Sweden

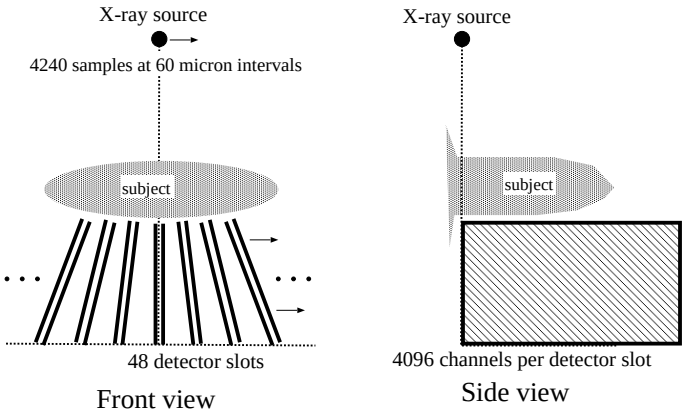


Fig. 1. The X-ray source is mounted on a frame with 48 slot detectors. Each slot can utilize up to four strips with each strip comprised of a row of 1024 bins of dimension (60 μm)^2. The source-detector assembly is translated linearly in 4240 steps of 60 μm. In effect each slot provides a projection consisting of a 4240 parallel fans. Because the slots look back at the source from different angles, they provide projections from views separated by 0.56 degrees, ranging from -13.54 to 13.54 degrees. From the side view, it can be seen that the X-ray illumination is a half-cone beam.

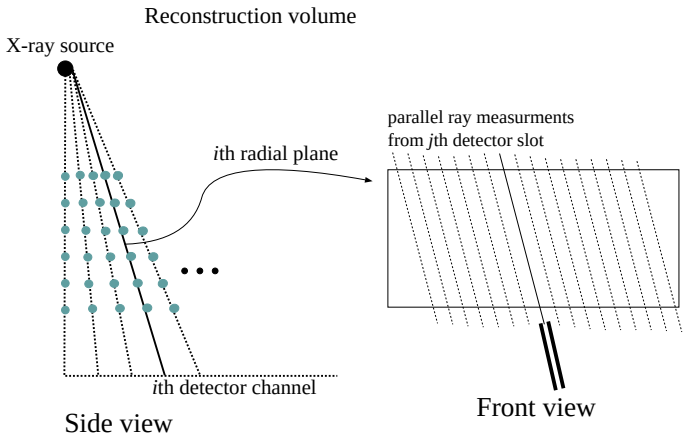


Fig. 2. Schematic of cylindrical volume used for image reconstruction. The ith radial plane contains the X-ray source trajectory and the ith channel from each of the 48 slots. The radial samples indicated by the light blue dots are arranged so that the samples from the different radial planes can be grouped together to form planes parallel to the detector plane. It is these parallel planes that contain the DBT images that will be shown. The projection data for the ith channel can be organized as a set of parallel beam projections (shown are the parallel rays for the jth slot) for the ith radial plane, where each slot contributes a different view angle.

used projection data sets with 11 views and 1800x2304 detectors at 100 μm resolution. For the current study, the system can provide 48 views and the equivalent detector dimensions would be 4096x4240 at 60 μm resolution. The former system divides the X-ray intensity among $\approx 46 \times 10^6$ ray measurements, while the latter system utilizes $\approx 834 \times 10^6$ ray measurements.

III. Implementation of ASD-POCS

The algorithm developed for DBT, within the ASD-POCS framework was presented in detail in Ref. [1], including a pseudo-code. Here, we employ the same algorithm aside from specifics related to the scanning geometry. The algorithm is guided by the following constrained, total variation (TV) minimization problem:

$$\vec{f}^* = \text{argmin}\|\vec{f}\|_{TV} \text{ subject to } |X\vec{f} - \tilde{g}|^2 \leq \epsilon^2, \quad (1)$$

where $\vec{f}$ represents the array of image voxel values; $\tilde{g}$ the array of ray measurements; X, the system matrix, models the forward projection of the discrete image array $\vec{f}$; and data-error tolerance ϵ is a parameter of the optimization problem. The norm $\|\cdot\|_{TV}$ is the discretized version of the following integral $\int_V d\mathbf{r}|\vec{\nabla} f(\mathbf{r})|$ over the continuous image function. While the algorithm put forth in Ref. [2] solves this optimization problem fairly accurately, it requires too many iterations for the present DBT systems, which have large data sets. Instead, the DBT algorithm in Ref. [1] is designed to run in ten iterations. Both algorithms alternate POCS (or ART), to enforce the data consistency constraint, and gradient descent, to minimize the image TV. The DBT version includes back-tracking on the gradient descent part to insure that each of these steps is close to the largest possible while actually decreasing image TV.

The main controlling parameter for the ASD-POCS DBT algorithm is the relaxation parameter β on the POCS, which is chosen in $(0, 1]$. Specifically, in the POCS loop, the image is updated after processing each ray measurement g_j:

$$\vec{f} \leftarrow \vec{f} + \beta\vec{X}_j \frac{g_j - \vec{X}_j \cdot \vec{f}}{|\vec{X}_j|^2}, \quad (2)$$

where $\vec{X}_j$ is the jth row vector of X yielding an estimate for the ray measurement g_j. When $\beta = 1$, the image estimate $\vec{f}$ is projected to exact consistency with the data measurement g_j, and smaller values changes the image estimate less. Because the gradient descent part of the algorithm tries to take large steps which do not change much in step-length, β is used to control the balance between the POCS step and the gradient descent step. The effect of smaller β is to regularize the image more.

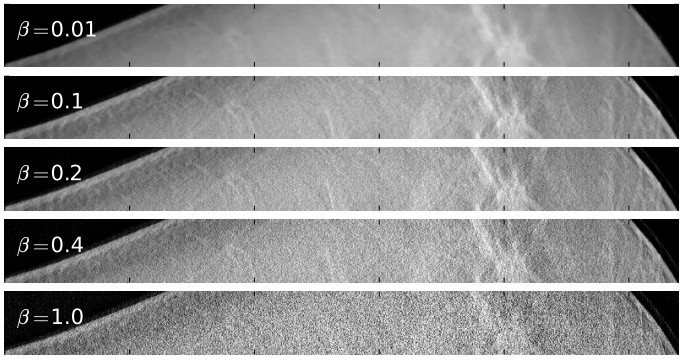


Fig. 3. DBT images reconstructed by the ASD-POCS algorithm for different values of β. The iteration number in each case was 10. The gray scale for each image is [0:0.5].

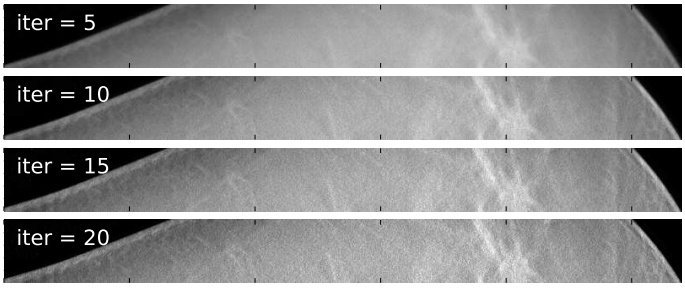


Fig. 4. DBT images from a basic implementation of EM with no explicit regularization. Regularization is achieved only through iteration truncation. The gray scale for each image is [0:0.5].

The published DBT algorithm also included a parameter p on the TV-norm that we do not investigate here. For the results presented here, β is varied and the iteration number is fixed at 10. We do not claim to solve Eq. (1) with this algorithm, but the resulting images can be of clinical utility.

Specific adaptations for the cylindrical geometry of the present DBT scanner include a substantial acceleration of the POCS-step. One iteration of POCS is performed on each radial plane independent of all other radial planes. This arrangement allows for the contribution of each measurement ray to be computed much more efficiently, because each ray is confined to a 2D plane. Also, on the POCS iteration, a computing cluster can be efficiently employed with each processing unit responsible for a single plane. The gradient of the TV-norm remains as a 3D computation, but its execution time is negligible in comparison with the POCS-step. For the present implementation simple finite differencing is used directly on the cylindrical grid without accounting for the coordinate change. Thus, we are not accurately modeling the image gradient, but this difference does not impact image quality of the resulting reconstructions.

IV. Results

The application of the ASD-POCS DBT algorithm to scanner data from the GE prototype appeared to be successful in that the parameter β could provide reasonable control over the trade-off between image noise and resolution. Here, we investigate this algorithm on a single data set from the XCounter system, which has many more measurements but more noise-per-measurement. The reconstruction volume, as stated earlier, lies on cylindrical grid. The radial sampling is adjusted so that planes parallel to the compression paddle lie on the grid, see Fig. 2. The parallel resolution is the same as the detector, while the radial resolution is selected so that the parallel planes are spaced 1 mm apart. A total of 100 parallel planes comprise the volume.

A sequence of images are shown in Fig. 3 for different values of β through a section of breast that contains a mass. For these results, only ten iterations of ASD-POCS were performed. As can be seen, the β parameter provides an effective algorithm control for image regularization. At the smallest values of β, the speckle noise is almost completely smoothed. As β increases the image sharpness improves, but the speckle noise becomes more apparent.

For comparison, the same data set is reconstructed by a basic implementation of the expectation maximization (EM) algorithm. This algorithm updates the image multiplicatively with the following update step:

$$\vec{f} \leftarrow \vec{f} \left(\frac{X^T \left(\tilde{g} / X\vec{f} \right)}{X^T \tilde{1}} \right), \qquad (3)$$

where $\tilde{1}$ symbolizes a data array filled with ones. The transpose of the system matrix X^T represents back-projection. We note here that for the present EM implementation X^T is the true transpose of X, which we use in order to preserve resolution. As an aside, EM is often implemented with a "ray-driven" forward projector and a "pixel-driven" back projector in order to remove conspicuous artifacts [3], but the use of mis-matched projectors, here, will compromise image resolution. Image regularization is achieved here with EM only by truncation of the iteration number. Iteration to convergence leads to the maximum likelihood image estimate, but such an image will be too noisy. Earlier iterations, however, can be of clinical use. In fact, the image reconstruction algorithm used for the GE scanner prototype at Massachusetts General Hospital (MGH) employs EM with 10 or fewer iterations [4].

The results of this basic EM implementation are shown in Fig. 4. As expected, the earlier iterations

yield smoother images where noise is washed out but so are some details of the underlying structure. As the iteration number increases more structures are apparent, but even by iteration 20, when the noise levels are quite noticeably higher, the image sharpness is not as high as the ASD-POCS results for β = 0.1 or 0.2. Such a comparison, however, is only meant to place some kind of context for the ASD-POCS algorithm. The EM implementation could possibly be improved by introducing Gibbs smoothing [5], where the TV norm could be introduced as an edge-preserving regularizer [6]. But our experience is that such regularization is difficult to control for underdetermined systems such as DBT. Images become blurred rapidly with modest values of the regularization parameter, and smaller values require the algorithm to be run to high iteration numbers.

V. Discussion

In this abstract we have presented results on DBT image reconstruction by ASD-POCS for a data set acquired on the XCounter DBT scanner, which incorporates a photon counting detector. This data set is interesting to test reconstruction algorithm on, because it divides the dose amongst an order-of-magnitude more detection channels than most other DBT systems. As a result, this data set is at one end of an extreme in terms of having a high-noise level and, on the other hand, high-resolution measurements.

From a technical point-of-view this particular scanner is interesting because of the straight-line scanning trajectory. This geometry allows for a cylindrical decomposition of the reconstruction volume that can substantially improved algorithm efficiency. We have implemented this non-standard grid in the present version of ASD-POCS. This non-standard grid would also allow for full utilization of the magnification factors of the diverging X-ray beam.

The results of the ASD-POCS image reconstruction appear promising. Even though only ten iterations were used, the parameter β appears to effectively control the trade-off between image-resolution and image-noise. At the conference, we will present results from other data sets taken with the XCounter scanner, showing performance on the challenging task of micro-calcification imaging. Furthermore, quantitative metrics will be developed that summarize resolution and noise-level. Comparisons with transmission-EM, which may have some advantage for high-noise data sets with photon counting detectors, will also be shown.

VI. Acknowledgment

E.Y.S., Y.D. and X.P. were supported in part by NIH R01 Grant Nos. CA120540 and EB000225. E.Y.S. was also supported in part by a Career Development Award from NIH SPORE grant CA125183-03. I.S.R. was supported in part by NIH Grant Nos. R33 CA109963 and R21 EB8801. Some computations for this work were performed on a cluster, partially funded by NIH Grant Nos. S10 RR021039 and P30 CA14599. The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

References

[1] EY Sidky, X Pan, IS Reiser, RM Nishikawa, RH Moore, and DB Kopans, "Enhanced imaging of microcalcifications in digital breast tomosynthesis through improved image-reconstruction algorithms," *Med. Phys.*, vol. 36, no. 11, pp. 4920–4932, NOV 2009.

[2] EY Sidky and X Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Phys. Med. Biol.*, vol. 53, no. 17, pp. 4777–4807, SEP 7 2008.

[3] B De Man and S Basu, "Distance-driven projection and back-projection in three dimensions," *Phys. Med. Biol.*, vol. 49, no. 11, pp. 2463–2475, JUN 7 2004.

[4] T Wu, RH Moore, EA Rafferty, and DB Kopans, "A comparison of reconstruction algorithms for breast tomosynthesis," *Med. Phys.*, vol. 31, no. 9, pp. 2636–2647, SEP 2004.

[5] K Lange, "Convergence of em image-reconstruction algorithms with gibbs smoothing," *IEEE Trans. Med. Imag.*, vol. 9, no. 4, pp. 439–446, DEC 1990.

[6] M Persson, D Bone, and H Elmqvist, "Total variation norm for three-dimensional iterative reconstruction in limited view angle tomography," *Phys. Med. Biol.*, vol. 46, no. 3, pp. 853–866, MAR 2001.

Investigating the Effect of Scattered Radiation in Dedicated Breast CT Imaging

Stephen J. Glick, Ph.D.
Department of Radiology
University of Massachusetts Medical School
Worcester, MA 01655
Stephen.Glick@umassmed.edu

Clay S. Didier, M.S.
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology
Cambridge, MA 02139

Abstract— Recently, there has been great interest in imaging the breast with a dedicated cone-beam CT scanner. This new imaging modality can provide 3D visualization of breast tissue without the structural overlap of breast tissue that hinders conventional mammography. Although great progress has been made in the development of dedicated flat-panel CT breast imaging, there are a number of important questions related to the application of these systems that should be addressed before they can be successfully used for routine clinical imaging. One of the primary concerns is how the rather large scattered radiation component affects the resulting image quality. In this study, we use computer simulation to investigate the degradation in microcalcification detection accuracy with and without the presence of scattered radiation. It was observed that for the imaging parameters studied in this simulation, scattered radiation introduced a significant penalty for the task of detecting small microcalcifications, with a difference in sensitivity of 15.6%.

I. Introduction

Although x-ray mammography has saved many lives and is considered the imaging modality of choice for early detection of breast cancer, it is far from perfect. One of the limiting problems with the conventional method for mammography is that the recorded image represents the superposition of a three-dimensional (3D) object onto a two-dimensional (2D) plane. Thus although breast compression is typically performed, normal anatomical structure like the parenchymal tissue can combine with useful diagnostic information (*e.g.*, a tumor) in such a way as to impede visualization and reduce lesion detectability. Although x-ray mammography has saved many lives and is considered the imaging modality of choice for early detection of breast cancer, it is far from perfect. One of the limiting problems with the conventional method for mammography is that the recorded image represents the superposition of a three-dimensional (3D) object onto a two-dimensional (2D) plane. Thus although breast compression is typically performed, normal anatomical structure like the parenchymal tissue can combine with useful diagnostic information (*e.g.*, a tumor) in such a way as to impede visualization and reduce lesion detectability.

Recently, there has been great interest in dedicated, cone-beam flat-panel breast CT imaging of the breast. In addition to our group at University of Massachusetts, a number of other academic groups are currently studying flat-panel CT breast imaging (CTBI) (Boone, Delson et al. 2001; Chen and Ning 2002; Tornai, McKinley et al. 2005; Yang, Kwan et al. 2007; Kalender 2009). Although progress has been made in investigating the feasibility of dedicated flat-panel CTBI, there are a number of important questions related to the application of these systems that should be addressed before they can be successfully used for routine clinical imaging. One of the primary concerns is how to reduce the degrading effects of the rather large scattered radiation component in the resulting images.

(Siewerdsen and Jaffray 2001) have reported that the measured detected scattered radiation with flat-panel cone-beam CT imaging increases with increasing cone angle, leading to image artifacts, a reduction in image contrast, and quantitative errors in measured CT numbers. (Kwan, Boone et al. 2005) have measured scatter properties on a prototype CT breast imaging scanner and reported scatter-to-primary ratios of up to 100% (depending on breast size). They conclude that scattered radiation in CTBI would likely impact image quality. Our group has also evaluated the characteristics of scatter using both experimental measurements and Monte Carlo simulation studies (Chen, Liu et al. March 2009) of a dedicated CTBI system and confirmed the large scatter component reported by Kwan et al. Based on these characterization studies, it appears as if scattered radiation can have important effects on image quality in flat-panel CTBI. If this is the case, it will be necessary to develop methods for reducing the scatter in CTBI.

To maximize the differentiation between malignant and benign lesions and thus improve the specificity of lesion analysis (with either non-contrast CTBI or contrast-enhanced CTBI), it is important to accurately estimate the CT number within suspected lesions. To date, no comprehensive studies using dedicated flat-panel breast CT systems have been conducted to analyze quantitative CT numbers that might have predictive value in assessing breast cancer. However, (Miyake, Hayakawa et al. 2005) have imaged the breast with a conventional CT scanner to examine lesion CT number, and to determine its' value for lesion analysis. Miyake et al reported a distribution of CT

values for both malignant (n=154 lesions) and benign (n=22 lesions) lesions that were studied, suggesting that a cutoff of 60 HU would provide good discrimination. This study was performed with a conventional CT scanner, where scattered radiation is much less of a problem. However, dedicated cone-beam CTBI systems have a larger scatter component due to the un-collimated large radiation field (up to 100% scatter-to-primary ratio). Scatter causes a "cupping artifact" in the reconstruction that results in the CT number of lesions to be spatially (within the breast) dependent. In addition, due to scatter the lesion CT number will be patient dependent, i.e., the lesion CT number will vary with breast size and shape. Because of this variability, scatter in flat-panel cone-beam CT can broaden benign and malignant CT value distributions, thus making lesion differentiation analysis more difficult.

It is well known that the detection of erroneously positioned scattered x-rays can degrade lesion contrast and signal-to-noise ratio (SNR) in CT imaging (Macovski 1983; Endo and Tsunoo 2001). Maximizing SNR in CT breast imaging is important for the accurate detection of masses as well as smaller microcalcifications. Although CTBI provides a superb portrayal of lesion morphology, some masses visualized in CTBI have poor SNR due to the low contrast between tumors and surrounding fibroglandular tissue for the energy range used in breast CT (40-80 kVp). This low contrast was reported many years ago by (Johns and Yaffe 1987) in a comprehensive study evaluating the x-ray absorption properties of normal and neoplastic breast tissue. Because of the inherent low contrast between lesions and normal fibroglandular tissue, it is especially important to reduce the additional contrast reduction caused by scatter.

Another degrading effect of scatter measured in CTBI is reduced microcalcification detection accuracy. The visualization of microcalcifications is very important for the accurate detection of ductal carcinoma in situ (DCIS). With conventional mammography, 90% of DCIS is identified on the basis of suspicious microcalcifications (Dershaw, Abramson et al. 1989). Since 14-50% of all DCIS eventually becomes invasive (Kopans 1998), detection of DCIS is important and can contribute to a decreased breast cancer mortality rate. In this study, we use computer simulation to investigate the degradation in microcalcification detection accuracy with and without the presence of scattered radiation. We hypothesize that reducing the large scatter component in tomographic breast imaging can improve detection of microcalcifications. If this hypothesis is true, then reduction of scatter would most likely improve the detection and diagnosis of breast cancer.

II. METHODS

To evaluate how scatter can affect detection accuracy of microcalcifications in CTBI, we performed a computer simulation study. Simulation is a powerful methodology for assessing the effect of scatter on image quality, because it is straightforward to generate simulated projection measurements with and without including scattered x-rays. For a number of years now, our group has been developing a computer simulation model for cone-beam CTBI (Vedula, Glick et al. 2003; Didier, O'Connor et al. 2009). This C++ language software models CTBI using a CsI based amorphous silicon flat-panel detector. Three stages are involved in simulating each system: 1) modeling the x-ray spectra, and scaling the x-ray fluence to provide the appropriate radiation dose, 2) determining the x-ray transmission through the breast phantom, and 3) modeling the signal and noise propagation through the CsI based detector. To model the scatter distribution incident on the detector, very long (i.e, with minimal quantum noise) Monte Carlo simulations were run for varying x-ray energies using the Penelope software [ ], and then the spatially dependent scatter-to-primary ratio (SPR) for each x-ray energy in the spectra were formulated. This SPR distribution generated from Monte Carlo simulation was then used in conjunction with the primary x-ray projection (generated using ray-tracing for each 1 keV energy interval) to estimate the scatter x-ray fluence incident on the detector.

Using this methodology, projection data were formed with and without scattered radiation of a 16 cm diameter cylindrical breast object with embedded spherical microcalcifications of size 360 microns. The x-ray absorption properties for the microcalcifications were modeled as hydroxyapatite. The background cylinder was modeled as a mixture composition of 99% fibroglandular tissue, and 1% adipose tissue. Although somewhat unrealistic, this composition was selected because it models a very dense breast, and makes for a challenging detection task. Microcalcifications embedded in less dense tissue (i.e., adipose tissue) would be easier to detect, however, most DCIS is found in more dense glandular tissue. Micro-calcifications were randomly placed at positions in a 5 x 5 grid (i.e, at each of the positions in a 5 x 5 grid, a "coin flip" determined whether the calcification was present or not)(see Figure 1). The exposure technique modeled was a 50 kVp tungsten anode spectra (with 2 mm Al filter), providing a mean glandular dose of 12 mGy. The simulation modeled the UMass bench-top CTBI prototype system using a Varian PaxScan 2520 with 127 μm pixel size and 2x2 binning on the readout, with the collection of 300 projection views. A 800 μm x-ray focal spot was modeled. Reconstruction was performed with Feldkamp's filtered backprojection algorithm (Feldkamp, Davis et al. 1984).

Three observers (medical physicists) were asked to indicate whether a calcification was present or absent at each of the 5 x 5 locations indicated by grid lines demonstrating possible locations (see Fig. 1). Prior to the reader experiment, a training session was conducted in which the observer viewed 50 microcalcifications (i.e., two 5 x 5 clusters) for each case, followed by a presentation of the true calcification locations. After the training session, each observer viewed 100 microcalcifications (i.e., four 5 x 5 clusters) for each case (scatter or no scatter). Shown in Fig

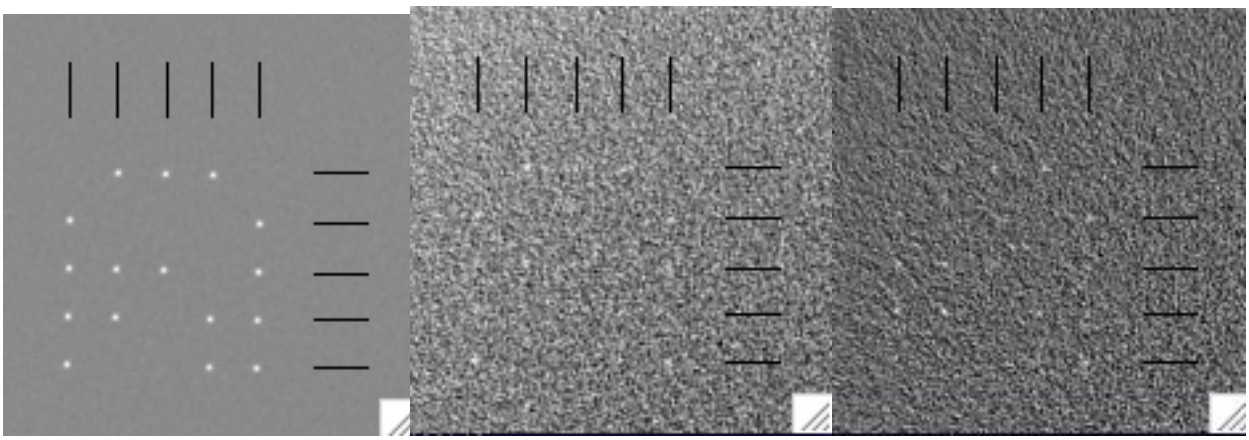
1 is one realization of calcifications randomly positioned on a 5 x 5 grid. Fig 1a shows the true locations of the microcalcifications for this realization (this is a reconstruction with very high dose 1000 mGy), and Figs 1b and 1c show simulated reconstructions with, and without scatter.

III. Results and discussion

For the simulation without scatter, three observers performed with a mean sensitivity of 83.6% (6.8%) (the standard deviation is reported in parenthesis) and average specificity of 98.1% (1.9%), whereas for the simulation with scatter, average sensitivity was 68.0% (7.4%) and average specificity was 94.9% (3.9%). In summary, it was observed that for the imaging parameters studied in this simulation, scattered radiation introduced a significant penalty for the task of detecting small microcalcifications, with a difference in sensitivity of 15.6%.

It is expected that similar performance differences would be observed for smaller microcalcifications embedded in less dense tissue. We plan to perform additional observer studies following the methods above to evaluate performance with different density backgrounds.

Detection of microcalcifications is very important for the accurate detection of ductal carcinoma in situ (DCIS). With conventional mammography, 90% of DCIS is identified on the basis of suspicious microcalcifications (Dershaw, Abramson et al. 1989). The results from this preliminary study suggest that scattered radiation in breast CT can introduce a significant penalty in terms of microcalcification detection accuracy. This result provides motivation for the development of techniques for reducing the effect of scatter in breast CT. These techniques could include such approaches as the development of new anti-scatter grids, or the development of new software based scatter compensation methods.



(a) (b) (c)

Fig 4.1. Example reconstructions of microcalcifications used for the observer study described herein. (a) shows a reconstruction at very high dose (essentially noise-free), (b) shows a CT reconstructed slice with x-ray scatter included in the simulation, and (c) shows a CT reconstructed slice without scatter modeled in the simulation. Observers were asked to record whether a calcification was present or absent at each location indicated by the intersection of the hash marks on the top and right sides. Each image is displayed using thresholding with the maximum and minimum pixel value.

Acknowledgment

This research was supported in part by the National Cancer Institute under grant number CA134128. The contents of this paper are solely the responsibility of the authors and do not represent the official views of the NIH.

REFERENCES

Boone, J. M., T. R. Delson, et al. (2001). "Dedicated Breast CT: Radiation Dose and Image Quality Evaluation." Radiology **221**(3): 657-677.

Chen, B. and R. Ning (2002). "Cone-beam volume CT breast imaging: Feasibility study." Med. Phys. **29**(5): 755-770.

Chen, Y., B. Liu, et al. (March 2009). "Characterization of scatter in cone-beam CT breast imaging: Comparison of experimental measurements and Monte Carlo simulation." Med Phys **36**(3): 857-869.

Dershaw, D., A. Abramson, et al. (1989). "Ductal carcinoma in situ: Mammographic findings and clinical implications." Radiology **170**: 411-415.

Didier, C. S., J. M. O'Connor, et al. (2009). "Enhanced computer simulation of CT mammography using a flat-panel imager." SPIE Medical Imaging **7258-18**.

Endo, M. and T. Tsunoo (2001). "Effect of scattered radiation on image noise in cone-beam CT." Med Phys **28**: 469-474.

Feldkamp, L. A., L. C. Davis, et al. (1984). "Practical cone-beam algorithm." J. Opt. Soc. Am. **1**: 612-619.

Johns, P. C. and M. J. Yaffe (1987). "X-ray characterisation of normal and neoplastic breast tissues." Phys. Med. Biol. **32**(6): 675-695.

Kalender, W. A. (2009). "X-Ray computed tomography: State of the art and future directions." presented at the World Congress of Medical Physics and Biomedical Engineering.

Kopans, D. B. (1998). Breast Imaging. Philadelphia, Lippincott, Williams, and Wilkins

Kwan, A. L. C., J. M. Boone, et al. (2005). "Evaluation of x-ray scatter properties in a dedicated cone-beam breast CT scanner." Med Phys **32**: 2967-2975.

Macovski, A. (1983). Medical Imaging Systems. Englewood Cliffs, NJ, Prentice Hall, Inc.

Miyake, K., K. Hayakawa, et al. (2005). "Benign or malignant? Differentiating breast lesions with computed tomography attenuation values on dynamic computed tomography mammography." J Comp Assisted Tomo **29**: 772-779.

Siewerdsen, J. H. and D. A. Jaffray (2001). "Cone-beam computed tomography with a flat-panel imager: Magnitude and effects of x-ray scatter." Med. Phys. **2001**(28): 220-231.

Tornai, M. P., R. L. McKinley, et al. (2005). "Design and development of a fully-3D dedicated x-ray computed mammotomography system." Proc. SPIE Physics of Medical Imaging **5745**.

Vedula, A. A., S. J. Glick, et al. (2003). "Computer simulation of CT mammography using a flat-panel imager." SPIE Physics of Medical Imaging, **vol. 5030**: 349-360.

Yang, K., A. L. C. Kwan, et al. (2007). "Evaluation of the spatial resolution of a dedicated breast CT system using computer simulation." SPIE Medical Imaging, **6510**.

4D micro-CT for Small Animal Imaging

Cristian T. Badea, Samuel M. Johnston, Yi Qi, G. Alan Johnson

***Abstract*—Preclinical imaging has become an important area of research in radiology. The focus of this work is on the development of 4D Micro-CT for functional cardiopulmonary imaging in the mouse. Sampling and reconstruction methods for 4D cardiac and perfusion imaging are presented. A dual tube/detector micro-CT system is used for image acquisition with retrospective gating. We achieve 4D cardiac micro-CT imaging with isotropic voxels of 88 microns and 10 ms temporal resolution. The reconstruction methods involve a point spread function deconvolution or fast iterative algorithms based on total variation implemented on a graphic processor unit. Perfusion studies are based on the use of multiple injections delivered via tail vein. We demonstrate 4D lung perfusion images in the mouse at 88 microns and 1.38 sec temporal resolution. The radiation dose associated with the proposed methods is in the range of typical micro-CT dose (<0.2 Gy). The 4D micro-CT imaging presented here can be applied in high throughput longitudinal studies and has immediate applications in a wide range of applications in drug safety and cardiopulmonary phenotyping.**

I. INTRODUCTION

Clinical imaging tools (CT,MRI and ultrasound) for human cardiopulmonary morphology and function are well developed. But there are considerable challenges in scaling these modalities to the temporal and spatial resolution required for the mouse, since a murine heart is only about 5 mm in diameter and can beat up to 600 times/minute. To date, the most frequently applied imaging modalities in murine models of cardiovascular disease are echocardiography and magnetic resonance imaging (MRI). The advantages of cardiac MRI include its tomographic nature, high spatial and temporal resolution and good soft tissue contrast. However, small animal MR studies are expensive, complex, and time-consuming. Many studies have therefore utilized echocardiography, at the expense of sensitivity. Echocardiography is very fast but 3D measurements are model based and very user-dependent. We were the first to demonstrate cine cardiac micro-CT[1]. Our initial method used prospective gating and was therefore somewhat slow in sampling. Other groups have performed cardiac micro-CT studies in mice using slip-ring flat panel based CT[2]. Recently, we have extended our dynamic imaging to include perfusion in the rodent[3]. The method involved an invasive jugular vein catheter that made longitudinal studies difficult. The focus of this work is to present our novel integrated Micro-CT solutions for 4D imaging in the mouse. The new developments address previous limitations. Sampling and reconstruction methods for fast high throughput 4D cardiac and perfusion micro-CT are presented. Cardiac micro-CT imaging is achieved via retrospective gating and supported by graphics processing unit (GPU) for fast reconstruction. Perfusion studies are based on the use of multiple injections delivered via tail vein.

Authors are with the Center For In Vivo Microscopy, Duke University, Durham, NC, 2771, USA (e-mail: Cristian. Badea @duke.edu).

II. DUAL TUBE/DETECTOR MICRO-CT SYSTEM

We use a dual tube/detector micro-CT system that we have developed explicity for these studies [4] and novel strategies in gating/sampling/reconstruction adapted for this hardware.

The x-ray tubes and the detectors are arranged orthogonally. A description of the components of this system follows (see Fig.1). The system involves:

-2 G-297 x-ray tubes (Varian Medical Systems, Palo Alto, CA) with 0.3/0.8 mm focal spot size

-2 Epsilon High Frequency X-ray generators produced by EMD Technologies (Quebec, Canada).

-2 CCD based detectors with a Gd2O2S phosphor (XDI-VHR 2 Photonic Science, East Sussex, UK) with pixels of 22 microns which are typically binned to 88 microns.

The sampling is controlled by a sequencing application written in LabVIEW (National Instruments). We use pulsed x-rays (80 kVp, 70mA, 10 ms per exposure). The dual tube/detector configuration acquires images with a sampling rate of 20 projections/sec and projections from the two imaging chains can be combined in a single reconstructed volume[5].

Fig. 1. A dual tube/detector micro-CT system in a specimen rotate geometry.

III. CARDIAC MICRO-CT

4D Cardiac micro-CT requires the use of contrast agents, a gating strategy and appropriate post processing and image reconstruction. To provide the necessary blood/tissue contrast differences we use a liposomal **blood pool contrast agent** containing 100 mg I/ml and delivered via a tail vein catheter[6] in a dose of 0.2 ml/25 g mouse.

A. *Retrospective Gating*

To provide cardiac images, a gating strategy incorporating information on respiration and heart motion is required. For high throughput 4D imaging, we have selected retrospective gating. The respiratory and ECG signals are recorded together with the exposure time and projection angles and used post-sampling in a retrospective gating strategy (Fig.2). We use a nose cone through which the animals freely breathe a mixture of air and isoflurane for anesthesia. Therefore, we avoid the intubation/ventilation approach used previously ensuring easier, faster set-up and higher throughput. This facilitates longitudinal studies, where repeated endotracheal intubation may be traumatic to the animals. Respiration information is obtained by a pneumatic pillow coupled to a pressure transducer that detects chest motion. ECG pads placed on the paws of the animals are used to detect heart motion. The signals are processed using a Coulbourn Instruments (Allentown, PA) LabVIEW (National Instruments, Austin TX) system. We collect 1800 projections over ~ 90 seconds. Two scanning protocol are possible: a single slow rotation over only 94^o rotation (due to the two orthogonal imaging chains the complete 180^o + fan angle) or multiple fast rotations over 360^o . The total radiation dose for a study is approximately 0.2 Gy. Post-sampling, the recorded ECG signal is used to assign projection images to their respective phase in the cardiac cycle. The clustering process is implemented in MATLAB (The MathWorks, Natick, MA) and starts by detecting the R peaks in the ECG signal. Next, each projection is registered with the ECG signal by finding the closest two R peaks to the projection sampling time. Each R-R interval in the ECG cycle is divided by the number of temporal intervals (each equal to 10%). The respiration signal is used to eliminate those projections associated with large respiratory motion.

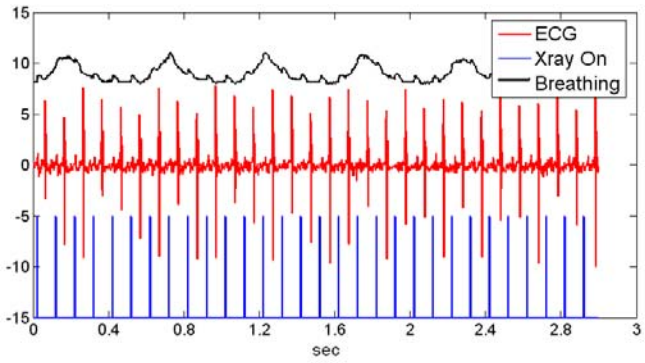


Fig. 2. Fast sampling is achieved via retrospective gating. The sampling sequence involves recording of the respiration, ECG signals as well as the sampling time and projection angles. Vertical lines display the sampling events. Data is acquired with a combined sampling rate of 20 projections/sec (due to the two imaging chains) during a continuous rotation.

B. *Reconstruction*

Sampling via retrospective gating challenges the traditional reconstruction. Both the irregular angular distribution of projections (represented as radial lines in Fourier space) associated with retrospective gating and their relatively low number (required to limit the dose) affect the image quality and introduce streaking artifacts and noise when using the analytical reconstruction algorithms, such as Feldkamp's[7]. In Fig.3, we illustrate this problem via simulations based on the dynamic Moby mouse phantom[8].

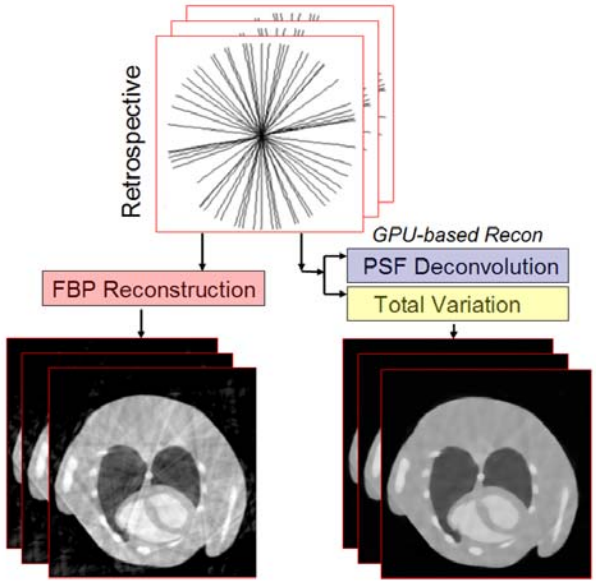


Fig.3. Retrospective gating is fast but its associated irregular angular distribution of projections creates streaking artifacts using FBP algorithms. Our solution is given either by a Point Spread Function (PSF) deconvolution or iterative CT algorithms using total variation implemented on GPU.

Our first approach is based on deconvolution with a **Point Spread Function** (PSF). It is generally accepted that 3D cone beam reconstruction can be approximated as a linear, shift-invariant process. Therefore, a reconstructed volume F can be expressed as a convolution in image space of the true volume T with the PSF P of the reconstruction algorithm. Mathematically, this is expressed as $T \otimes P = F$. In the Fourier domain, the convolution corresponds to multiplication: $\Im\{T\}\Im\{P\} = \Im\{F\}$. T can be obtained by dividing the Fourier transform of the original reconstructed volume by the Fourier transform of the PSF and then taking the inverse Fourier transform of the result: $T = \Im^{-1}\{\Im\{F\}/\Im\{P\}\}$. In practice, a PSF for a given algorithm is obtained by constructing a volume containing a single non-zero point at its center, simulating cone beam projections with the same sampling geometry and set of angles as the cardiac phase to be reconstructed, and reconstructing the point volume with the algorithm. The deconvolution approach works because the PSF volume is affected by the same artifacts as the reconstructed image. Some additional processing is needed in the Fourier domain to prevent singularities caused by the division. An example of a micro-CT images of a C57BL/6 mouse reconstructed via PSF deconvolution are shown by Fig. 4. The voxel size is 88 microns and the temporal resolution is 10 ms.

As an alternative to PSF deconvolution we use iterative algorithms for cone beam CT reconstruction based on sparseness prior. In particular, minimum **total variation** (TV) is implemented as the sparseness prior[9]. The TV algorithm is implemented using a sparseness transform. The 3D TV term of an image f, in this work is defined as follow:

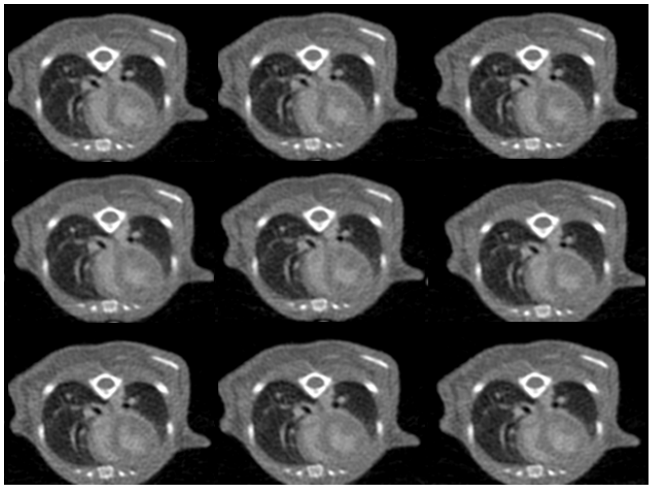


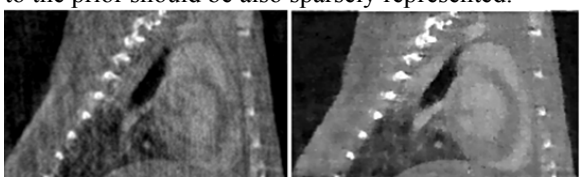
Fig.4. The same axial slice reconstructed during nine different cardiac phases. A PSF deconvolution approach was used. The images show no streaking artifacts.

$$\left\|\varphi(f)\right\|_{l_1} = \int |\nabla f| dxdydz = \int \sqrt{\nabla f_x^2 + \nabla f_y^2 + \nabla f_z^2}\, dxdydz$$

where φ takes the finite difference of the 3D image and || . ||l1 takes the l1 norm of the transform coefficients. The reconstruction problem is treated as an optimization problem and solved by minimizing a cost function that will contain two terms corresponding to: 1) data fidelity i.e. the measure of agreement between the reconstructed image and the projections acquired and 2) the TV-norm that will account for a solution with sparseness representation. The nonlinear conjugate gradient (CG) method is used to solve the optimization problem iteratively. We have also investigated the use of a prior provided by reconstruction using Feldamp's algorithm using all projections acquired retrospectively or a prospectively gated set, if available. This *fp* image prior displays low noise and high image quality, but contains no information on the particular cardiac phase that we wish to reconstruct. Formulated as a constrained optimization problem we solve for f:

$$\min[\mu_1 \left\|\varphi_1(f - f_p)\right\|_{l1} + \mu_2 \left\|\varphi_2(f)\right\|_{l1}] \ \textit{subject to} \ Af = b$$

where μ_1 , μ_2 are regularization factors, *A* is the projection matrix and *b* the projections. The cost function forces the selection of a solution *f* that should be sparse and its difference to the prior should be also sparsely represented.



(A) (B)

Fig.5. Comparison between FBP (A) and TV-CT (B) reconstructions using only 95 retrospectively- gated projections.

Fig.5 illustrates the effectiveness of the TV-CT (after 5 iterations) compared to a FBP reconstruction when only 95 projections with irregular angular distributions were used.

Both PSF deconvolution and TV-CT are computationally demanding. We address this problem by using a Graphics Processing Unit (GPU), which consists of several hundred processors capable of performing the same function in parallel. This architecture is ideal for performing backprojection, reprojection, and construction of the 3D TV gradient, in which the same memory and arithmetic operations must be performed over every voxel in the volume or every pixel in the projection images. Based on recent tests, the GPU enables acceleration of the bakprojection operation on the order of 300× compared to CPU implementations. To exploit this acceleration while minimizing transfers between disk and memory and remaining flexible in our choice of algorithms, we have built a client-server reconstruction framework. The server runs persistently in the background, maintains large arrays containing the projection and reconstruction data, and performs computationally demanding tasks on the GPU, as well as other arithmetic and utility functions on the CPU. The server receives instructions from clients via sockets, and the clients can be created dynamically from Matlab, allowing us to implement our reconstruction algorithms as Matlab functions. The server is written in C/C++ and the GPU is programmed with the Compute Unified Device Architecture (CUDA). The GPU used in this study is a GeForce GTX 285 (NVIDIA, Santa Clara, CA).

IV. PERFUSION MICRO-CT

Recently we have proposed an approach to measuring lung perfusion in rats using micro-CT[3] with a temporal sampling equivalent to one heart beat. We have based our approach on the paradigm that multiple carefully controlled injections of contrast agent create reproducible time attenuation curves. The method involved the use of a conventional radiographic contrast agent such as Isovue 370, (Bracco Diagnostics, NJ) delivered via a jugular vein catheter. Unfortunately this approach limits one's ability to perform longitudinal studies. We have addressed this limitation using a tail vein catheter for delivery of contrast agent and another novel sampling strategy. The sampling strategy is illustrated in Fig. 6. A single 360^o rotation is achieved in 22 sec and permits the acquisition of 440 projections (20 projs/sec). This projection data allows the reconstruction of 4 time points by using a different quadrant of projections from each tube/detector. This may be inadequate for accurate perfusion estimation. The number of time points (and therefore the temporal resolution) during first pass perfusion can be increased by using multiple injections/rotations. We use four (*m=4*) different injections and full rotations. The total volume of contrast agent is limited to 0.4 ml and is delivered using a LabVIEW controlled micro-injector as 4 injections (each 0.1 ml) separated by 2 minutes intervals to allow for clearance. Each rotation is started with an angular offset of $90^o/m$ i.e. 22.5^o in case of 4 injections and rotations. The temporal resolution is increased *N=4m* i.e. 16 times to 1.38 sec. Post sampling, data is selected from the complete set of 440x4=1760 projections and used for reconstruction using Feldkamp's algorithm. Note that the temporal resolution can be further increased if subsets of projections are used, but in such situations an iterative reconstruction such as TV-CT is preferred. Fig. 7 presents results from a lung perfusion study in a C57BL/6 mouse. Images were reconstructed with Feldkamp's algorithm at 88

microns voxel size and 1.38 s temporal resolution. The enhancement in a selected region of interest in the lung parenchyma represents the time attenuation curve and can be used to compute perfusion maps. The associated dose of such a study was less than 0.2 Gy.

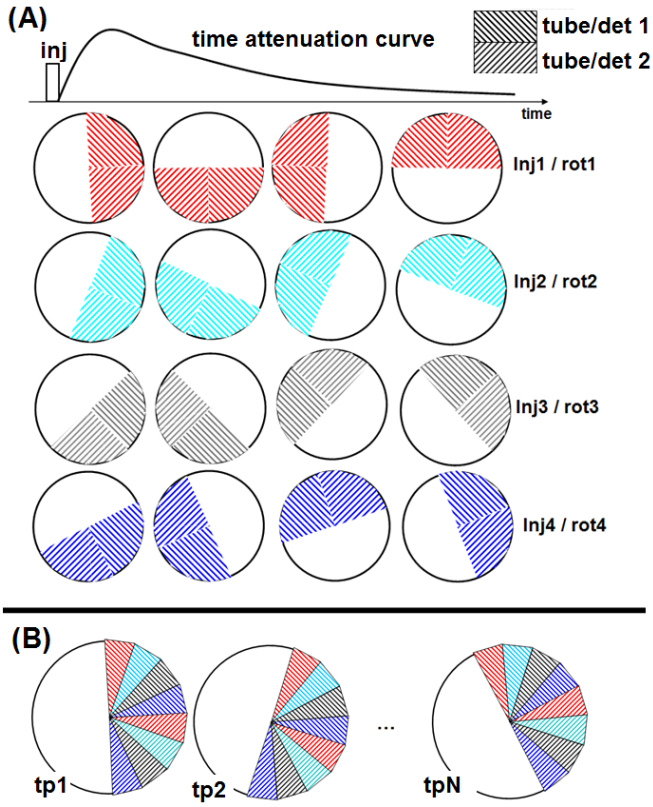


Fig.6. (A) The perfusion sampling involves the use of multiple injections and rotations ($m=4$). The first pass perfusion is represented by a generic time attenuation curve. The successive rotations involve starting with an angular offset of 90°/m. (B) The temporal resolution can be increased by combining projections from multiple acquisition. This is achieved by selecting sectors with projections corresponding to the same time point during the time attenuation curve.

V. Conclusion

These results demonstrate how micro-CT can be used not only for robust morphological but also functional imaging requiring 4D data. While our micro-CT system is in many ways unique, the sampling methods and software that we describe here can be used with commercial systems[2]. These novel new acquisition and reconstruction strategies show extraordinary promise for high throughput and longitudinal studies with immediate applications in a wide range of preclinical studies such as in cardiopulmonary safety or cancer research

Acknowledgment

All work was performed at the Duke Center for In Vivo Microscopy, NCRR National Biomedical Technology Research Center (P41 RR005959), with additional support from NCI (U24 CA092656 and KO8 CA114176). Liposomal contrast agent was provided by Ketan Ghaghada and Ananth Annapragrada, Univ. of Texas.

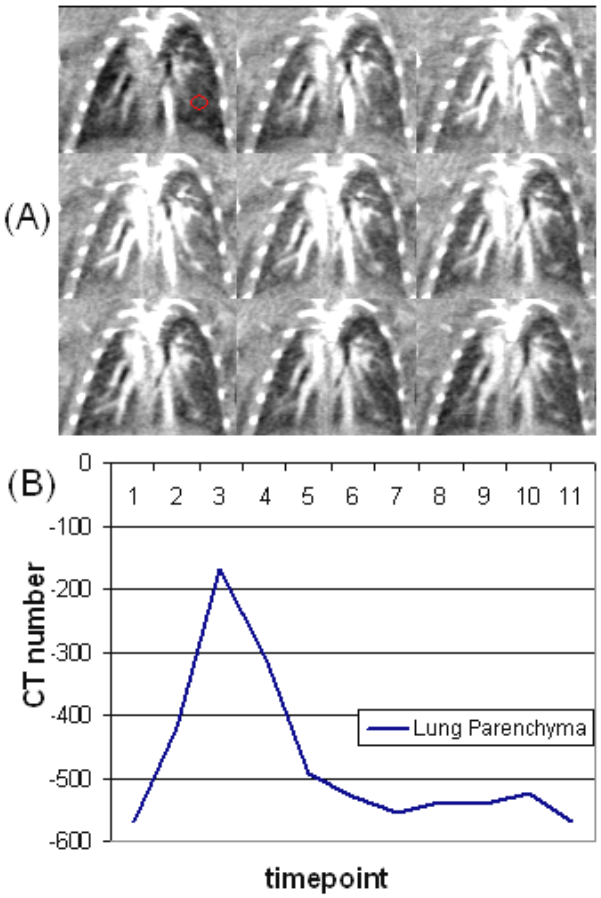


Fig.7. (A) The same coronal micro-CT slice over nine successive time points in the perfusion sequence. A lung parenchyma region of interest was selected (red circle) and the time attenuation curve is shown in (B).

References

[1] C. Badea, B. Fubara, L. Hedlund, and G. Johnson "4D micro-CT of the mouse heart," *Molecular Imaging,* vol. 4, pp. 110-6, 2005.

[2] M. Drangova, N. L. Ford, S. A. Detombe, A. R. Wheatley, and D. W. Holdsworth, "Fast Retrospectively Gated Quantitative Four-Dimensional (4D) Cardiac Micro Computed Tomography Imaging of Free-Breathing Mice," *Invest Radiol,* vol. 42, pp. 85-94, Feb 2007.

[3] C. Badea, S. M. Johnston, E. Subashi, Y. Qi, L. Hedlund, and G. Johnson, "Lung Perfusion Imaging in Small Animals using 4D Micro-CT at Heart Beat Temporal Resolution," *Medical Physics,* vol. 37, pp. 54–62, 2010.

[4] C. Badea, S. Johnston, B. Johnson, M. De Lin, L. W. Hedlund, and G. A. Johnson, "A Dual Micro-CT System for Small Animal Imaging " in *SPIE, Medical Imaging*, San Diego, CA, 2008, pp. 691342-10.

[5] S. Johnston, G. A. Johnson, and C. T. Badea, "Geometric Calibration for a Dual Tube/Detector Micro-CT System " *Medical Physics,* vol. 35, pp. 1820-1829 2008.

[6] S. Mukundan, K. Ghaghada, C. Badea, L. Hedlund, G. Johnson, J. Provenzale, R. Bellamkonda, and A. Annapragada, "A Nanoscale, Liposomal Contrast Agent for Preclincal MicroCT Imaging of the Mouse," *AJR,* vol. 186, pp. 300–307, 2006.

[7] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *J. Opt. Soc. Am.*, vol. 1, pp. 612-19, 1984.

[8] W. P. Segars, B. M. W. Tsui, E. C. Frey, G. A. Johnson, and S. S. Berr, "Development of a 4-D digital mouse phantom for molecular imaging research," *Molecular Imaging & Biology,* vol. 6, pp. 149-159, May-June 2004.

[9] J. Song, Q. H. Liu, G. A. Johnson, and C. T. Badea, "Sparseness prior based iterative image reconstruction for retrospectively gated cardiac micro-CT," *Med Phys,* vol. 34, pp. 4476-83, Nov 2007.

Total Variation regulated iterative algorithms for microCT

Xuan Liu[1], Michel Defrise[2], Alexander Sasov[1]

***Abstract*— We investigate the applicability of total variation (TV) regulated iterative algorithms for microCT. Two types of algorithms are investigated: the projection onto convex sets (POCS) type algorithms with minimization of TV, and maximum likelihood (ML) reconstruction with a TV penalty. The goal is to find a practical iterative algorithm for applications on microCT scanners when reduced scanning time and/or reduced radiation dose are preferred. We have done numerical simulations to study the properties of the algorithms. However, our focus is on studying the behavior of the algorithms when they are applied to real data. Our preliminary results from both simulations and real data indicate that the ML-type reconstruction gives better reconstruction in terms of robustness and noise property, but converges much slower than the POCS-type reconstruction.**

***Index Terms* — microCT, tomography, iterative reconstruction, total variation**

I. Introduction

Micro computerized tomography (microCT/nanoCT) technique is being used in an increasing range of applications such as biomedical, geology and material research. With a large array area detector, single circular orbit is the standard scanning geometry due to its simplicity and efficiency. For the same reasons, filtered back-projection algorithms, in particular, the Feldkamp cone-beam reconstruction (FDK)[1], are the algorithms of choice for reconstruction.

It is well known that iterative reconstruction algorithms may improve image quality due to possibilities of modeling the imaging system and data noise, and of incorporating prior knowledge on the object. The advantages over the FBP-type algorithms are much more obvious when the acquired data are coarsely sampled, incomplete, or distorted by physical phenomena such as scattering or beam-hardening effect, if these effects can be properly modeled. In emission tomography, e.g., Position Emission Tomography (PET), Single Photon Emission Tomography (SPECT), where the measurements have low photon counts, the iterative algorithms, especially the ML-type algorithms, are widely studied and validated as valuable reconstruction algorithms. In transmission X-ray CT, the iterative algorithms are not yet commonly known, though impressive progresses have been made in the last years [2]. Roughly, there are 2 major reasons. First of all, in most applications of microCT, the signal-to-noise ratios in the scans are much higher than those in an emission scan. Also the line integral model is a better approximation in CT. Therefore, FBP is usually quite effective. This is particularly true for *in vitro* scanning and material scanning, when radiation dose is not an issue. Secondly, the image array in high resolution cone-beam microCT is huge compared to PET or SPECT, in the order of 500^3~8000^3 voxels. This imposes an enormous challenge for iterative algorithms which require holding almost the entire image matrix in memory during forward-projection. The reconstruction time is also prohibitively long. Recent developments on iterative algorithms with prior image models based on total variation (TV) [3, 4] have shown impressive reconstructions from a small number of projections. In parallel, processing hardware such as Graphics Processing Units (GPU) is advancing rapidly and the investigations on using them to accelerate reconstruction are getting more and more popular [2]. These progresses have renewed the interest for iterative algorithms for microCT scanners.

We have implemented two types of iterative algorithms and evaluated them with both numerical simulations and real data. The first algorithm is the adaptive steepest descent POCS (ASD-POCS) algorithm by constrained TV minimization proposed by Sidky and Pan [3]. This method approximates image reconstruction as a finite linear system, and seeks a solution with minimum TV within the set of solutions satisfying a given data fidelity condition. The minimization of TV is achieved with a steepest descent algorithm with adaptive step size. The POCS step, which enforces the data fidelity, is achieved by the basic Algebraic Reconstruction Technique (ART) algorithm combined with the non-negativity condition. We refer to this algorithm as ART-TV. The underlying assumptions are that the system is underdetermined and that most objects are piece-wise constant and therefore the corresponding images have sparse gradient-magnitude images. These authors have demonstrated that the algorithm is capable of accurate reconstruction from few angular projections. The second algorithm is the maximum-likelihood (ML) algorithm of O'Sullivan and Benac [5]. We refer to this algorithm as ML or ML-TV if a TV penalty term is used. In this work, the polychromaticity of the X-ray spectrum and scattering has not been modeled.

The purpose of this investigation is to find a practical algorithm for microCT data. More specifically, we try to identify situations where iterative algorithms outperform the conventional FBP-type algorithms and are practical as far as the computer resources and reconstruction times are concerned. A few situations are in our mind: 1. Animal imaging, with the aim of reducing radiation dose for longitudinal studies when the animals are scanned repeatedly during a certain period, or in dynamic scans (respiratory, cardiac dynamic imaging) when signal-to-noise ratio may be low; 2. Samples which may change quickly due to temperature change or drying out, and for which short scans are preferred; 3. Low signal-to-noise scans, for

[1] SkyScan N.V., Kartuizersweg 3B, B-2550, Kontich, BELGIUM. E-mail: Xuan.Liu@skyscan.be.

[2] Dep. Of Nuclear Medicine, Vrije Universiteit Brussel, Brussels, BELGIUM. E-mail: mdefrise@vub.ac.be

example, ultra-high-resolution scans (<1 μm) when the power of x-ray source is limited. Reconstructions of both numerical simulations and real scans are presented.

II. Reconstruction algorithms

A. Imaging system and notations

We consider a cone-beam CT system with a single circular orbit. All scanners in this study are capable of a full 360-degree scan. Each system contains an X-ray source and a 2D camera consisting of **U**x**V** imaging elements, where **U** and **V** are in order of 512…4000. The imaging system rotates during a scan in a step-and-shoot mode, yielding **P** projection images, one at each angular position.

We denote the detected X-ray intensity at each imaging element as $\mathbf{y_j}$, j = 1,…,**D**, where **D**=**P**x**U**x**V** is the total number of elements (also referred to as lines of response (LOR's)).

We denote the non-attenuated X-ray intensity in each LOR as $\mathbf{I_j}$. This is more often referred as air intensity. The values are obtained from the LOR's at detector edges where the X-rays do not pass through the object.

We denote the linear attenuation coefficient in voxel i as $\mathbf{x_i}$, i=1, …, **N**, where **N**≈**U**x**U**x**V** is the number of voxels. $\mathbf{x_i}$'s are the unknowns a reconstruction algorithm seeks.

The linear imaging system considered by the POCS-type reconstruction can be described as:

$$\sum_{i=1}^{N} s_{j,i} x_i = \ln(I_j / y_j) \tag{1}$$

where <$\mathbf{s_{j,i}}$> denotes the system matrix.

The ML-type reconstruction assumes that the measurements are Poisson variables. The mean value in LOR j is

$$\langle y_j \rangle = I_j e^{-\sum_{i=1}^{N} s_{j,i} x_i} \tag{2}$$

The summation in eq. (1) and (2) is the projection operator. Due to the large data size, it is impossible to pre-calculate the system matrix. Instead, the $s_{j,i}$ are calculated on the fly. We use a ray-driven method for projection [8]. For each ray, the coefficients are cached in the memory during projection and are used later in the back-projection process.

The total variation of an image **x** is defined as:

$$TV(x) = \sqrt{\|\nabla x\|^2 + \varepsilon_{\min}} \tag{3}$$

The term ε_{min} ($=10^{-8}$) avoids singularity when the gradients of TV are calculated. We use a 2-point discretization of each component of the gradient.

B. Constrained, TV-minimization reconstruction with POCS

A detailed description of this algorithm can be found in [3]. Only an outline and some implementation details are given here. The solution of the linear system described by eq. (1) is obtained by minimizing the total variation of the image subject to 2 constraints: non-negativity constraint and data consistency condition (eq. (4)).

$$\sqrt{\sum_{j=1}^{D} \left(\sum_{i=1}^{N} s_{j,i} x_i - \ln(I_j / y_j)\right)^2} \le \varepsilon \tag{4}$$

where ε is a upper bound on the error defined as the square root of the differences between estimated data and the measurements. This inequality implies that data inconsistencies due e.g. to a simplified data model and noise, are tolerated to a certain level ε. A non-negative solution satisfying (4) is obtained by the ART algorithm. The ART updating is done per LOR, and the procedure of updating over all LOR's is one iteration. For LOR j, the updating of image voxel i is given by

$$x_i^{k+1} = x_i^k + \beta \frac{s_{j,i}}{\sum_{i'=1,N} s_{j,i'}^2} (\ln I_j / y_j - \sum_{i'=1}^{N} s_{j,i'} x_{i'}^k) \tag{5}$$

where $\beta \in (0,1]$ is the relaxation parameter, which is reduced by a factor $\beta_{red} \in (0,1]$ after each iteration. The minimization of TV is done using steepest-descent algorithm with adaptive step size. A drawback of this algorithm is the optimization of the 6 parameters in use.

Alternatively, ART can be replaced by the simultaneous ART (SART) where the image is updated per projection instead of per ray in ART.

C. Maximum likelihood reconstruction with TV penalty

In maximum-likelihood approach, the measured X-ray intensities are assumed to have Poisson distribution. Note that this Poisson model is approximate because the projection data usually undergo various corrections first. Based on this assumption, the log-likelihood of a dataset is given by:

$$\Phi(x) = \sum y_j \log(< y_j >) - < y_j > - \log(y_j!) \tag{6}$$

A maximum-likelihood reconstruction is a solution that maximizes the log-likelihood objective function Φ(x). We use the maximum-likelihood (ML) algorithm of O'Sullivan and Benac [5]. A TV penalty term can be added to form a new objective function Ψ(x)= Φ(x)- γ' TV(x). Maximization of Ψ(x) can be achieved using the one-step-late [6] modification of the ML algorithm. The updating equation is given by:

$$x_i^{k+1} = x_i^k - \frac{1}{Z_i} \log \frac{b_i + \gamma' \left.\frac{\partial TV(x)}{\partial x_i}\right|_{x=x^k}}{\sum_{j=1}^{D} s_{j,i} I_j e^{-\sum_{l=1}^{N} s_{j,l} x_l^k}} \tag{7}$$

where $z_i \approx \left(\sum_{i=1}^{N} s_{j,i}\right)_{\max}$, and $b_i = \sum_{j=1}^{D} s_{j,i} y_j$ is the back-projection of the measurements and can be pre-calculated. The regulating parameter $\gamma'=\gamma.\max(b_i)$, where γ is usually very small, in the order of 10^{-3}~10^{-5}. Non-negativity condition is applied after each updating. We use a uniform zero-image as starting image. In this algorithm, the forward-projection is ray-driven and the back-projection is voxel-driven. Acceleration using ordered subsets(OS) was applied with a rule of thumb that each subset should contain at least 30 projections.

III. 3D Simulation

We have done 3D numerical simulation of the 3D Shepp-Logan phantom using the geometry of the *in vivo* scanner SkyScan-1076. To save computation time we

down-sampled the projections from 1Kx1K to 256x256 by averaging, and used a large rotation step of 10 degrees.

Fig. 1 shows one slice of the reconstructions of a simulated noise-free 3D Shepp-Logan phantom using different algorithms. Other simulations have also been performed with Poisson noise, more views and fewer views down to 18 (results not shown). The ML-TV algorithm gives the most accurate reconstruction after 2000 iterations. As also pointed out by other authors, we observe that ART-TV and SART-TV reach a good reconstruction after about 100 iterations, while ML and ML-TV yield a reasonable reconstruction only after more than 1000 iterations. The TV penalty reduces noise and streaking artifacts effectively.

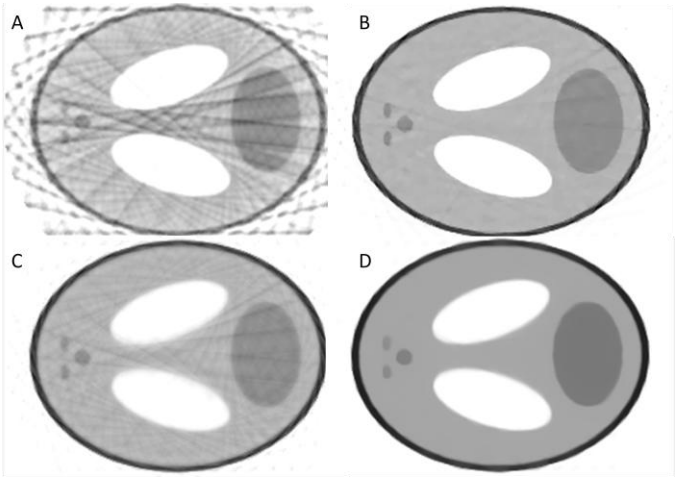


Fig.1. Reconstructions of simulated 3D Shepp-Logan phantom: 36 projections over 360 degrees. A. FDK; B. SART-TV (99 iter); C: ML(2000 iter, 1 subset); D. ML-TV (2000 iter, 1 subset, γ=0.0001).

IV. Results on real data

SkyScan produces several types of microCT and nanoCT scanners [7], each of them can operate at different resolutions. Due to the large memory requirement and time consuming nature of the iterative algorithms, we have restricted us to applications with data up to 1Kx1K.

Two examples are given. One is a small piece of a car catalytic converter with a regular pattern of very fine structures. Another is an adult mouse scanned while respiratory motion signals were recorded for retrospective synchronization [9]: projections were sorted offline into 4 respiratory phases. Each phase is less affected by motion blurring; however, the noise level is higher. To study the behavior of the algorithms with coarser angular sampling, the scans were also down-sampled in angular direction.

During acquisition, the projections were processed for geometrical distortion correction and flat field correction.

A. Sample scan example

The catalytic converter was scanned on a SkyScan-1172 scanner: 275 views, 0.7-degree step, projection size of 1000x524 pixels, Al+Cu filter was used to reduce beam-hardening. Each projection is the average of 20 frames of 474 ms each, resulting in a good signal-to-noise ratio. The reconstructed volume contains 524x1000x1000 voxels. Ring artifacts correction has been applied before reconstruction.

The top row in fig. 2 shows reconstructions of the original data set (275 views). FDK is quite effective and the reconstruction is almost artifact-free. ART-TV gives comparable results as FDK after just a few iterations. The ML algorithm also gives comparable image after 20 iterations (11 subsets). With TV penalty, ML reconstruction is somewhat blocky (image not shown).

The bottom row in fig. 2 shows reconstructions from only 19 projections extracted from the original dataset. As can be seen, ART-TV and ML resolve far more details than FDK but nevertheless do not allow recovering the resolution obtained with the 275-view scan. The ML-TV image becomes blocky without resolving more details and a similar blocky look is also observed with ART-TV when the iteration exceeds a certain stage.

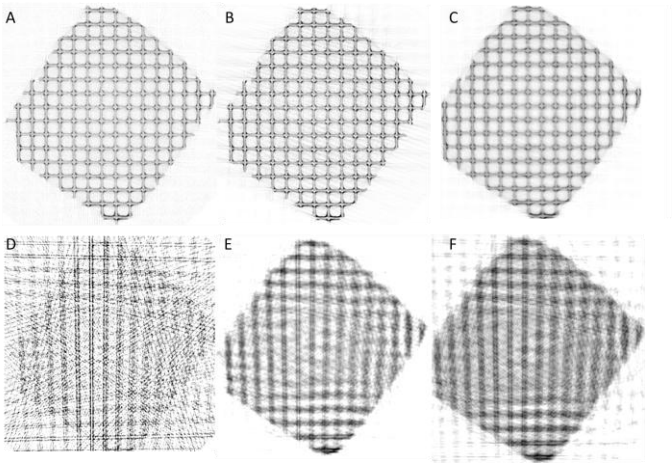


Fig.2. Reconstructions of the catalytic converter sample. Top row: 275 views.; Bottom row: 19 views. A. FDK; B. ART-TV(19 iter.); C. ML (21 iter., 11 subsets); D. FDK; E. ART-TV (19 iter); F. ML (101 iter, 1 subset)
.

B. Animal scan example

The second example is one phase of a 4D scan with retrospective synchronization [9]. An adult mouse was scanned on a SkyScan-1076 scanner: 451 views with 0.8-degree step; projection size 1000x524; 0.5mm Al filter; 8 frames with 158ms exposure time per angular position. The total scan duration was 16 minutes. After sorting into 4 time bins, each bin contains ~2 frames per view. This scan is relatively noisy, and it is still affected by motion despite the synchronization. The reconstructed volume is 524x1000x1000.

The reconstructions using FDK, ART-TV and ML are shown in fig. 3. Compared to FDK, ART-TV seems to recover part of the blurring artifacts (ribs), but the soft tissues (blood vessels in the lung, heart) are not well distinguishable. The ML algorithm converges slowly but the soft tissues are better resolved than FDK, while the bone structures are not as sharp as with FDK.

A subset of the dataset with 76-views was obtained. The reconstructions are shown in fig. 4. ART-TV has the poorest image quality for low-contrast soft tissues. The soft tissues are

distinguished much better with ML and ML-TV (not shown) than the other 2 algorithms.

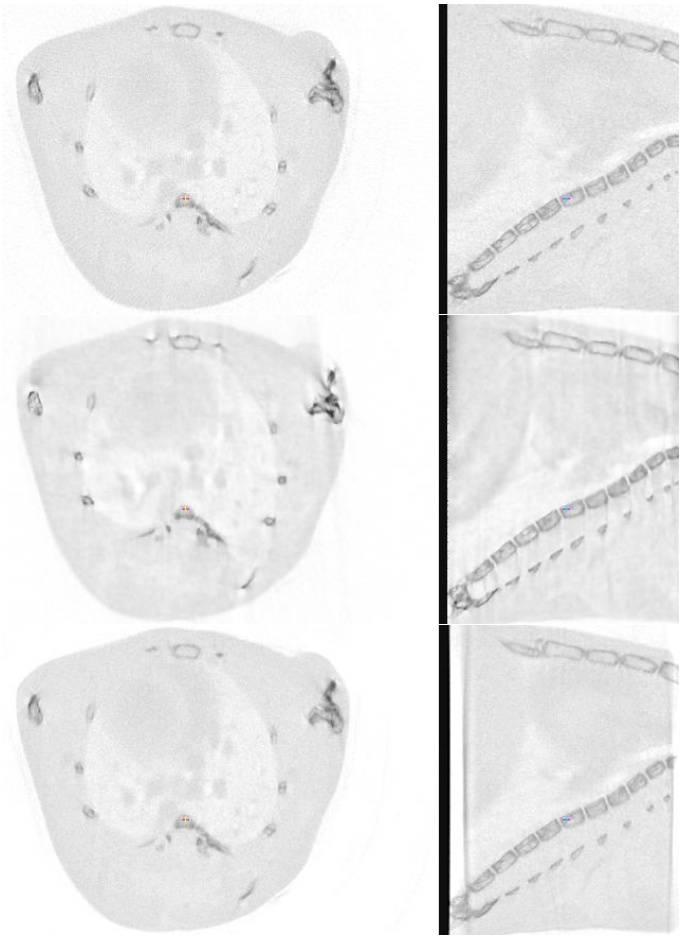


Fig.3. 451-view reconstruction. Left: transaxial slice; right: sagittal view(rotated by 90°). Top: FDK; middle: ART-TV(16 iter); bottom:ML (21 iter, 15 subsets).

V. Discussions

We have investigated the potential benefits of 2 iterative algorithms for the SkyScan microCT and nanoCT scanners. Though numerical simulations show remarkable improvements when using these algorithms instead of the standard FDK algorithm, these advantages are not always confirmed with real data. This might be due partly to the fact that the scatter and the polychromatic X-ray spectrum have not been modeled in our study.

For data with sufficient sampling and low noise level, FDK remains a robust, efficient and effective algorithm for most applications. ART-TV gives a good reconstruction with just a few iterations, but even with the TV minimization streaking artifacts and salt-and-pepper noise are still hard to get rid of completely. This might require further optimization of the 6 parameters used in this algorithm. Low-contrast objects, such as soft tissues, are not very well resolved with ART-TV. The ML algorithm is easier to use in terms of parameter optimization and generates better reconstructions especially for low-contrast objects and *in vivo* scans. However, it converges slowly. The ordered-subset method does accelerate convergence but cannot be applied when the dataset only contains a few views. The TV penalty used in both algorithms does reduce streaks and noise, but tends to generate blocky images, which might in practice be more misleading than a noisy but unbiased FDK image.

In both algorithms, more accurate forward/back-projectors are worth investigation. In ART-TV or SART-TV, this may reduce the streaking artifacts caused by an inaccurate projection model. In ML, when a ray-driven back-projector was used, oscillations could be observed beyond certain iteration stage. This problem can be avoided by a voxel-driven back-projector which is more accurate.

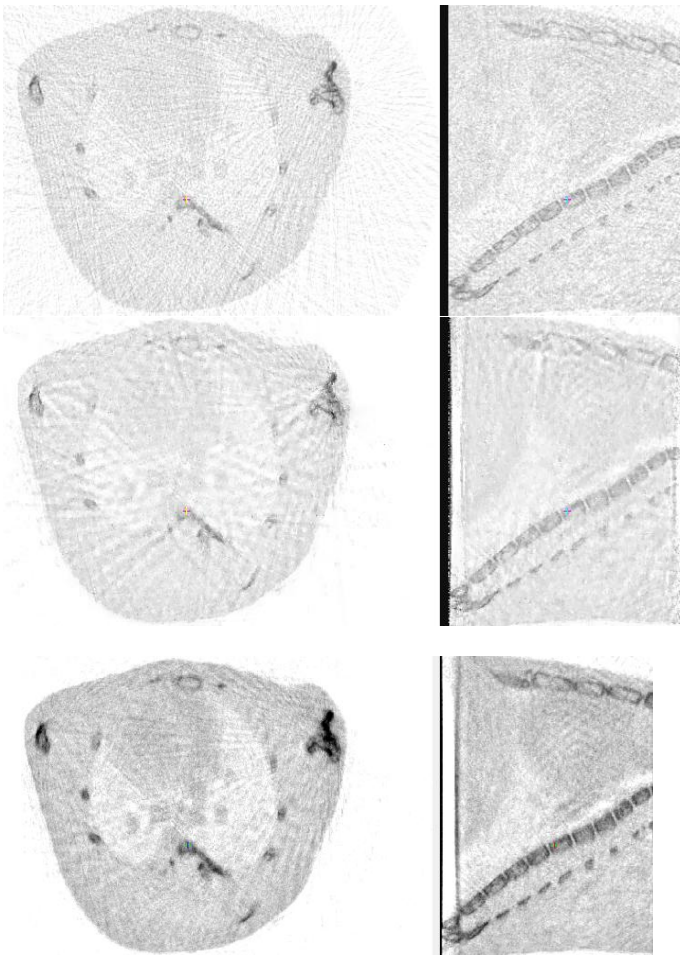


Fig.4. 76-view reconstruction. Top: FDK; middle: ART-TV(89 iter); bottom: ML-TV(181 iter, 2 subsets)

References

[1] L.A. Feldkamp, L.C. Davis, J.W. Kress, *J. Opt. Soc. Am. A*, 1, 6 (1984), pp.612-619.
[2] http://www.fully3d2009.org/download/proceedings_2009.pdf
[3] E.Y. Sidky and X. Pan, *Phys. Med. Biol.*, vol. 53, pp. 4777-4807, 2008.
[4] V. Y. Panin et al, *IEEE Trans. Nucl. Sc*. 46, 2202-2210, 1999.
[5] O'Sullivan and Benac, *IEEE Trans Med Imag*, vol 26, p. 283, 2007.
[6] P. Green, , *IEEE Trans. Med. Imag.*, vol. 9, pp. 84-93, 1990.
[7] http://www.skyscan.be/products/all_products.htm
[8] Joseph P, *IEEE Trans. Med. Imaging* 1 192-6, 1983.
[9] X. Liu, F. Nadeem, P.L. Salmon, A. Sasov, *Developments in X-Ray Tomography VI* , SPIE vol.7078, 707809, 2008.

A 3-DIMENSIONAL DISCRETE TOMOGRAPHY APPROACH FOR SUPERRESOLUTION OF MICRO-CT IMAGES: APPLICATION TO FOAMS

Wim van Aarle[*], *Gert Van Gompel*[†], *K. Joost Batenburg*[*], *Elke Van de Casteele, Jan Sijbers*[*]

[*]IBBT-Visionlab, University of Antwerp, Antwerp, Belgium
[†] Dept. of Radiology, UZ Brussels, Brussels, Belgium

ABSTRACT

In μCT imaging, structures that are small w.r.t the image resolution are hard to segment because of partial volume effects. In this paper, we show that if the object consists of a small number of different materials, each having a constant density, the resolution of the tomographic reconstructions can be dramatically improved by using prior information on the grey values of the scanned objects, resulting in much more accurate segmentations. The proposed method is based on an upsampling of the reconstruction grid, combined with the discrete algebraic reconstruction technique (DART) [1], in which the scanned object is assumed to be composed of homogeneous materials. Experiments on simulated CT data of foams show that the proposed method indeed generates significantly better segmentations compared to conventional methods.

Index Terms— μCT, computed tomography, superresolution, discrete tomography, metal foams

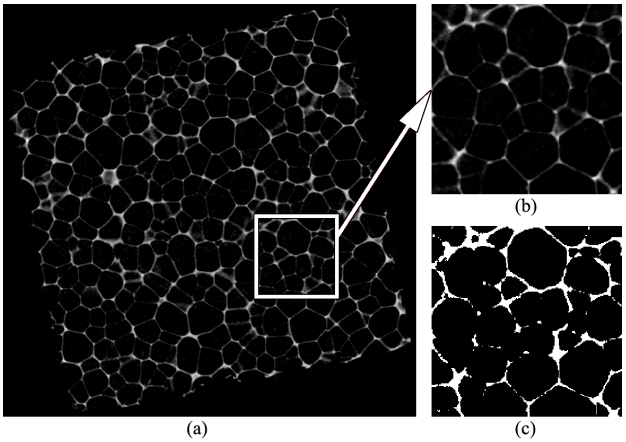


Fig. 1. Reconstruction of a polyurethane foam.
(a) FBP reconstruction (b) zoomed in FBP reconstruction
(c) Otsu segmentation of (b)

1. INTRODUCTION

High quality μCT reconstructions of very small structures, such as bone trabeculae or cell edges in porous materials, often prove difficult to obtain. The use of a high resolution detector to increase CT acquisition quality, requires longer scanning times and a higher radiation dose, risking damage to the object. Alternatively, zooming in on a specific region of interest in the object as to increase the spatial resolution of that region has the drawback of projection truncation, leading to cupping artifacts in the resulting reconstructions.

In our approach, the aim is to reconstruct low resolution μCT images on an upsampled, high resolution grid. As upsampling makes the problem highly underdetermined, the discrete algebraic reconstruction technique (DART) [1] is used to include prior knowledge to resolve the non-uniqueness of the solution. The DART technique is an iterative reconstruction method that assumes the objects to be piecewise uniform with known densities. An additional advantage of DART is that it results in a segmented image. Indeed, DART aims to find the piecewise constant image that minimizes the projection distance (i.e. the difference between the measured data and the Radon transform of the reconstructed image). This is essentially different from finding the optimal segmentation of a continuous reconstruction from the dataset.

An important application of μCT imaging lies in the field of materials science where foam objects often need to be characterized. Using segmented X-ray tomographic reconstructions, various structural parameters (such as pore size distribution and pore interconnectivity) and mechanical parameters (such as stiffness and strength) can be estimated non-destructively [2]. It is, however, critical that these segmentations are of an adequate resolution. Fig. 1(a) shows a typical FBP reconstruction of a polyurethane foam taken with a SkyScan 1172 μCT scanner at a pixel resolution of 17μm. The zoomed-in image in Fig. 1(b) shows that the edges of the pores are very thin compared to the resolution of the image. Indeed, the thickness of some edges is even smaller than a single pixel. This leads to partial volume effects, i.e. a pixel might belong in part to one partition and in part to another, making image segmentation a very daunting task. Fig. 1(c) shows the segmentation of this reconstruction created by se-

This work was financed by IBBT Flanders. K. J. Batenburg also acknowledges the financial support of FWO, Flanders.

lecting a global threshold using the well-known method of Otsu [3]. It is clear that very small structures are not found and that some are too thick. This will result in inaccurate analysis of the foam object.

In this paper, the proposed discrete tomography technique for the reconstruction on an upsampled grid is applied to simulated CT datasets of metal foams as to increase the effective spatial resolution of the reconstructions and segmentations.

The paper is organized as follows: Section 2 explains the theory behind our approach. Section 3 describes the experimental setup and presents reconstruction results on a two dimensional image. Section 4 then discusses how the proposed algorithm can be extended to three-dimensional volumes. Finally, conclusions are drawn in Section 5.

2. SUPERRESOLUTION BY DISCRETE TOMOGRAPHY

Consider a sinogram consisting of M projections with N radial samples at sampling distance $\Delta s = 1$. Such a dataset is typically reconstructed on an $N \times N$ grid, with sampling distance $\Delta t = \Delta s = 1$ (see Fig. 2(a)).
Algebraic reconstruction methods consider tomographic reconstruction as the problem of solving a system of linear equations

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{p} \tag{1}$$

where $\boldsymbol{x} = (x_j)$ are the unknown attenuation values on the grid in image domain, $\boldsymbol{p} = (p_i)$ are the measured projection values, and $\boldsymbol{A} = (a_{i,j})$ is the linear projection operator. Consider an ideal experiment, without noise, where a sufficient number of projections is available to ensure that the system in Eq. (1) has a unique solution. Now assume that the uniform density of the object is known in advance. The combination of the complete dataset and the prior knowledge then represents an overdetermined reconstruction problem. Hence, parts of the projection data contain redundant information. To optimally exploit this redundant data, we transform the reconstruction problem on the $N \times N$ grid to a limited data reconstruction problem on an $aN \times aN$ grid with sampling distance $\Delta\tilde{t} = 1/a$ for some integer $a > 1$. Note that the sampling distance of the refined grid is now smaller than the sampling distance of the detector pixels. Each detector pixel can be seen as the sum of contributions of a subdetector pixels with sampling distance $\Delta\tilde{s} = \Delta s/a = 1/a$ which corresponds to the sampling distance of the refined grid (see Fig. 2(b)). Compared to the system in Eq. (1), the system

$$\tilde{\boldsymbol{A}}\boldsymbol{y} = \boldsymbol{p} \tag{2}$$

corresponding to the upsampled reconstruction grid has the same number of equations, while the number of unknowns has increased by a factor a^2.

To solve the system in Eq. (2) under the constraint that each of the y_i can only take values in a prescribed set $R = \{\rho_1, \ldots, \rho_k\}$, we apply DART. Here, we only provide a short overview of this iterative heuristic algorithm and refer to [1] for more details. First, a conventional reconstruction is computed. Then, a number of DART iterations are performed, each containing the following steps:

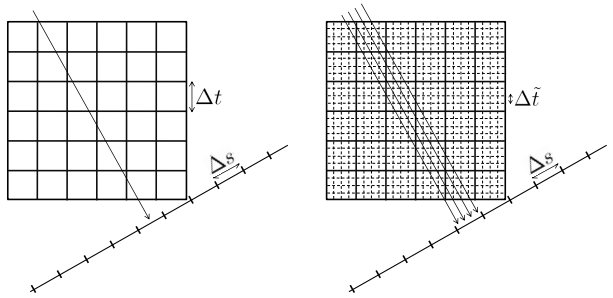
1. The current reconstruction is subdivided into partitions by thresholding and each of the partitions is assigned the known grey level value.

2. The difference between the actual projection data and the forward projection of the segmentation is computed.

3. Pixels that are on the boundary of a partition are identified.

4. A regular iterative technique such as SIRT [4] is applied to reconstruct the projection difference in the pixels on the boundaries. The other pixels are kept fixed at the grey level of their partition.

The motivation for the DART-approach is that segmentation errors, in general, occur near the edges of the different partitions. By assuming that the pixels in the center of a partition are already perfectly reconstructed, the number of unknowns in Eq. (2) drops while the number of equations remains the same, making it again uniquely solvable.



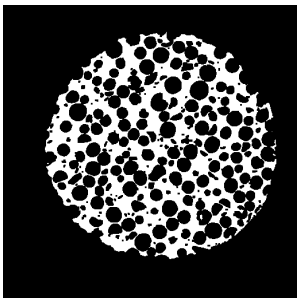
(a) coarse grid (b) fine grid

Fig. 2. (a) reconstruction grid used for conventional CT. (b) reconstruction grid in the proposed discrete tomography approach

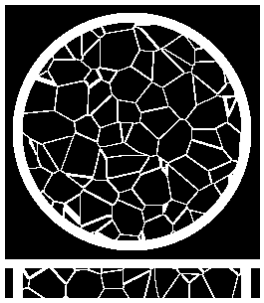
3. 2D SIMULATION EXPERIMENT

To demonstrate the proposed method, a 512×512 pixel phantom image was simulated, shown in Fig. 3(a). This binary image is based on an FBP reconstruction of a porous material in a real CT-scanner and contains many small holes and very thin edges that are difficult to segment.

A Poisson distributed, parallel beam sinogram image was computed with $M = 360$ projection angles, equiangularly



(a) Phantom A, 512×512 pixels, used in section 3



(b) Phantom B, $256 \times 256 \times 30$ pixels, used in section 4, up: single slice, bottom: vertical cut

Fig. 3. Simulated phantoms of foams.

distributed between $[0, \pi)$ and 512 detector pixels. The partial volume effect was then simulated by downsampling each projection with a factor $d = 4$, i.e. by summing the radial bins 4 by 4. This sinogram, with $N = 128$ detector pixels, was then used in the following experiments:

(a) Standard FBP reconstruction with Otsu's segmentation method [3] on the $N \times N$ grid.

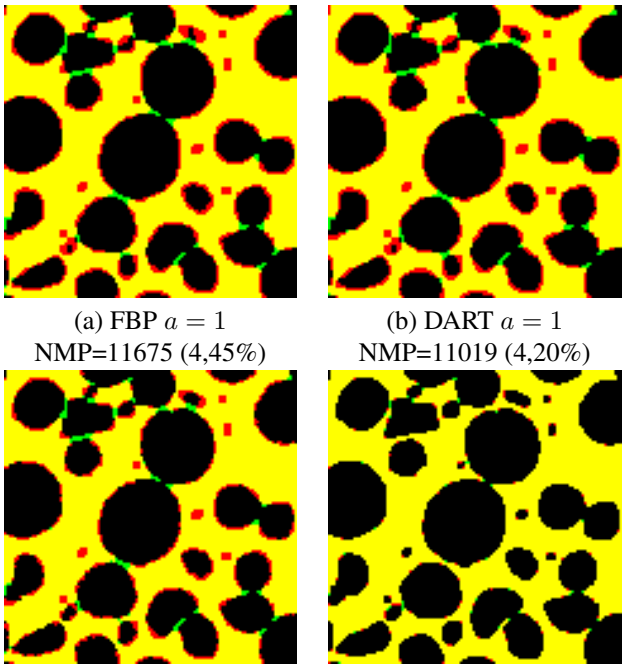
(b) DART reconstruction on the $N \times N$ grid.

(c) FBP reconstruction on a refined $aN \times aN$ grid (with $a = 4$) from an upsampled $aN \times M$ sinogram. The upsampling of the sinogram is performed in the radial direction by 1D linear interpolation such that each projection consists of aN pixels with pixel width $\Delta t = 1/a$ and $a > 1$ an integer.

(d) DART reconstruction on the refined $aN \times aN$ grid (with $a = 4$) from the original $N \times M$ sinogram according to the approach explained in Section 2.

To quantify the reconstruction quality, the number of misclassified pixels (NMP) w.r.t. the ground truth image in Fig. 3(a) was computed. Since the experiments were performed at varying resolutions, the reconstructions were rescaled to the size of the ground truth images using bilinear interpolation.

Fig. 4 (a-d) depict the above mentioned reconstruction experiments (a-d). In each subfigure, the reconstruction is displayed in red and the ground truth image in green. Where both images overlap, the corresponding pixel is colored in yellow. In our experiments, reconstruction on a larger grid is clearly advantageous. The improvement, however, is more profound when using DART with grid upsampling than when using FBP with sinogram upsampling. Using the method of Section 2, the edges of the structure are reconstructed more accurately and nearly all small holes are found.



(a) FBP $a = 1$
NMP=11675 (4,45%)

(b) DART $a = 1$
NMP=11019 (4,20%)

(c) FBP $a = 4$
NMP=9726 (3,71%)

(d) DART $a = 4$
NMP=2393 (0,91%)

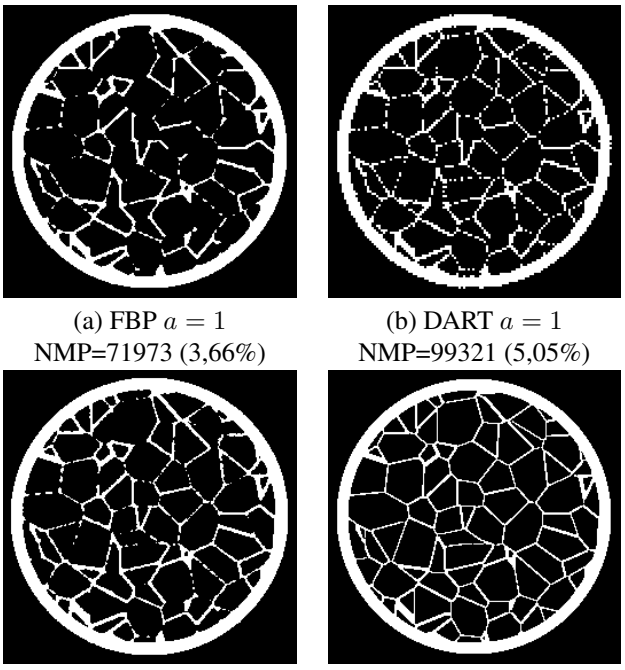
Fig. 4. Thresholded FBP and DART reconstructions of Phantom A, with ($a = 4$) and without ($a = 1$) upsampling.

4. 3D SIMULATION EXPERIMENT

The update operations in the DART algorithm involve pixels at the object boundary. In classical reconstruction techniques for parallel beam and fan beam data, each slice of a 3D volume is reconstructed independently. For DART, however, such a strategy does not exploit the boundary concept in the third dimension. Therefore, one can expect that using a 3D connectivity window for determining the boundary pixels leads to improved accuracy compared to a purely 2D approach. For 2D DART reconstructions, we use a 8-connectivity window to determine the boundary pixels, i.e. a pixel is considered on the boundary if any of its 8 neighbours belong to a different partition. For 3D DART reconstructions, we have found that it is best to use 6-connectivity, i.e. for each dimension only the two most adjacent ones are considered.

To demonstrate the superresolution technique in 3D, a $256 \times 256 \times 30$ phantom image was generated, representing a polyurethane foam (Fig. 3(b)). In these kind of foams, it is often important to detect fractured cell walls. The volume was created by computing a Voronoi diagram on a set of randomly chosen points, which ensures that all walls are closed (unfractured).

For each slice of the phantom volume, a sinogram was created using 256 detector pixels and $M = 180$ projection angles, equiangularly distributed between $[0, \pi)$. These were



(a) FBP $a = 1$
NMP=71973 (3,66%)

(b) DART $a = 1$
NMP=99321 (5,05%)

(c) FBP $a = 4$
NMP=89902 (4,57%)

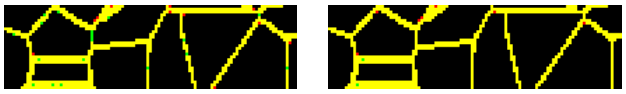
(d) DART $a = 4$
NMP=3268 (0,17%)

Fig. 5. Thresholded FBP and DART reconstructions of Phantom B, with ($a = 4$) and without ($a = 1$) upsampling.

then downsampled with a factor $d = 2$ ($N = 128$), resulting in a $128 \times 180 \times 30$ sinogram volume. Note that this creates an image resolution that is lower than the thickness of the cell walls.

Experiments (a-d) from Section 3 were repeated for Phantom B with the 3D version of DART. For each reconstruction, Fig. 5 depicts a single slice of the volume (the same one as shown in Fig. 3(b)). Note in Fig. 5(b) that the DART reconstruction on an $N \times N$ grid does not necessarily improve the reconstruction accuracy compared to the FBP reconstruction for $a = 1$. This is because DART suffers particularly from partial volume effects, as no intermediate grey levels are allowed in the binary reconstruction. Also note in Fig. 5(c) that using the sinogram upsampling method not always improves the reconstruction accuracy. This is caused by the incorrect assumption that the projection data is a continuous curve. It is, however, clear that in Fig. 5(d) a drastic improvement in reconstruction quality is obtained. Indeed, where the other reconstructions all show fractured cell walls, the proposed method accurately recovers most cell walls.

To demonstrate the effect of using the 3D connectivity window, a slice-by-slice DART reconstruction volume was also computed. Fig. 6 shows a vertical cut through the different slices of the volume. Although both methods generate very good reconstructions compared to the conventional methods, employing the 3D boundary concept lowers the number of misclassified pixels even more.



(a) Slice-by-slice DART
NMP=6198 (0,32%)

(b) 3D DART
NMP=3268 (0,17%)

Fig. 6. Vertical cut through the slices of DART reconstructions ($a = 4$) of Phantom B.

5. CONCLUSIONS

To improve the detection of small structures in CT reconstructions, we proposed a method that increases the spatial resolution without increasing the detector resolution. The method assumes that the object is composed of homogeneous materials and is piecewise constant. This prior knowledge is exploited by reconstructing the object on an upsampled grid using the discrete algebraic reconstruction technique DART, effectively increasing the spatial resolution without increasing the detector resolution. Our experiments on simulated data have shown that the proposed method generates reconstructions with significantly more detail compared to conventional reconstruction algorithms. We extended this method to parallel-beam 3D, as to include the prior knowledge of piecewise uniformity of the object in the third direction as well, thereby achieving better reconstructions than a slice-by-slice approach. In future work we will further validate the proposed method on real data.

6. REFERENCES

[1] K.J. Batenburg and J. Sijbers, "DART: a fast heuristic algebraic reconstruction algorithm for discrete tomography," *Proceedings of ICIP*, vol. 4, pp. 133–136, 2007.

[2] A. Elmoutaouakkil, L. Salvo, and E. Maire, "2D and 3D characterization of metal foams using X-ray tomography," *Advanced Engineering Materials*, vol. 4, no. 10, pp. 803–807, 2002.

[3] N. Otsu, "A threshold selection method from gray level histograms," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 9, pp. 62–66, 1979.

[4] J. Gregor and T. Benson, "Computational analysis and improvement of SIRT," *IEEE Transactions on Medical Imaging*, vol. 27, no. 7, pp. 918–924, 2008.

Semi-automatic misalignment correction for a small animal micro-CT scanner with variable geometry

Daniele Panetta[1], Nicola Belcari[2], Alberto Del Guerra[2] and Piero A. Salvadori[1]

[1]Department of PET and Radiopharmaceutical Chemistry, CNR Insitute of Clinical Physiology, Pisa, Italy
[2]Department of Physics "E. Fermi", University of Pisa and Center of Excellence AmbiSEN, Pisa, Italy

***Abstract*—An engineered micro-CT scanner for *in vivo* imaging of small animals (*Xalt*, which stands for *X*-ray *A*nima*L* *T*omograph) has been recently built at the Institute of Clinical Physiology of the National Research Council (IFC-CNR, Pisa, Italy), in partnership with the University of Pisa. The Xalt scanner is a cone beam micro-CT based on a 2-dimensional flat-panel detector and microfocus x-ray source. The magnification and detector orientation can be manually varied to select the spatial resolution, field of view (FoV) and sensitivity that match the needs of several different imaging protocols for mice, rats and test tubes. Even with high precision mechanical assembly, recalibration for geometric misalignments is necessary after each modification of the scanner geometry. To avoid the burden of geometric recalibration, a semi-automatic cone beam calibration method has been applied and integrated in the reconstruction software. The calibration method does not require measurements on specific phantoms, as it is based on the recovery of redundancies in the projection data of generic objects that are lost when the scanner components are out of alignment. By using this method, our variable geometry micro-CT can be realigned at each scan by analyzing the projection data set (or part of it), without additional phantom measurements. To assess the reliability of our method we have made several acquisitions on a phantom, each carried after mechanical stress of the positioning system. We showed that the semi-automatic calibration method is an effective aid for a fast and reliable scanner realignment, without additional acquisitions.**

***Index Terms*—MicroCT, geometrical calibration, variable geometry scanners**

I. Introduction

High resolution x-ray micro-computed tomography (micro-CT) is an unrivaled tool for morphological imaging of small sized animals and specimen. Many mechanical/geometrical designs have been developed, with spatial resolutions ranging from 100 μm down to less than 1 μm, depending on the characteristics of the detector and the x-ray source. The sharpness of the reconstructed image relies strongly on the relative alignment between the focal spot, the axis of rotation (AoR) and the detector; hence, the accurate measurement of geometrical misalignment of the system is necessary to ensure proper image quality. Several authors have investigated the problem of the geometrical calibration in cone beam CT, developing algorithms for the measurement of the misalignment parameters of the system [1]-[3], and for image reconstruction with known geometric misalignments [4]. In case of variable geometry systems, allowing for selection of source-to-axis distance (SAD) and axis-to-detector distance (ADD) in a given range, the problem of geometric calibration arises each time one or both of the above distances are changed.

Fig. 1. The Xalt micro-CT scanner at IFC-CNR, Pisa, Italy

The aim of this work is to assess the capability of a previously published cone beam calibration method [1] to reduce the burden of recalibration after modifications of the geometry setup. The method has been applied on a novel small animal micro-CT scanner with variable geometry, which was recently built in the framework of a collaboration between the Institute of Clinical Physiology (IFC-CNR) and the Functional Imaging and Instrumentation Group (FIIG) of the Department of Physics "E. Fermi" - University of Pisa. The micro-CT scanner was designed to be compatible with the YAP-(S)PET scanner (ISE s.r.l., Vecchiano, Italy), which is installed in the same molecular imaging laboratory at IFC-CNR. This enables us to perform efficient multimodal PET/SPECT/CT imaging on small animals.

II. Meterials and methods

A. Variable geometry micro-CT scanner

The *Xalt* scanner (*X*-ray *A*nima*L* *T*omograph) is a dedicated *in vivo* small animal cone beam micro-CT, designed to provide flexible geometry setup and scanning modality. The x-ray source is a microfocus tube with 50 kV maximum voltage, 1 mA maximum current and 35 μm nominal focal spot size. The source power supply is equipped with a system for fast

on/off swicth via external trigger. The flat panel detector is composed by a CMOS sensor with 2048×1024 square pixels of size 48 μm, for a total active area of 10×5 cm^2; the scintillator is a Gd_2O_2S:Tb (Gadox) layer with plastic backing. The intrinsic spatial resolution of the flat panel detector is 8.5 lp/mm @ 10% MTF.

The source and the detector are mounted on the rotating gantry in such a way that the SAD and the ADD can be changed independently by a simple manual operation; this allows to change the magnification in the range 1.13-2.60. The mechanical system for the shift of the detector and the source is ready to be motorized, to allow for a future implementation of software-controlled geometry modifications. The rectangular detector can also be skewed by 90 degrees about its central axis: this permits to select between wide transaxial field of view (FoV) and long axial FoV, according to the needs of the specific imaging protocol. Depending on the geometry setup, the maximum transaxial FoV is 80 mm in diameter, the maximum longitudinal FoV is 79 mm (for a single circular scan) and the minimum voxel size is 18.5 μm. Both continuous rotation and step-and-shoot scanning modes are possible. The device is fully shielded for x-ray leakage: the dose rate at each point of its external surface is always less than 1 μSv/h as requested by official radiation protection regulatory statements.

The scanner is controlled by dedicated software with user-friendly graphical interface for control/acquisition and analysis/reconstruction. The acquisition program allows to edit the imaging protocol parameters and animal information, to acquire scout views at arbitrary angles before tomographic acquisitions and to acquire flat-field and offset calibration data. It also provides continuous monitoring of the hardware status.

B. Data preprocessing and tomographic reconstruction

The analysis/reconstruction program is a modular program for preprocessing of the raw data, automatic geometric calibration and volumetric reconstruction. In the preprocessing module, the raw data are corrected for dark signal and flat field normalization. The output of this module is a corrected dataset that is ready for tomographic reconstruction. Before reconstruction, the corrected dataset can be used as input of the geometric calibration module for the semi-automatic measurement of the misalignment parameters of the system, as described below. After the geometric calibration, the volumetric reconstruction module is used; this module implements a modified Feldkamp algorithm with perspective correction [4]. The misalignment parameters used in the reconstruction module are those obtained in the geometric calibration module. Before starting the recontruction of the whole volume, three preview images along the main orthogonal planes can be checked to ensure that all parameters are correct. The preview slices can be reconstructed at arbitrary distance from the median planes. The program allows the choice of reconstruction filter, and provides ring artifact reduction, cupping correction and Hounsfield Unit (HU) normalization. The volumetric reconstruction module is optimized for cache memory usage, it exploits symmetries to reduce the backprojection time and it was parallelized with OpenMP to take advantage of modern multi-core processors. All voxels outside of the acquisition FoV are excluded from the backprojection to reduce further the computation time. The backprojection time from a dataset of 720 projection of 2048×1024 pixels is <2 minutes for a volume of 512^3 voxels and <30 minutes for a volume of 1024^3 voxels, using a desktop PC equipped with 2 Intel Xeon 3.0 GHz quad-core CPUs.

C. Semi-automatic cone beam geometric calibration

The method for cone beam geometric calibration is described in detail in [1]. The overall misalignemt of the cone beam system is parametrized with the 6-tuple $\boldsymbol{\delta} = (\delta u; \delta v; \delta D; \phi; \eta; \lambda)$. The parameters δu and δv represent the transversal shift and longitudinal shift of the detector, respectively; the parameter δD is the deviation of the actual SDD from the nominal value; the three angles ϕ, η and λ are the detector skew, slant and tilt, respectively. The misalignment parameters of the cone beam scanner are obtained by minimizing a geometry-dependent cost function:

$$c(\boldsymbol{\delta}) = \sum_{\theta, \mathbf{u}_t, i} e_\theta^2(\mathbf{u}_t, \boldsymbol{\delta})|_{\mathrm{ROI}_i}. \quad (1)$$

In the above equation, the function $e_\theta(\mathbf{u}_t, \boldsymbol{\delta})$ is the *error projection* at angle θ which is computed as follows:

$$e_\theta(\mathbf{u}_t, \boldsymbol{\delta}) = g_\theta(\mathbf{u}_t) - g_{\theta+\pi-2\gamma}(\mathbf{H}_{\boldsymbol{\delta}}\mathbf{u}_t). \quad (2)$$

In practice, the error projection is computed by subtracting from each line integral $g_\theta(\mathbf{u}_t)$, measured at a given point $\mathbf{u}_t$ of the misaligned detector plane and for a gantry angle θ, the corresponding redundant line integral $g_{\theta+\pi-2\gamma}(\mathbf{H}_{\boldsymbol{\delta}}\mathbf{u}_t)$. The operator $\mathbf{H}_{\boldsymbol{\delta}}$ is a generalized reflection operator that depends on $\boldsymbol{\delta}$, as explained in [1]. The computation of c is restricted to a small number of ROI's in the error projection, for the reasons explained below. It has been shown that the redundancy of the projection data from a full scan, well known in 2D fan-beam geometry for all object functions [5], exists also in cone

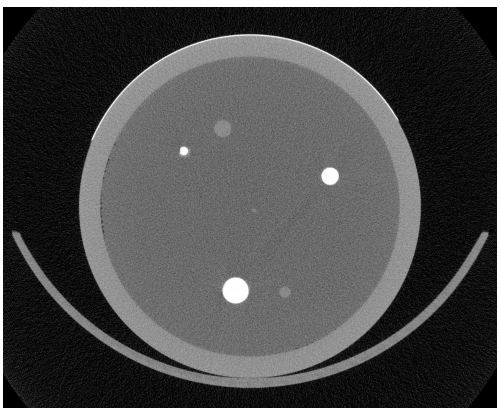


Fig. 2. Reconstructed slice of the multimodal CT-PET mouse phantom (level = 750 HU; window = 3500 HU; voxel size = 37 μm)

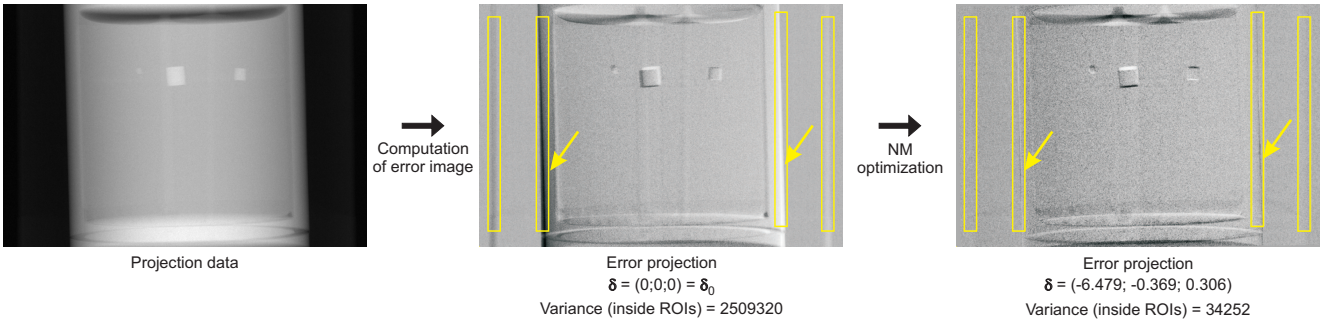


Fig. 3. Computation of the error projection and selection of the ROI's to be included in the computation of the cost function. The arrows indicate sharp edges on the error projection, that disappear when the optimal misalignment parameters of the system are found. When the optimal parameters are found, the cost function (strictly related to the ROI variance) reaches a minimum. The images refer to experiment n. 4.

beam geometry for objects that are uniform along z and/or invariant for rotations about the AoR. For those objects, the error projection computed at an arbitrary gantry angle with the true misalignment parameters $\boldsymbol{\delta}_{\text{true}}$ is a null function (for noiseless and continuously sampled data). For real data from generic objects it was shown that, if c is computed in some properly selected ROI's on the error projection, a minimum is obtained at $\boldsymbol{\delta} = \boldsymbol{\delta}_{\text{true}}$. Hence, the geometric calibration of the system can be set as an optimization problem. We use the simplex method of Nelder and Mead [6] for the minimization of the cost function (1). The user can select the ROI's and the number of projection angles to be included in the computation of the cost function via a graphical user interface.

D. Phantom measurements

In order to study the impact of geometry setup modifications on the misalignment parameters of the system, we have done six experiments on a cold multimodal PET/CT mouse phantom (4 cm diameter) (Isotope Products Laboratories, Valencia, CA). A reconstructed slice of the phantom is shown in Figure 2. The main parameters of each experiment are reported in Table I. Four experiments have been performed with the same setup, each after mechanical stress (repositioning) of the system. The mechanical stress consisted in moving back and forth the x-ray tube and the detector supports, and by temporarily changing the detector orientation, in order to check the reproducibility of the mechanical positioning. The other two experiments were done on different positions. Among all the possible geometric configurations, only three were selected to get a FoV size suitable for the chosen test object (i.e., with transaxial FoV size > 5 cm). In all experiments, the detector orientation was in wide transaxial FoV mode (large side in transverse direction). All scans were done with the following parameters: 50 kV, 0.5 mA, 2 mm Al filtration, 720 projections over 360 degrees, 0.5 s/projection.

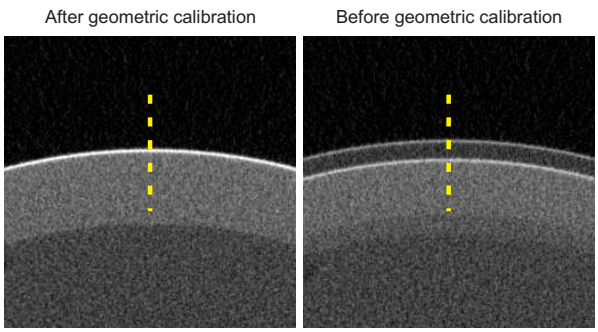


Fig. 4. Details of the reconstructed image of the phantom (experiment n. 1), before and after geometric calibration. The sharp thin detail is a label tape placed on the external surface of the phantom.

The calibration module was used to find the misalignment parameters of the system. As explained in [1], only the three parameters $(\delta u; \phi; \eta)$ were determined. The remaining parameters δv, δD and λ) were all fixed to 0. Because these latter parameters have less impact in the final image quality, the accuracy of the mechanical positioning system was sufficient to leave them uncorrected without significant image artifacts. Four ROI's on the error projections were placed upon sharp edges arising from the phantom and the bed, as shown in Figure 3. In all experiments, the cost function (1) was computed using $N = 2 \cdot 10^5$ points from a single error projection, at $\theta = 0$. A greater amount of data could be selected for the computation of the cost function in a full 2π scan. We have chosen to include only the data from the first error projection for two reasons: first, the error projection at $\theta = 0$ is the only one that is available also for short scans; second, a reduced number of line integrals included in the computation of (1) lead to a speed-up of the calibration process, even if this could reduce the robustness with respect to noise. In all cases, the initial point of the iterative minimization process was set to $\boldsymbol{\delta}_0 = (0; 0; 0)$; hence, no prior estimates of the actual geometric misalignment of the system were used.

For each experiment n, two reconstructions with Ram-Lak filter were done: the former was reconstructed with the misalignment parameters $\boldsymbol{\delta}_{n-1}$ measured on the dataset of the preceding experiment (no recalibration), and the latter with the parameters $\boldsymbol{\delta}_n$ measured on the dataset of experiment n itself (with recalibration). This was useful to evaluate the impact of the mechanical stress between two consecutive experiments, as well as the capability of the calibration method in identifying the correct misalignment parameters each time the system is modified. For the first experiment, $\boldsymbol{\delta}_{n-1} = \boldsymbol{\delta}_0 = (0; 0; 0)$.

TABLE I
GEOMETRIC SETUP OF PHANTOM EXPERIMENTS AND CALIBRATION RESULTS

Exp. number n	SAD/ADD (mm)	Vox. size (μm)	Transv. FoV. size (mm)	FoV length (mm)	$\boldsymbol{\delta}_{\mathrm{n}} = (\delta u_{\mathrm{n}}; \phi_{\mathrm{n}}; \eta_{\mathrm{n}})$	Calib. time (s)
1	270/56	39.7	80.5	34.6	(-6.075; -0.365; 0.282)	28
2					(-4.751; -0.364; 1.634)	29
3					(-5.526; -0.362; 0.313)	26
4					(-6.479; -0.369; 0.306)	41
5	186/76	34.0	62.4	29.0	(-4.001; -0.367; 0.837)	60
6	186/56	36.9	60.8	31.6	(-6.006; -0.405; -0.098)	34

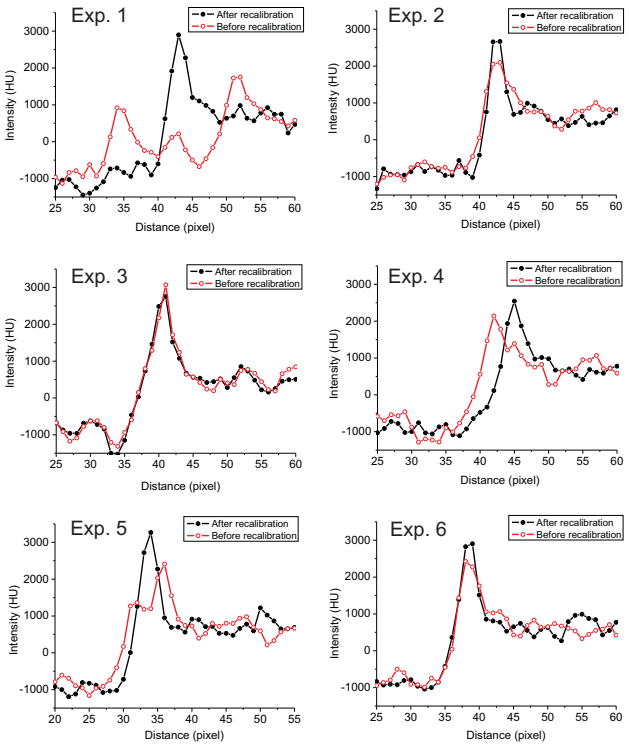


Fig. 5. Profiles of the reconstructed images along the lines shown in Figure 4, before and after geometric calibration.

After each reconstruction, a line profile along a thin detail of the phantom (Figure 4) was evaluated to check the correctness of the measured misalignment parameters. The profiles were made all on slices far from the midplane and the AoR, in order to highlight residual artifacts due to wrong values of skew and slant, if present.

III. RESULTS

The optimal misalignment parameters from each dataset n are reported on Table I. For the first four experiments at SAD = 270 mm and ADD = 56 mm, the measured δu was in the range [-6.479 pixels; -4.751 pixels]; ϕ varied in the range [-0.369 degrees; -0.362 degrees] and η was in the range [0.282 degrees; 1.634 degrees]. The variations of the misalignment parameters are on the same order of magnitude also for the other two experiments. These variations, even if small, have lead to a deterioration of the image quality when the system was not recalibrated. This is apparent in Figure 5, were the line profiles highlighted in Figure 4 are compared for $\boldsymbol{\delta}_{n-1}$ (before recalibration) and $\boldsymbol{\delta}_n$ (after recalibration). Except for $n = 3$, the peak of the thin detail after recalibration appears always sharper and higher than that obtained before recalibration. In the third experiment, the profile appeare almost identical before and after recalibration. In all cases, an average time of 36 s was spent for the iterative minimization of the cost function.

IV. DISCUSSION AND CONCLUSIONS

We have built a micro-CT scanner with variable geometry for high resolution *in vivo* imaging on small animals. Due to the strong sensitivity of the image quality upon small variations of the misalignment parameters, a recalibration of the system is necessary after modifications of the geometry setup. We have shown that our semi-automatic calibration method is effective in reducing the burden of recalibration. Less than 1 minute was spent to find the optimal misalignment parameters on each dataset. This time could be reduced by parallelization of the algorithm for the computation and minimization of the cost function. The only task requested to the user is the selection of the ROI's on the error projections to be included in the computation of the cost function. As it is expected, the accuracy of the calibration result can depend on the choice of these ROI's. A possible solution is to add fixed sharp details on the acquisition FoV (e.g., on the animal bed) on which the ROI's could be automatically placed by the software that implements the method.

REFERENCES

[1] D. Panetta, N. Belcari, S. Moehrs and A. Del Guerra, "An optimization-based method for geometrical calibration in cone beam CT without dedicated phantoms", *Phys. Med. Biol.* vol. 53, pp. 3841-3861, 2008.

[2] K. Yang, A.L.C. Kwan, D.F. Miller and J.M. Boone, "A geometric calibration method for cone beam CT systems" *Med. Phys.* vol. 33, pp. 1695-1706, 2006.

[3] F. Noo, R. Clackdoyle, C. Mennessier, T.A. White and T.J. Roney, "Analytic method based on identification of ellipse parameters for scanner calibration in cone beam tomography", *Phys. Med. Biol.*, vol. 45, pp. 3489-3508, 2000.

[4] M. Karolczak, S. Schaller, K. Engelke, *et al*, "Implementation of a cone beam reconstruction algorithm for the single-circle source orbit with embedded misalignment correction using homogeneous coordinates", *Med. Phys.* vol. 28, pp. 2050-2069, 2001.

[5] A.C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*, (Piscataway, NJ: IEEE).

[6] J.A. Nelder and R. Mead "A simplex method for function minimization" *Comput. J.*, vol. 7, pp. 308-313, 1965.

Online cone beam reconstruction with displaced flat panel detector

Michael Manhart[a], Frank Dennerlein and Holger Kunze[b]

***Abstract*— In 3D medical imaging, the size of the reconstructable field-of-view (FOV) is a factor that significantly defines the clinical benefit of medical X-ray systems. Due to the small detector size, the diameter of the FOV of C-arm devices is limited, but it is known that the FOV can be enlarged by using a horizontally displaced detector for data acquisition in a full rotation. For such geometries recently new reconstruction algorithms were proposed, with which the image quality of the reconstructions could be improved compared to previously known ramp filtered backprojection algorithms [3]. The proposed algorithms however require a rebinning step involving complementary rays, making an online reconstruction almost impossible. We present a new algorithm, modifying the algorithm of Kunze et. al. [4] allowing an online reconstruction.**

***Index Terms*—displaced detector, cone-beam CT, large volume**

I. Introduction

C-arm computed tomography has become an integral part of interventional radiology procedures. It is a useful tool during liver lesion embolisation to visualize feeding arteries or during drainage insertions to guide the placement of tubes[1 ,2].

However, conventional C-arm devices often suffer from a limited reconstructable field of view (FOV) which prohibits the imaging of complete organs or both the target and the entry point of needle applications.

It is known that the diameter of the FOV can be almost doubled by performing an acquisition with a shifted detector in a full rotation. Various algorithms are known for this geometry:

In [3], a FDK algorithm with a specific detector weighting scheme applied before the filtering step (W-FDK) is proposed which computes an artifact-free reconstruction for the plane in which the tube moves, but for large cone angles severe artifacts occur.

These artifacts can be reduced by the algorithms proposed in [4] and [5] which are based on the reconstruction scheme introduced by Noo [6].

Compared to the differentiation backprojection filtration algorithm described in [7] these algorithms have the advantage that smaller regions in the FOV can be reconstructed in high resolution without the need of reconstructing long stripes of the FOV to perform the Hilbert transform in the reconstruction domain.

Unlike the W-FDK algorithm, the algorithms described in [4] and [5] have the disadvantage that they require a rebinning to synthesize virtual projections of approximately double width from the original data to perform a high quality Hilbert transform. This property of the reconstruction algorithm prohibits an online reconstruction during data acquisition. However the ability of online reconstruction is an important feature for reconstruction algorithms in an interventional environment, as the result has to be available shortly after the last projection was acquired to influence the progress of the procedure.

Therefore, we have reviewed the data extension step of [4] to provide an algorithm which enables online reconstruction without the disturbing reconstruction artifacts.

II. Reconstruction Algorithm

A. Large Volume Cone-Beam Reconstruction Formula

This section starts with a review of the algorithm Kunze et al. described in [4] for the reconstruction of large FOV from data acquired with a displaced, equally spaced flat panel detector. This algorithm can be described as a generalization of the fan-beam inversion formula suggested in [6] to reconstruct the x-ray linear attenuation coefficient $f(\vec{x})$ with $\vec{x} = (x, y, z)$ from the cone beam projections

$$g(\lambda, \vec{\alpha}) = \int_0^\infty dt f(\vec{a}(\lambda) + t\vec{\alpha}).$$

$\vec{a}(\lambda)$ is the source position parameterized by the polar angle λ and $\vec{\alpha}$ the unit vector directing from the source to the detector coordinate (u, v)

$$\vec{\alpha}(\lambda, u, v) = \frac{-D\vec{e}_w(\lambda) + u\vec{e}_u(\lambda) + v\vec{e}_v(\lambda)}{\sqrt{u^2 + v^2 + D^2}},$$

with $\vec{e}_u(\lambda)$ a unit vector in row direction, $\vec{e}_v(\lambda)$ a unit vector in column direction and $\vec{e}_w(\lambda)$ a unit vector orthogonal to the detector, see figure 1. The detector coordinate (0, 0) corresponds to the orthogonal projection of the source point onto the detector. The distance between detector and source is D.

In the following we assume that the detector is displaced in positive u direction. The detector could be totally displaced in one direction, so that only data for $u \geq 0$ is considered. But for stability reasons, we use a small overlap $2u_\varepsilon$ so that $g(\lambda, u, v)$ is measured in the interval $[-u_\varepsilon; u_m]$.

[a] M. Manhart: Pattern Recognition Lab, University of Erlangen-Nuremberg, Martensstraße 3, 91058 Erlangen, Germany

[b] F. Dennerlein, H. Kunze: Siemens Healthcare, 91301 Forchheim, Germany. E-Mail: holger.hk.kunze@siemens.com.

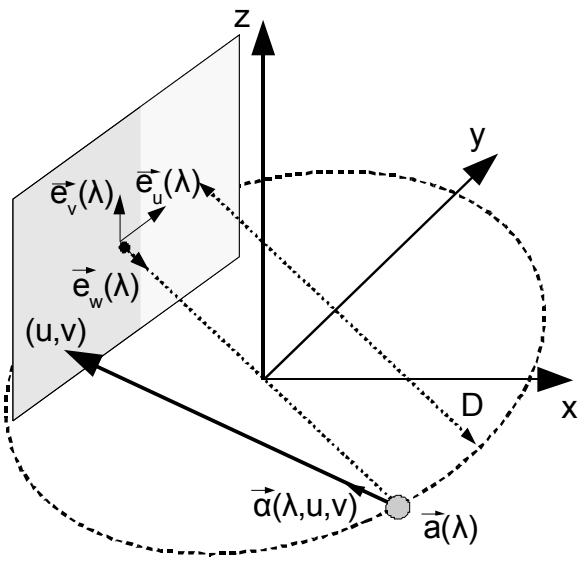


Fig. 1: Illustration of the cone-beam geometry

Given a set of such projection data, the object density can be obtained according to [4] as follows:

1. Compute the extended projection using the rebinning formula

$$g_E(\lambda,\vec{\alpha}(\lambda,u,v)) = \begin{cases} g\left(\lambda_E,\vec{\alpha}(\lambda_E,-u,v)\right), u < -u_\varepsilon \\ g(\lambda,\vec{\alpha}(\lambda,u,v)) \quad , u \geq -u_\varepsilon \end{cases}$$

with

$$\lambda_E = \lambda + \pi - 2\tan^{-1}\frac{u}{D}$$

2. Calculate the partial derivative of the extended data with respect to the parameter of the source trajectory

$$g_D(\lambda,u,v) = \frac{d}{d\mu} g_E(\mu,\vec{\alpha}(\lambda,u,v))\Big|_{\mu=\lambda}$$

3. Multiply the differentiated data with a length correction weight

$$g_C(\lambda,u,v) = g_D(\lambda,u,v)\frac{D}{\sqrt{D^2+u^2+v^2}}$$

4. Perform a Hilbert transform

$$g_F(\lambda,u,v) = \int_{-u_m}^{u_m} d\tilde{u} h(u-\tilde{u}) g_C(\lambda,\tilde{u},v)$$

Multiply the Hilbert transformed data with a smooth weighting function to suppress artifacts due to the data extension.

$$g_W(\lambda,u,v) = g_F(\lambda,u,v) w(\lambda,u)$$

$$w(\lambda,u) = \begin{cases} 0 & u < -u_\varepsilon \\ \sin^2\left(\frac{u+u_\varepsilon}{4u_\varepsilon}\pi\right) & -u_\varepsilon \leq u \leq u_\varepsilon \\ 1 & u > u_\varepsilon \end{cases}$$

5. Back project the filtered data using a linear distance weighting

$$f(\vec{x}) = \frac{1}{2\pi}\int_0^{2\pi} d\lambda \frac{g_W(\lambda,u^*,v^*)}{(\vec{a}(\lambda)-\vec{x})\vec{e}_w(\lambda)}$$

where (u^*,v^*) are the coordinates of the cone beam projection of $\vec{x}$ onto the detector.

B. Separate reconstruction of measured and extended data

To achieve the goal of online reconstruction, it is necessary to process the measured data of a projection and its data extension separately, each at the moment when they can be retrieved.

For that reason the extended projection $g_E(\lambda,u,v)$ needs to be written as a linear combination of the measured data $g(\lambda,u,v)$ and the data extension $g_e(\lambda,u,v)$

$$g_E(\lambda,u,v) = \phi(\lambda,u,v)g(\lambda,u,v) + (1-\phi(\lambda,u,v))g_e(\lambda,u,v)$$

with $\phi(\lambda,u,v)$ a weighting function blending from the original data to the data extension. In the above reviewed algorithm $\phi(\lambda,u,v)$ is the Heaviside step function in the parameter u shifted by $-u_\varepsilon$.

Due to the linearity of the reconstruction algorithm we can process both components of $g_E(\lambda,u,v)$ separately obtaining the volumes $f_m(\vec{x})$ of the measured projections and $f_e(\vec{x})$ of the projection extension.

C. Approximate data extension

In the algorithm described in section A, the data extension was done using a rebinning formula. This algorithm guarantees for the plane z = 0 an artifact free reconstruction, but is suboptimal with respect to computational effort and memory requirements: To be able to extend one projection several complementary projections are required.

Therefore, we propose a different data extension scheme. Due to the large extend of truncation and due to the fact that many high frequent structures are typically located at the position of truncation in such a geometry, the commonly used data extension algorithms based on the fitting of the projections of circles or ellipses are not good choices [8].

In our case, however, we can make use of the additional knowledge about the object structure contained in the opposite projections. We thus propose to use the data acquired at the position $\lambda+\pi$ to extend the current projection. This extension can then be written as

$$g_e(\lambda,u,v) = g(\lambda+\pi,-u,v).$$

Thus only data from one single projection is used for the projection extension.
Of course, this approach yields only a coarse approximation of the missing projection data. Note, however, that the extended data values are only required for the Hilbert filtering and not used during backprojection, so that the impact on image quality caused by this approximation is very low.

D. Online reconstruction algorithm

Using the linearity property of the reconstruction algorithm described in section B and the approximate data extension proposed in section C we propose the following online algorithm for a large FOV reconstruction using a displaced detector:
First, the reconstruction algorithm of section A is applied for the weighted projection $\phi(\lambda,u,v)g(\lambda,u,v)$ skipping step 1. Thus only the measured data for the projection at source position $\vec{a}(\lambda)$ are processed.
In a second step, the same projection data are flipped to obtain the projection extension $g_e(\lambda-\pi,u,v)$ for the complementary source position $\vec{a}(\lambda-\pi)$. Note that λ is a polar angle and thus $\vec{a}(\lambda)$ is a 2π periodic function. After weighting with $(1-\phi(\lambda-\pi,u,v))$, they are processed once again by the algorithm of section A without step 1, but this time using the geometry of the source position $\vec{a}(\lambda-\pi)$.
For $\phi(\lambda,u,v)$, a cosine square function is chosen, ramping down from 1 to 0 within the range from $[-u_\varepsilon;0]$.

III. NUMERICAL EVALUATION

We compared our new algorithm of section II D to the algorithm that has been suggested in [4].
In both algorithms, we implemented the differentiation (step 3) according to the scheme described in [9]. The Hilbert transform was computed using the half pixel shift formula; see for instance [10].
As test object we chose the Schaller head phantom [11]. Cone beam projections were simulated using the parameters listed in table 1.

source isocenter distance	R = 570 mm
source detector distance	D = 1200 mm
detector pixel size	Δu = Δv = 1 mm
# of projections	720
overlap	u_ε = 25 mm

Table 1: Geometry parameters used for data simulation

Figure 2 presents two slices (at z = 0cm and z = -2.4cm) through the reconstructions of the head phantom. Note that the visible stripes are a result of the low number of the projections and are not caused by the displaced detector.

Additionally to simulation studies we compared the reconstruction result using scanned phantoms. In figure 3 the reconstruction result of an abdomen phantom is shown.

IV. CONCLUSION

We presented a new algorithm for full scan cone beam reconstruction using a displaced detector. The results are comparable to these of the algorithm described in [4] however the new algorithm enables an online reconstruction of the investigated object. This property is gained for the computational cost of one additional back projection; however the rebinning step could be skipped which allows online reconstruction, eases data handling and reduces memory requirements.

REFERENCES

[1] B.C. Meyer, B.B. Frericks, M. Voges, M. Borchert, P. Martus, J. Justiz, K.-J. Wolf and F. K. Wacker: "Visualization of Hypervascular Liver Lesions During TACE: Comparison of Angiographic C-Arm CT and MDCT", *AJR*, 190: W263 - W269, 2008

[2] C.Becker, T. Waggershauser, K. Krimmel, G. Michalski, U. Lechel, O. Meissner: "LARGE VOLUME C-arm CT: Liver Coverage, Image Quality, and Radiation Exposure", *RSNA* 2008, SSJ24-02

[3] G. Wang: "X-ray micro-CT with a displaced detector array", *Med. Phys.* 29 (7), pp. 1634 – 1636, Jul. 2002

[4] H. Kunze, M. Manhart, F. Dennerlein: " Cone beam reconstruction with displaced flat panel detector", Proc. Int. Conf. Fully 3D Image Reconstruction, pp 138-145

[5] D. Schäfer, M. Grass: " Cone-beam filtered back-projection for circular X-ray tomography with off-centerdetector"", Proc. Int. Conf. Fully 3D Image Reconstruction, pp 86-89

[6] F. Noo , M. Defrise, R. Clackdoyle, H. Kudo: "Image reconstruction from fan-beam projections on less than a short-scan", *Phys. Med. Biol.* 47, pp. 2525-2546, 2002

[7] S. Leng, T. Zhuang, BE. Nett, GH. Chen, Phys. Med. Biol., Vol. 50, 1805–20, 2005.

[8] J. Hsieh, E. Chao, J. Thibault, B. Grekowicz, A. Horst, S. McOlash, T.J. Myers: "A novel reconstruction algorithm to extend the CT scan field-of-view." *Med. Phys.* 31 (9), pp. 2385-2391, 2004

[9] F. Noo, S. Hoppe, F. Dennerlein, G. Lauritsch, J. Hornegger: "A new scheme for view-dependent data differentiation in fan-beam and cone-beam computed tomography", *Phys. Med. Biol.*, Vol 52, pp. 5393-5414, 2007

[10] F. Noo, J. D. Pack, D. Heuscher: "Exact helical reconstruction using native cone-beam geometries." *Phys. Med. Biol.*, Vol 48, pp. 3787-3818, 2003

[11] S. Schaller, Th. Flohr, P. Steffen: "An Efficient Fourier Method for 3-D Radon Inversion in Exact Cone-Beam CT Reconstruction", *IEEE Transactions on Medical Imaging*, Vol. 17 (2), pp. 244-250, Apr. 1998

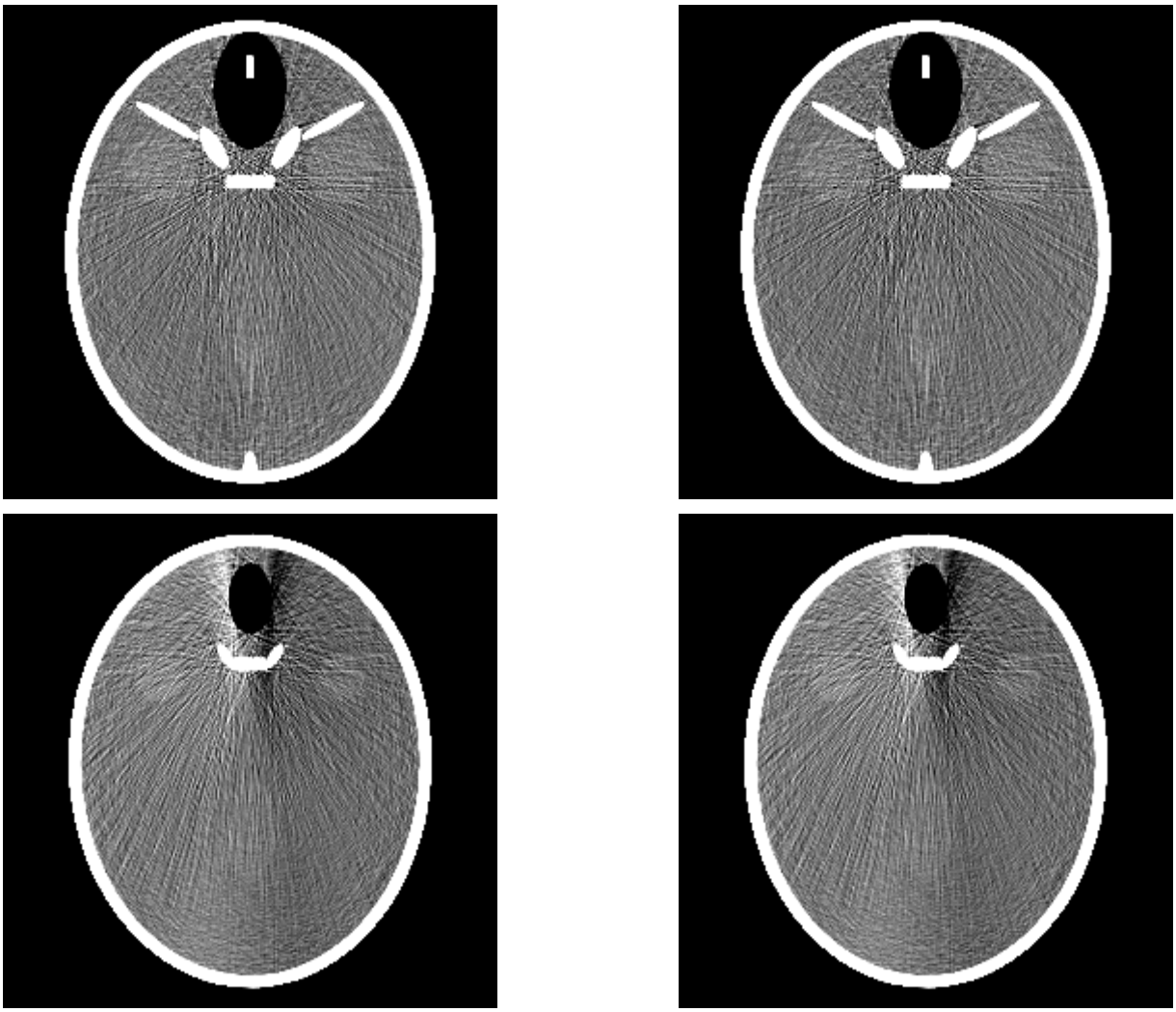


Fig. 2: Comparison of the reconstruction result using the proposed method (left) and method proposed in [4] (right) for z = 0 cm (top) and z = -2.4 cm (bottom) in the gray scale window [-100 HU; 100 HU]

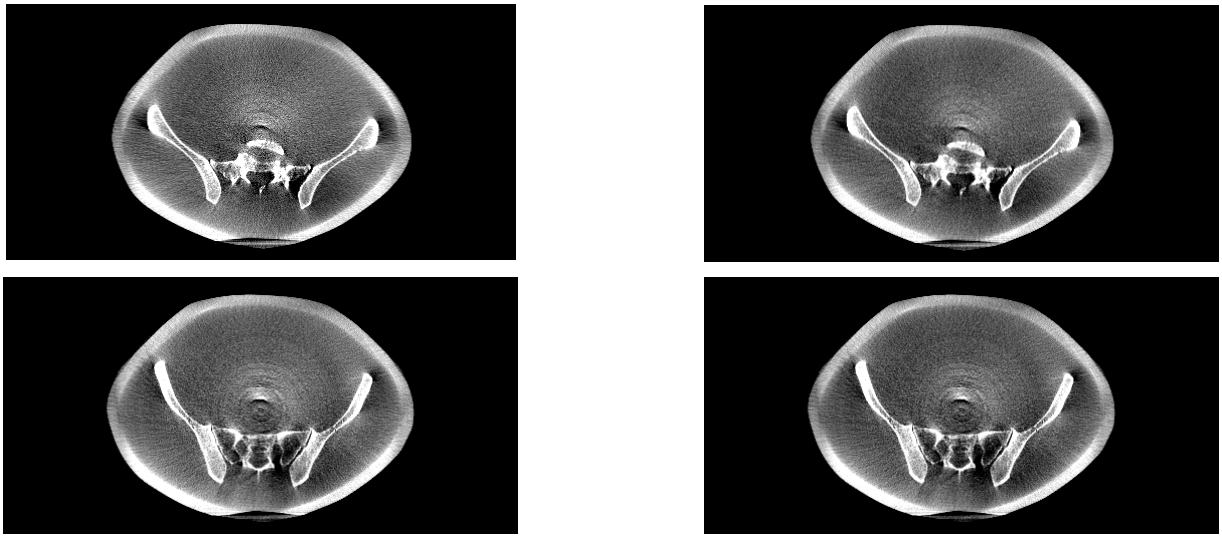


Fig. 3: Comparison of the reconstruction result using the proposed method (left) and method proposed in [4] (right) for z = 0 cm (top) and z = -2 cm (bottom) in the gray scale window [-200 HU; 400 HU]

Cone-beam Artifact Evaluation of the Factorization Method

Frank Dennerlein and Frédéric Noo

***Abstract*— We present a quantitative and qualitative CB artifact investigation of the factorization approach recently suggested for image reconstruction in circular cone-beam computed tomography. This investigation involves simulated data and for the first time also phantom and clinical CB data acquired with a Siemens C-arm scanner. For comparison, the results obtained with the short-scan FDK method (for the simulated and real data) and full-scan FDK (only for simulated data) are presented.**

I. Introduction

It is well-known that CT image reconstruction in the short-scan circular cone-beam geometry is an ill-posed problem, because circular data acquisition does not yield a complete set of the 3D Radon values of the investigated object. This complete set, however, would be required for an exact and stable reconstruction of the object density [1]. Missing Radon data causes CB artifacts in the reconstruction results. The strength and appearance of these artifacts depend on the structure of the investigated object, but typically also differ significantly from one reconstruction algorithm to another.

Many attractive algorithms have been suggested over the years for short-scan CB reconstruction, see for instance [2]–[4]. One recently introduced method is the factorization approach [5]. In this algorithm, the 3D problem is decomposed into a set of independent 2D inversion problems using analytical steps. A solution to each 2D problem is then estimated iteratively, involving a gradient-descent scheme with early stopping regularization [5]. First numerical results indicated that the factorization approach is fairly robust with respect to the missing 3D Radon information [5].

In this paper, we will further investigate the CB artifact behavior of the factorization approach in a typical C-arm geometry, both from simulated and for the first time also from real phantom as well as clinical CB data. For comparison, we will show the reconstructions obtained with the widely-used short-scan FDK method and also give results obtained with full-scan FDK, as a benchmark, for the simulation studies.

II. Preliminaries

In this section, we briefly summarize the geometrical set-up of our study and present some details about the implementation of the considered reconstruction methods. For an in-depth description of these algorithms, however, we redirect the reader to [6], [2] and [5], resp.

F. Dennerlein is with Siemens AG, Healthcare Sector, Forchheim, Germany and F. Noo is with UCAIR, Dept. of Radiology, University of Utah, Salt Lake City, USA. The work of F. Noo was supported by the NIH grant R21 EB009168. The concepts and information presented in this paper are based on research and are not commercially available.

A. Geometrical Set-Up

In the following sections we will present reconstructions from several distinct CB data sets. Each data set was acquired (or, simulated) using a planar X-ray source trajectory of circular shape. The symbol R denotes the trajectory radius, and λ gives the angular position of the source during the scan. Relative to the source trajectory, we introduce a right-handsided $x-y-z$ coordinate system such that the plane of the scan is at $z = 0\,\text{mm}$. CB projections are collected/simulated in the interval $[\lambda_c - \lambda_s/2, \lambda_c + \lambda_s/2]$, with λ_s describing the short-scan length and λ_c giving the central location on the source trajectory. The polar angle increment between two adjacent projections is denoted as $\Delta\lambda$. We use a flat panel detector with pixel size Δu in horizontal and Δv in vertical direction and use the quantity D to describe the distance between X-ray source and detector. See figure 1 for an illustration and table I for the values of the introduced quantities applied during the evaluation.

B. Implementation Details

For the results presented below, we applied the FDK methods with sinc-apodization on the ramp-filter kernel. Short-scan FDK involved the Parker weighting scheme described in [7] to approximately cope with redundant data.

The implementation of the Factorization approach is identical to that suggested in [5]. The algorithm described in that paper requires the selection of five parameters, which we set to $\epsilon = 0.01$, $\alpha = 0.01$, $\sigma = 0.7$, $\gamma_{thres} = 0.002$ and $\gamma_{max} = 400$ for the presented evaluation.

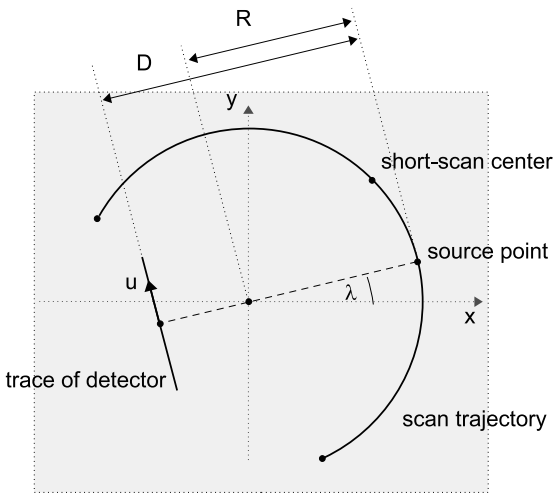


Fig. 1. Top-view illustration of the considered CB acquisition geometry.

TABLE I

GEOMETRY PARAMETERS FIXED DURING EVALUATION

	simulation	real data
trajectory radius [mm]	$R = 750$	$R = 750$
detector-source distance [mm]	$D = 1200$	$D = 1200$
angular increment [°]	$\Delta\lambda = 0.4$	$\Delta\lambda = 0.4$
detector pixel size [mm]	0.8	0.616
detector area [mm× mm]	336×400	382×296

III. CONE-BEAM ARTIFACT ASSESSMENT

A. Quantitative studies

This section presents a quantitative assessment of CB artifacts on simulated data sets of the FORBILD head phantom without noise and without truncation. Artifact quantification is based on comparing the reconstruction result f^e to the true object density function f using the concept of relative reconstruction error that is here defined for each point $\underline{x} = (x, y, z)$ at which $f(\underline{x}) \neq 0$ as

$$\epsilon(\underline{x}) = \frac{f^e(\underline{x}) - f(\underline{x})}{f(\underline{x})}. \tag{1}$$

The distribution of this error over specific object regions will be used to estimate the CB artifact level, as described below.

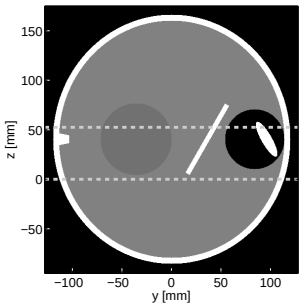


Fig. 2. Slice $x = 0$ mm though the FORBILD head phantom. The two dashed lines indicate the plane of the scan as well as the plane investigated in Fig. 3.

In our study, we shifted the phantom so that its central slice is at $z = 40$ mm; see figure 2 and simulated data for a full scan (to enable the benchmark reconstruction with full-scan FDK) and for several short-scan realizations, all of them of length $\lambda_s = 210°$.

For the artifact study, we investigated image quality on a slice by slice basis, using slices that are orthogonal to the z-axis. Within each slice, a region $\mathcal{S}_z$ was defined that consists of all points at which the true density f takes values between 30 HU and 80 HU and which are at least 0.5 mm away from the discontinuities in the object. CB artifacts were then quantified using the distribution of the relative reconstruction error $\epsilon(\underline{x})$ for $\underline{x} \in \mathcal{S}_z$; we here represent this distribution by its mean $\bar{\epsilon}_z$ and its standard deviation ϵ_z^σ. Note that a mean value different from $\bar{\epsilon}_z = 0$ indicates a reconstruction bias while ϵ_z^σ describes the heterogeneity of the CB artifact structure.

Figure 3 presents results for the short-scan centered at $\lambda_c = 0°$, next to the full-scan benchmark reconstruction. The left column shows the reconstructed densities within the slice $z = 52.5$ mm. The corresponding distribution of ϵ within $\mathcal{S}_z$ is presented in the histogram on the right. The two FOMs $\bar{\epsilon}_z$ and ϵ_z^σ are indicated with a black circle and a black horizontal line in each histogram, respectively. We observe that short-scan FDK yields a fairly wide-spread error histogram, while in the Factorization results, the distribution is much more compact and closer to the reference reconstruction obtained with full-scan FDK. Visually, the remaining artifacts of the Factorization method tend to be of streak-like structure, thus affecting only localized regions within the reconstructions. This behavior is different from that of short-scan FDK, which comes with low-frequency, more globally-distributed CB artifacts.

In short-scan geometries, CB artifacts typically vary with the scan orientation. To address this issue, the quantification described above was carried out for 6 short-scans with varying scan center $\lambda_c = \{0°, 60°, 120°, 180°, 240°, 300°\}$. Doing so, we obtain for each slice 6 values $\bar{\epsilon}_z$ for the reconstruction bias. These 6 values are summarized by their mean and standard deviation. The top of figure 4 shows these two quantities as a function of the slice z. The artifact heterogeneity is quantified in a similar way, using for each slice the mean and standard deviation of ϵ_z^σ across the 6 experiments; see the bottom of figure 4 for an illustration of the artifact heterogeneity FOM.

As expected, all methods perform well close to the plane of the scan, but farther away from this plane, in particular where the object density varies significantly in z, the strength of CB artifacts increases noticeably. For short-scan FDK, we observe a strong bias (reaching up to 1.2%), a strong artifact

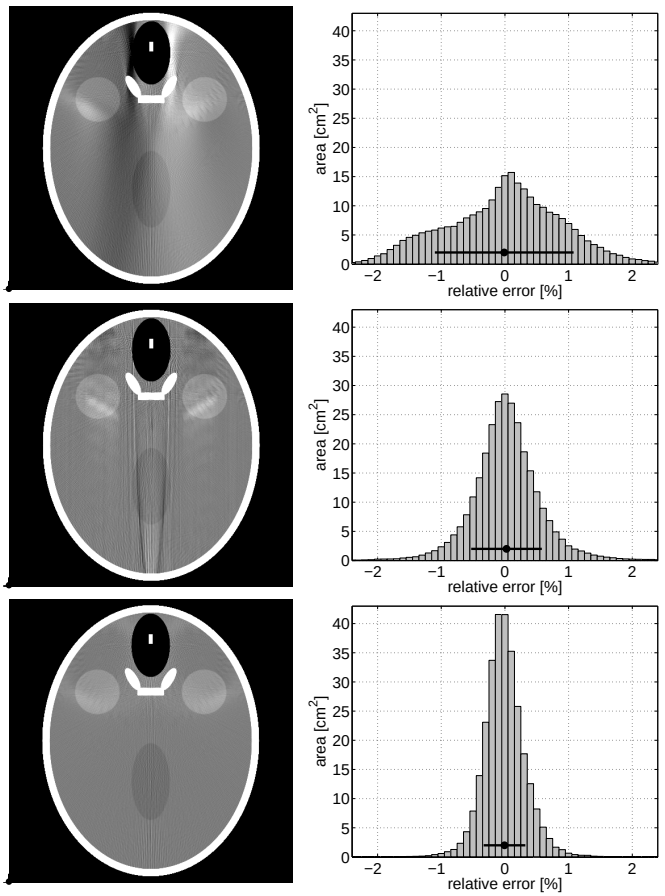


Fig. 3. CB artifact assessment in the reconstructions of the FORBILD head phantom. (Left) The slice $z = 52.5$ mm in the gray-scale window [0 HU, 100 HU] and (right) histograms of the relative reconstruction error inside the region $\mathcal{S}_z$. The horizontal bar in each histogram shows the first moment (circle) and second moment (half-width of the bar) of the histogram.

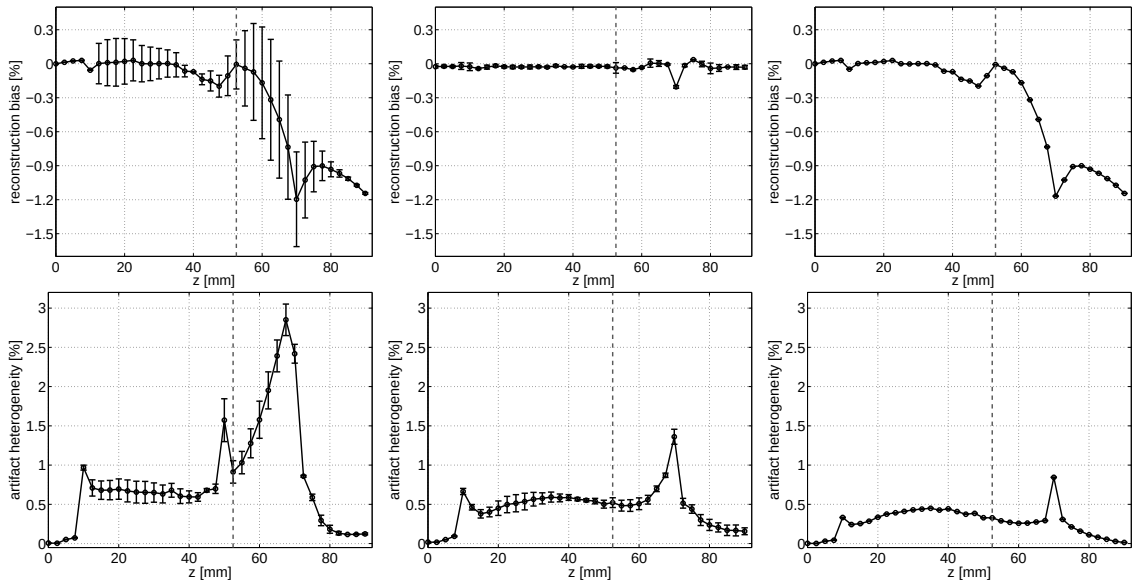


Fig. 4. The CB artifact FOMs (top) reconstruction bias and (bottom) artifact heterogeneity, as a function of z. The black circles indicate the mean of these FOMs across 6 CB data sets with different short-scan center λ_c, while the errorbars illustrate the FOM standard deviations across these 6 different data sets. From (left) to (right): the short-scan FDK method, the Factorization method and full-scan FDK.

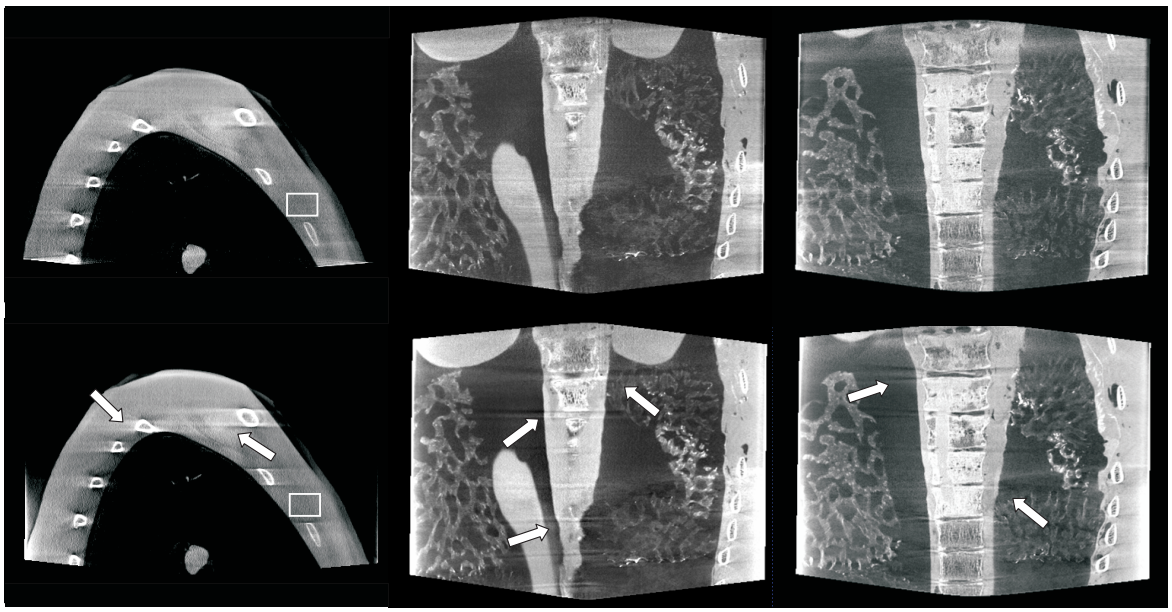


Fig. 5. Reconstruction from the real C-arm data sets in scenario A (left) and scenario B (center and right), obtained with (bottom) short-scan FDK and (top) the factorization approach. The results are presented selecting grayscale-windows such that the mean of the reconstructed values in specific ROIs gave identical gray values in both images. One such ROI is illustrated in the images on the left side. The arrows highlight some regions where short-scan FDK shows strong CB artifacts.

heterogeneity of maximum 3% and a high sensitivity to the scan orientation; see the long error-bars in the corresponding plots. In the factorization results, the strength of CB artifacts is fairly independent of the actual scan realization, and artifact heterogeneity is always less than 1.5%. The most sigificant advantage of the factorization approach, however, is that it yields a very low reconstruction bias, even at higher z. In this property, it outperforms both, full-scan and short-scan FDK.

B. Real Data Reconstruction

We now present several reconstructions from real short-scan CB data collected on commercially-available C-arm systems (Siemens AXIOM Artis [8]) for visual inspection of image quality. Compared to the simulated data, the real data contains inconsistencies due to physical effects as well as potential deviations from the ideal circular acquisition geometry. Our focus is to see if the factorization method can handle these inconsistencies robustly and retain its image quality benefits compared to short-scan FDK.

In a first, pre-clinical study, we scanned a fairly wide physical thorax phantom using a short-scan of length $\lambda_s = 215^\circ$ and using two distinct scenarios of truncation. In scenario A, the projection of the thorax was never truncated across the upper detector boundary and transaxial truncation occurred only

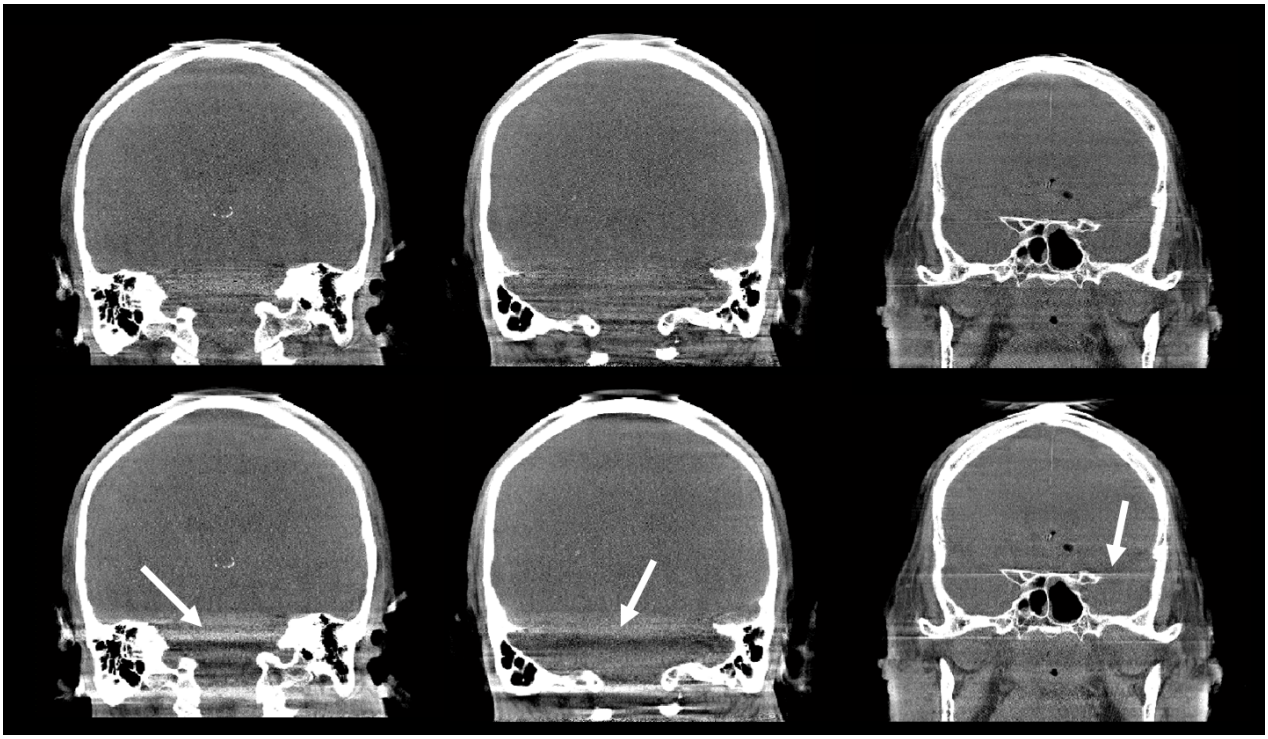


Fig. 6. Clinical data sets, reconstructed with (bottom) short-scan FDK and (top) the factorization approach, in identical grayscale windows. The white arrows again highlight artifact regions in the short-scan FDK results.

towards the ends of the short-scan. In scenario B, CB data were always truncated in axial direction, but transaxial truncation occurred closer to the short-scan center only. Figure 5 presents the reconstruction results in both scenarios on vertical planes, each of which was obtained by averaging 3 planar results that are spaced by $0.25\,\mathrm{mm}$. The averaging reduces noise and contributes to a better visibility of potential CB artifacts in the images.

Our second study involved two clinical head data sets acquired at the Department of Neuroradiology, University of Erlangen[1]. Figure 6 presents three vertical slices of thickness $3\,\mathrm{mm}$ through the reconstructions as obtained with the two algorithms under comparison.

From these results, we observe that the short-scan FDK approach yields fairly strong artifacts tangent to the bony object structures, as indicated with the white arrows. The factorization approach shows a clear reduction of these artifacts and also more tolerance with respect to transaxial truncation (as can be seen in the thorax reconstructions) and thus, altogether, a noticeable improvement of image quality. This improvement becomes particularly visible in the subcranial regions in the clinical data sets.

IV. Conclusions

In this paper, we investigated in more details the CB artifact behavior of the factorization method recently suggested for short-scan circular CB CT. We carried out a quantitative artifact assessment that demonstrated that the Factorization approach comes with very low reconstruction bias compared to short-scan FDK and that it even outperforms full-scan FDK in that property. In addition, the Factorization method turns out to be less sensitive to the short-scan orientation than Feldkamp's approach.

The image quality improvements observed in the simulation study also hold for medical data sets, as demonstrated on the reconstructions of a physical thorax phantom and of neuro data sets acquired with Siemens AXIOM Artis C-arm systems. These benefits for clinical applications were not obvious, since the real CB data contains various physical effects that are inconsistent with the data model that the Factorization approach strongly relies on. Despite these inconsistencies, the Factorization approach significantly reduced the CB artifacts that are present in the short-scan FDK results close to the vertebrae and in the subcranial regions. The Factorization approach might thus be a viable alternative to achieve 3D image reconstruction on commercial scanners.

[1]The authors thank Dr. med. Tobias Struffert, Department of Neuroradiology, University of Erlangen, for providing the clinical patient data sets.

References

[1] H. K. Tuy, "An inversion formula for cone-beam reconstruction," *SIAM J. Appl. Math.*, vol. 43, no. 3, pp. 546–552, 1983.
[2] G. Wang, Y. Liu, T. H. Lin, and P. C. Cheng, "Half-scan cone-beam X-ray microtomography formula," *Scanning*, vol. 16, no. 4, pp. 216–220, 1994.
[3] H. Yu and G. Wang, "Feldkamp-type VOI reconstruction from super-short-scan cone-beam data," *MP.*, vol. 31, no. 6, pp. 1357–1362, 2004.
[4] M. Grass, T. Koehler, and R. Proksa, "3D cone-beam CT reconstruction for circular trajectories," *PMB*, vol. 45, no. 2, pp. 329–347, 2000.
[5] F. Dennerlein, F. Noo, H. Schoendube, J. Hornegger, and G. Lauritsch, "A factorization approach for cone-beam reconstruction on a circular short-scan," *IEEE Trans. Med. Imag.*, vol. 27, no. 7, pp. 887–896, 2008.
[6] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *J. Opt. Soc. Am. A*, vol. 1, no. 6, pp. 612–619, 1984.
[7] M. D. Silver, "A method for including redundant data in computed tomography," *Med. Phys.*, vol. 27, pp. 773–774, 2000.
[8] "AXIOM Artis dBA: Biplane C-arm system with flat detector for angiography, *Siemens AG, Medical Solutions*."

GEL'FAND-GRAEV'S RECONSTRUCTION FORMULA IN THE 3D REAL SPACE

Yangbo Ye[1,2], Hengyong Yu[3] and Ge Wang[3]

1. Department of Mathematics, University of Iowa, Iowa City, Iowa 52242, USA
2. School of Mathematics, Shandong University, Jinan, Shandong 250100, China
3. Division of Biomedical Imaging, VT-WFU School of Biomedical Engineering and Sciences, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, 24060, USA

yangbo-ye@uiowa.edu, hengyong-yu@ieee.org, ge-wang@ieee.org

ABSTRACT

We present a classical reconstruction formula of I.M. Gel'fand and M.I. Graev in a 3D setting for easy access and related research. Simulation will be reported to show effects of the pseudo-differential operators, evaluate quality of reconstructed images and serve as a basis for extensions.

Index Terms— X-ray transform, computed tomography (CT), image reconstruction, pseudo-differential operator.

1. INTRODUCTION

In a series of papers [1-3] and a classical book [4] I.M. Gel'fand, M.I. Graev and others proved inversion formulas for x-ray transforms of various dimensions in different spaces. In particular, Gel'fand-Graev's inversion formula [1] is based on a fundamental relationship which expresses projection data as a Hilbert transform. This finding was re-discovered in recent years; see [5-9]. It has wide and crucial applications in CT reconstruction, including truncated reconstruction [10-11], backprojection filtration (BPF) [12], interior tomography [13-17], and limited-angle tomography [18]. For a survey of these results, please see [19-20].

Despite its high originality and information density, Gel'fand-Graev's inversion formula [1] was cast in high dimensions and specialized terms and is rather difficult even for some dedicated engineers to follow. In this report, we will focus on [1] and represent their inversion formula and its proof for the 1D x-ray transform in a 3D real space in a relatively engineer-friendly language. Clearly, this 3D setting is the most relevant to biomedical imaging

2. NOTATION

We follow the notations in [1]. Set $n = 3$, $k = 1$, $\alpha = \alpha_1 \in \mathbb{R}^3$ non zero, and $\beta \in \mathbb{R}^3$. Then, $h: x = \alpha t + \beta$, $t \in \mathbb{R}$, is a straight line. Let f be a function on $\mathbb{R}^3$. Then, the 1D x-ray projection is an integral transform $f \mapsto \mathcal{R}f$ given by

$$\mathcal{R}f(\alpha,\beta) = \int_{\mathbb{R}} f(\alpha t + \beta)\, dt. \tag{1}$$

Let us describe the relation between α, β and $\mathcal{R}f$ in a formal way.

Denote by $H_{1,3}$ the **manifold of lines** h in $\mathbb{R}^3$. Then, $\dim H_{1,3} = 4$. Denote by $G_{1,3}$ the **Grassman manifold of 1D subspaces** in $\mathbb{R}^3$, which defines a natural mapping $\pi: H_{1,3} \to G_{1,3}$ by translating a line $h \in H_{1,3}$ to the line $a \in G_{1,3}$ through the system origin. Clearly, $\dim G_{1,3} = 2$.

We denote $\mathbb{R}^3 - \{O\}$ by $E_{1,3}$, the manifold of 1-frames, in the sense that any $\alpha \in E_{1,3}$ defines a 1D coordinate system in the line $a: x = \alpha t$, $t \in \mathbb{R}$, in $G_{1,3}$. Then any $(\alpha,\beta) \in E_{1,3} \times \mathbb{R}^3$ defines a line $h \in H_{1,3}$ by $x = \alpha t + \beta$, $t \in \mathbb{R}$ as before, and we denote this by $\sigma: (\alpha,\beta) \mapsto h$. Two $(\alpha,\beta), (\alpha',\beta') \in E_{1,3} \times \mathbb{R}^3$ define the same h if and only if

$$\alpha' = A\alpha,\ \beta' = \beta + \alpha t_0 \text{ for some } A \in \mathbb{R} - \{0\},\ t_0 \in \mathbb{R}. \tag{2}$$

Consider a function $\varphi = \mathcal{R}f$ on $H_{1,3}$. Denote by σ^* the mapping $\sigma^*: \varphi(h) \mapsto \varphi(\alpha,\beta) = \varphi\big(\sigma(\alpha,\beta)\big)$. The image of σ^* is the set of functions on $E_{1,3} \times \mathbb{R}^3$ which are invariant under (2).

Define a **differential 1-form** $\varkappa\varphi$ on $E_{1,3} \times \mathbb{R}^3$ by

$$\varkappa\varphi = \sum_{j=1}^{3} \frac{\partial\varphi(\alpha,\beta)}{\partial\beta^j} d\alpha^j, \tag{3}$$

where $\alpha = (\alpha^1, \alpha^2, \alpha^3)$ and $\beta = (\beta^1, \beta^2, \beta^3)$. We note that

$$\mathcal{R}f(\alpha',\beta') = \int_{\mathbb{R}} f(\alpha' t + \beta')\, dt$$

$$= \int_{\mathbb{R}} f(A\alpha t + \beta + \alpha t_0)\, dt = \int_{\mathbb{R}} f(\alpha(At + t_0) + \beta)\, dt$$

$$= \frac{1}{A}\int_{\mathbb{R}} f(\alpha t + \beta)\, dt = \frac{1}{A}\, \mathcal{R}f(\alpha,\beta).$$

Thus,

$$\sum_{j=1}^{3} \frac{\partial\varphi(\alpha',\beta')}{\partial\beta'^j} d\alpha'^j = \frac{1}{A}\sum_{j=1}^{3} \frac{\partial\varphi(\alpha,\beta)}{\partial\beta'^j} d\alpha'^j$$

$$= \sum_{j=1}^{3} \frac{\partial\varphi(\alpha,\beta)}{\partial\beta^j} d\alpha^j,$$

i.e., the differential form $\varkappa\varphi$ depends on $h = \sigma(\alpha,\beta)$ only and is independent of the choice of (α,β). In other words, $\varkappa\varphi$ is a differential form on $H_{1,3}$.

Let H_0 be the **manifold of oriented 1D lines** in $\mathbb{R}^3$ through the origin. Since the oriented lines have directions,

H_0 is diffeomorphic to the unit sphere S^2 in $\mathbb{R}^3$. Fix a 1D oriented submanifold $\gamma \subset H_0$. Then, γ is an oriented curve on S^2. For any vector $x \in \mathbb{R}^3$, denote by $x+\gamma$ the manifold of oriented lines shifted from those in γ to x. $x+\gamma$ is diffeomorphic to an oriented curve on the unit shpere centered at x. Define an **operator $\mathcal{J}_\gamma$** on $\varphi = \mathcal{R}f$ by

$$(\mathcal{J}_\gamma \varphi)(x) = \frac{1}{2\pi i}\int_{x+\gamma} \varkappa\varphi. \tag{4}$$

Note that $\mathcal{J}_\gamma = 0$ for any closed curve γ, and hence $\mathcal{J}_\gamma$ depends only on the end points of γ.

Take $\xi \in \mathbb{R}^3 - \{O\}$. Denote by G_ξ the **manifold of all subspaces** $h \in H_0$ of $\mathbb{R}^3$ belonging to the plane $\langle \xi, x\rangle = 0$. Then, G_ξ is diffeomorphic to the large circle on the sphere S^2 intersecting the plane $\langle \xi, x\rangle = 0$, oriented by the right-hand rule with ξ pointing to the direction of the thumb. Note that $\dim \gamma = \dim G_\xi = 1$.

Now let us define the **intersection index** $\gamma \cdot G_\xi$ of γ and G_ξ. An intersection of γ and G_ξ contributes 1 to the intersection index if γ and ξ point to the same side of G_ξ looking from the origin. An intersection contributes -1 if if γ and ξ point to opposite sides of G_ξ looking from the origin. Define the **Crofton function** by $\mathrm{Crf}_\gamma(\xi) = \gamma \cdot G_\xi$. Note that $\mathrm{Crf}_\gamma(\xi) = 0, \pm 1$, and $\mathrm{Crf}_\gamma(-\xi) = -\mathrm{Crf}_\gamma(\xi)$.

3. PSEUDO-DIFFERENTIAL OPEARATORS

Let $P(D) = \sum_n a_n D^n$ be a linear differential operator with constant coefficients a_n, acting on smooth functions on $\mathbb{R}^3$. Here $P(\xi) = \sum_n a_n \xi^n$ is a polynomial of three variables with

$$n = (n_1, n_2, n_3),\ \xi = (\xi_1, \xi_2, \xi_3),\ \xi^n = \xi_1^{n_1}\xi_2^{n_2}\xi_3^{n_3},$$
$$D = (-i\partial_1, -i\partial_2, -i\partial_3),\ D^n = (-i\partial)_1^{n_1}(-i\partial)_2^{n_2}(-i\partial)_3^{n_3}.$$

Then, by the Fourier inversion formula

$$P(D)u(x) = \frac{1}{(2\pi)^3}\int_{\mathbb{R}^3} d\xi \int_{\mathbb{R}^3} e^{i(x-y)\cdot\xi} P(\xi)u(y)\, dy.$$

A **pseudo-differential operator** $C(D)$ on $\mathbb{R}^3$ is defined by

$$C(D)u(x) = \frac{1}{(2\pi)^3}\int_{\mathbb{R}^3} d\xi \int_{\mathbb{R}^3} e^{i(x-y)\cdot\xi} C(\xi)u(y)\, dy$$
$$= \frac{1}{(2\pi)^{3/2}}\int_{\mathbb{R}^3} e^{ix\cdot\xi} C(\xi)\hat{u}(\xi)d\xi \tag{5}$$

for a certain function $C(\xi)$ which is called the **symbol of the pseudo-differential operator** $C(D)$.

Theorem 1. (Gel'fand and Graev [1]) *The composition operator $\mathcal{J}_\gamma\mathcal{R}$ defined by (1) and (4) is a pseudo-differential operator on $\mathbb{R}^3$ with symbol* $\mathrm{Crf}_\gamma(\xi)$.

Proof. By (1), (3) and (4),

$$(\mathcal{J}_\gamma\mathcal{R}f)(x) = \frac{1}{2\pi i}\int_\gamma \sum_{j=1}^3 d\alpha^j \int_{\mathbb{R}} \frac{\partial f(\alpha t + x)}{\partial x^j} dt. \tag{6}$$

Since

$$\hat{f}(\xi) = \frac{1}{(2\pi)^{3/2}}\int_{\mathbb{R}^3} e^{-ix\cdot\xi} f(x)dx,$$

$$f(x) = \frac{1}{(2\pi)^{3/2}}\int_{\mathbb{R}^3} e^{ix\cdot\xi}\hat{f}(\xi)d\xi,$$

we have

$$\sum_{j=1}^3 \frac{\partial f(\alpha t+x)}{\partial x^j} d\alpha^j$$
$$= \frac{i}{(2\pi)^{3/2}}\sum_{j=1}^3 \left(\int_{\mathbb{R}^3} e^{i(\alpha t+x)\cdot\xi}\hat{f}(\xi)\xi^j d\xi\right) d\alpha^j.$$

Back to (6), we have

$$(\mathcal{J}_\gamma\mathcal{R}f)(x)$$
$$= \frac{1}{(2\pi)^{5/2}}\int_{\mathbb{R}} dt \int_\gamma \langle \xi, d\alpha\rangle \int_{\mathbb{R}^3} e^{i(\alpha t+x)\cdot\xi}\hat{f}(\xi)d\xi \tag{7}$$
$$= \frac{1}{(2\pi)^{3/2}}\int_\gamma \int_{\mathbb{R}^3} e^{i(\alpha t+x)\cdot\xi}\hat{f}(\xi)\delta(\langle\xi,\alpha\rangle)d\xi \wedge \langle \xi, d\alpha\rangle.$$

Here we used the fact that

$$\frac{1}{2\pi}\int_{\mathbb{R}} e^{i\alpha\cdot\xi t} dt = \delta(\langle\xi,\alpha\rangle).$$

Therefore, by (7) $\mathcal{J}_\gamma\mathcal{R}$ is a pseudo-differential operator with symbol

$$C(\xi) = \int_\gamma \delta(\langle\xi,\alpha\rangle)\langle\xi, d\alpha\rangle = \frac{1}{2\pi}\int_{\mathbb{R}} dt \int_\gamma e^{i\alpha\cdot\xi t}\langle\xi, d\alpha\rangle.$$

We want to prove that $C(\xi) = \mathrm{Crf}_\gamma(\xi)$ for $\xi \in \mathbb{R}^3 - \{O\}$.

To that effect, we only need to prove $C(\xi) = \mathrm{Crf}_\gamma(\xi)$ for a small local piece of γ. Now, we can use a local coordinate system for γ. Let s, $s_1 \le s \le s_2$, be a variable. For each oriented line $h(s) \in \gamma$ denote by $\alpha(s)$ its basis, depending on s smoothly. Set $u(s) = \langle\xi, \alpha(s)\rangle$. Then

$$C(\xi) = \int_{s_1}^{s_2} \delta\big(u(s)\big)\frac{du}{ds} ds = \frac{1}{2\pi}\int_{\mathbb{R}} dt \int_{s_1}^{s_2} e^{itu(s)}\frac{du}{ds} ds.$$

If γ does not intersect G_ξ, $u(s)$ is never zero, and hence

$$\int_{s_1}^{s_2} \delta\big(u(s)\big)\frac{du}{ds} ds = 0.$$

If γ intersects G_ξ once, from

$$\int_{s_1}^{s_2} \delta\big(u(s)\big)\frac{du}{ds} ds = \int_a^b \delta(u)du,$$

where $a = u(s_1) = \langle\xi, \alpha(s_1)\rangle$ and $b = u(s_2) = \langle\xi, \alpha(s_2)\rangle$, we get $C(\xi) = 1$ if $a < 0 < b$, and $C(\xi) = -1$ if $a > 0 > b$, which is exactly $\mathrm{Crf}_\gamma(\xi)$.

Alternatively, we may compute

$$C(\xi) = \frac{1}{2\pi}\int_{\mathbb{R}} dt \int_{s_1}^{s_2} e^{itu(s)}\frac{du}{ds} ds$$
$$= \frac{1}{2\pi}\int_{\mathbb{R}} dt \int_a^b e^{itu} du = \frac{1}{2\pi i}\int_{\mathbb{R}} \frac{e^{ibt} - e^{iat}}{t} dt, \tag{8}$$

which is a Froullani integral. Eqn.(8) is $= 0$ if a and b are of the same sign, $= 1$ if $a < 0 < b$, and $= -1$ if $a > 0 > b$. Consequently, we see $C(\xi) = \mathrm{Crf}_\gamma(\xi)$, and Theorem 1 follows. ■

4. INERSION FORMULA

First, we have a corollary.

Corollary. (Gel'fand and Graev [1]) *For two oriented 1D submanifolds γ_1 and γ_2, the symbol of the pseudo-differential operator $(\mathcal{J}_{\gamma_1}\mathcal{R})(\mathcal{J}_{\gamma_2}\mathcal{R})$ is*

$$C(\xi) = \mathrm{Crf}_{\gamma_1}(\xi)\mathrm{Crf}_\gamma\ (\xi).$$

Proof. This is the composition law of pseudo-differential operators. In fact, by (5) we have

$$C_1(D)C_2(D)u(x)$$
$$= \frac{1}{(2\pi)^{3/2}} \int_{\mathbb{R}^3} e^{ix\cdot\xi} C_1(\xi)(C_2(D)u)^\wedge(\xi)d\xi$$
$$= \frac{1}{(2\pi)^{3/2}} \int_{\mathbb{R}^3} e^{ix\cdot\xi} C_1(\xi)C_2(\xi)\hat{u}(\xi)d\xi. \quad \blacksquare$$

A 1D submanifold $\gamma \subset H_0$ is called a quasicycle if $\left|\mathrm{Crf}_\gamma(\xi)\right|$ is a constant function for almost every ξ. In particular, γ is a quasicycle if $|\gamma \cap G_\xi| = 1$ for almost all $\xi \in \mathbb{R}^3 - \{O\}$. This is the case when γ is a smooth curve on S^2 whose end points are diametrically opposite: $\gamma(s_2) = -\gamma(s_1)$. For a quasicycle γ, denote the constant $\left|\mathrm{Crf}_\gamma(\xi)\right|$ by $c(\gamma)$. Note that in this case $\mathrm{Crf}_\gamma(\xi)$ itself is not a constant in general.

By the Corollary, if γ is a quasicycle, then the symbol of $\left(\mathcal{J}_\gamma\mathcal{R}\right)^2$ is

$$\left(\mathrm{Crf}_\gamma(\xi)\right)^2 = \left|\mathrm{Crf}_\gamma(\xi)\right|^2 = c(\gamma)^2,$$

which is a constant. A pseudo-differential operator with a constant symbol $c(\gamma)^2$ is indeed $c(\gamma)^2E$, where E is the identity operator, by the Fourier inversion formula. Therefore, we have proved an inversion formula for the x-ray transform $\mathcal{R}$.

Theorem 2. (Gel'fand and Graev[1]) *If γ is a quisicycle in H_0, then $\left(\mathcal{J}_\gamma\mathcal{R}\right)^2 = c(\gamma)^2E$. Thus, for the integral transform $f \mapsto \varphi = \mathcal{R}$f, one has the inversion formula*

$$\mathcal{J}_\gamma\mathcal{R}\mathcal{J}_\gamma\varphi = c(\gamma)^2 f. \quad (9)$$

We remark that (9) is essentially the BPF reconstruction formula for any scanning curve that is a quasicycle.

5. HILBERT TRASNFORM

Back to (7) and (8), we have

$$(\mathcal{J}_\gamma\mathcal{R}f)(x) = \frac{1}{(2\pi)^{3/2}} \int_{\mathbb{R}^3} e^{ix\cdot\xi} \hat{f}(\xi)d\xi \times \frac{1}{2\pi i} \int_{\mathbb{R}} \left(e^{it\langle\xi,\alpha(s_2)\rangle} - e^{it\langle\xi,\alpha(s_1)\rangle}\right)\frac{dt}{t}.$$

By the Fourier inversion formula,

$$(\mathcal{J}_\gamma\mathcal{R}f)(x) = \frac{1}{2\pi i}\mathrm{PV}\int_{\mathbb{R}} \left(f(\alpha(s_2)t + x) - f(\alpha(s_1)t + x)\right)\frac{dt}{t} \quad (10)$$

which shows a Hilbert transform. Our computation in (8) shows that the symbol of $\mathcal{J}_\gamma\mathcal{R}$ is

$$C(\xi) = \frac{1}{2}(\mathrm{sgn}\langle\xi,\alpha(s_2)\rangle - \mathrm{sgn}\langle\xi,\alpha(s_1)\rangle).$$

Here the PV integral takes Cauchy principal values.

After [1], Eqn. (10) was re-proved by Rullgård in the case of exponential Radon transform for single photon emission computed tomography [5], and also in the standard and non-standard helical cone-beam CT cases [7-10, 21-24]. Note that as in [10], two boundary terms will appear in (10) if the projection (1) is replaced by an integral on $[0, \infty)$. Important related results were also reported [25-32]. For more details, please refer to [33].

6. NUMERICAL SIMULATION

Because the Gel'fand and Graev reconstruction formula has major implications for biomedical tomography, we performed a numerical study to evaluate its computational effects and reconstruction results. Our simulator was implemented in MatLab. To demonstrate its validity, we performed several numerical tests assuming a circular scanning locus of radius 57.0 cm in fan-beam geometry. The object was a modified Shepp-Logan phantom in a compact support with a radius of 10.0 cm. We used an equi-spatial virtual detector array of length 20.0 cm. The detector was centered at the system origin and made perpendicular to the direction from the system origin to the x-ray source. For each of our selected numbers of projections over a full-scan range, we equi-angularly acquired the corresponding projection dataset. We compared reconstructions obtained using the Gel'fand and Graev formula with the involved pseudo-differential operators and our previously implemented BPF software, respectively. Our results are in agreement but with different numerical textures and noise performance.

7. DISCUSSIONS AND CONCLUSION

The Gel'fand and Graev formula is a reconstruction strategy valid not only in the real 3D world but also higher dimensions. Numerical implementation of this formula in higher dimensions would be an interesting undertaking and may have practical applications. Potentially, this approach could also be a powerful modeling tool for some complicated phenomena.

The Gel'fand and Graev formula assumes that the Radon transform is not truncated. Recently, our group and others are actively working on interior tomography to deal with truncated projections[13, 34-35]. A possibility is that we may extend the Gel'fand and Graev formula into an "interior" reconstruction framework. We are investigating along this direction.

In conclusion, we have interpreted the Gel'fand and Graev formula in the 3D setting. Hopefully, the deep insights and novel tools developed by them may inspire further work in the CT field. This case study also shows the importance of mathematics for tomographic imaging, the synergy between mathematicians and engineers, and the practical benefits of a pure theoretical finding, which, at the first glance, may appear quite irrelevant to daily healthcare.

8. ACKNOWLEDGEMENT

This work is partially supported by NIH/NIBIB grants (EB002667, and EB011785).

9. REFERENCES

[1] I.M. Gel'fand, and M.I. Graev, *Crofton function and inversion formulas in real integral geometry.* Functional Analysis and its Applications, 1991. 25(1): p. 1-5.

[2] I.M. Gel'fand, and S.G. Gindikin, *Nonlocal inversion formulas in real integral geometry.* Functional Analysis and its Applications, 1977. 11: p. 173-179.

[3] I.M. Gel'fand, M.I. Graev, and Z.Y. Shapiro, *Integral geometry on manifolds of k-dimensional planes.* Soviet Mathematics - Doklady, 1966. 7: p. 801-804.

[4] I.M. Gel'fand, M.I. Graev, and N.Y. Vilenkin, *Generalized functions. Vol. 5: Integral geometry and representation theory*. 1966, New York-London: Academic Press.

[5] H. Rullgard, *An explicit inversion formula for the exponential radon transform using data from 180 ~.* Ark. Mat.,, 2004. 42: p. 353-362.

[6] Y. Zou, and X.C. Pan, *An extended data function and its generalized backprojection for image reconstruction in helical cone-beam CT.* Phys. Bio. Med., 2004. 49(22): p. N383-N387.

[7] Y.B. Ye, S.Y. Zhao, H.Y. Yu, and G. Wang, *Exact reconstruction for cone-beam scanning along nonstandard spirals and other curves*. *Proceedings of SPIE,Vol. 5535, p:293-300*, 2004, Denver, CO, United States.

[8] S.Y. Zhao, H.Y. Yu, and G. Wang. *A family of analytic algorithms for cone-beam CT. Proceedings of SPIE, Vol. 5535, p:318-328*, 2004, Denver, CO, United States.

[9] J.D. Pack, F. Noo, and R. Clackdoyle, *Cone-beam reconstruction using the backprojection of locally filtered projections.* IEEE Transactions on Medical Imaging, 2005. 24(1): p. 70-85.

[10] F. Noo, R. Clackdoyle, and J.D. Pack, *A two-step Hilbert transform method for 2D image reconstruction.* Phys. Bio. Med., 2004. 49(17): p. 3903-3923.

[11] M. Defrise, et al., *Truncated Hilbert transform and image reconstruction from limited tomographic data.* Inverse Problems, 2006. 22(3): p. 1037-1053.

[12] Y. Zou, X.C. Pan, and E.Y. Sidky, *Image reconstruction in regions-of-interest from truncated projections in a reduced fan-beam scan.* Phys. Bio. Med., 2005. 50(1): p. 13-27.

[13] Y.B. Ye, H.Y. Yu, Y.C. Wei, and G. Wang, *A general local reconstruction approach based on a truncated Hilbert transform.* International Journal of Biomedical Imaging, 2007. Article ID: 63634, 8 pages.

[14] Y.B. Ye, H.Y. Yu, and G. Wang, *Exact interior reconstruction with cone-beamCT.* International Journal of Biomedical Imaging, 2007. Article ID: 10693, 5 pages.

[15] H.Y. Yu, Y.B. Ye, and G. Wang, *Local reconstruction using the truncated Hilbert transform via singular value decomposition.* Journal of X-Ray Science and Technology, 2008. 16(4): p. 243-251.

[16] H. Kudo, et al., *Tiny a priori knowledge solves the interior problem in computed tomography.* Phys. Bio. Med., 2008. 53(9): p. 2207-2231.

[17] M. Courdurier, et al., *Solving the interior problem of computed tomography using a priori knowledge.* Inverse Problems, 2008. 24, Article ID 065001 , 27 pages.

[18] Y.B. Ye, H.Y. Yu, and G. Wang, *Exact interior reconstruction from truncated limited-angle projection data.* International Journal of Biomedical Imaging, 2008. Article ID: 427989, 6 Pages.

[19] Y.B. Ye, H.Y. Yu, and G. Wang, *Interior tomography: Mathematical analysis*, in *Biomedical mathematics: Promising directions in imaging, therapy planning, and inverse problems*, Y. Censor, M. Jiang, and G. Wang, Editors. 2010, Medical Physics Publishing.

[20] G. Wang, H.Y. Yu, and Y.B. Ye, *Interior tomography: Practical applications*, in *Biomedical mathematics: Promising directions in imaging, therapy planning, and inverse problems*, Y. Censor, M. Jiang, and G. Wang, Editors. 2010, Medical Physics Publishing.

[21] Y. Zou, and X.C. Pan, *Exact image reconstruction on PI-lines from minimum data in helical cone-beam ct.* Phys. Bio. Med., 2004. 49(6): p. 941-959.

[22] Y.B. Ye, S.Y. Zhao, H.Y. Yu, and G. Wang, *A general exact reconstruction for cone-beam CT via backprojection-filtration.* IEEE Transactions on Medical Imaging, 2005. 24(9): p. 1190-1198.

[23] T.L. Zhuang, et al., *Fan-beam and cone-beam image reconstruction via filtering the backprojection image of differentiated projection data.* Phys. Bio. Med., 2004. 49(24): p. 5489-5503.

[24] S.Y. Zhao, H.Y. Yu, and G. Wang, *A unified framework for exact cone-beam reconstruction formulas.* Medical Physics, 2005. 32(6): p. 1712-1721.

[25] A. Katsevich, *Theoretically exact filtered backprojection-type inversion algorithm for spiral CT.* SIAM J. Appl. Math., 2002. 62(6): p. 2012-2026.

[26] A. Katsevich, *A general scheme for constructing inversion algorithms for cone-beam CT.* Int. J. of Math. and Math. Sci., 2003. 21: p. 1305-1321.

[27] A. Katsevich, *An improved exact filtered backprojection algorithm for spiral computed tomography.* Advances in Applied Mathematics, 2004. 32(4): p. 681-697.

[28] J.D. Pack, and F. Noo, *Cone-beam reconstruction using 1D filtering along the projection of m-lines.* Inverse Problems, 2005. 21(4): p. 1105-1120.

[29] Y.B. Ye, and G. Wang, *Filtered backprojection formula for exact image reconstruction from cone-beam data along a general scanning curve.* Medical Physics, 2005. 32(1): p. 42-48.

[30] Y. Zou, X. Pan, and E.Y. Sidky, *Theory and algorithms for image reconstruction on chords and within regions of interest.* J. Opt. Soc. Am. A, 2005. 22(11): p. 2372-2384.

[31] T.L. Zhuang, and G.H. Chen, *New families of exact fan-beam and cone-beam image reconstruction formulae via filtering the backprojection image of differentiated projection data along singly measured lines.* Inverse Problems, 2006. 22(3): p. 991-1006.

[32] H. Y.Yu, and G. Wang, *Cone-beam composite-circling scan and exact image reconstruction for a quasi-short object.* International Journal of Biomedical Imaging, 2007. Article ID: 87319, 10 pages.

[33] G. Wang, Y.B. Ye, and H.Y. Yu, *Approximate and exact cone-beam reconstruction with standard and non-standard spiral scanning.* Phys. Bio. Med., 2007. 25: p. R1-R13.

[34] H.Y. Yu, and G. Wang, *Compressed sensing based interior tomography.* Phys. Bio. Med., 2009. 54(9): p. 2791-2805.

[35] H.Y. Yu, J.S. Yang, M. Jiang and G. Wang, *Supplemental analysis on compressed sensing based interior tomography.* Phys. Bio. Med., 2009. 54(18): p. N425-N432.

Evaluations of a 2D inverse Hilbert Transform

Harald Schöndube*, Karl Stierstorfer*, and Frédéric Noo†

***Abstract*—We evaluate a new two-dimensional inverse Hilbert transform formula for the use in reconstruction algorithms based on the differentiated backprojection (DBP). As opposed to using traditional one-dimensional inverse Hilbert transform approaches this new formula would allow greater flexibility in the arrangement of the backprojection grid. In this paper, we introduce the new formula and compare its performance with using a one-dimensional approach, especially in terms of their discrete implementations. Since our focus is specifically on evaluating the properties of the inverse Hilbert transform, which are independent from the data acquisition geometry, we limit our considerations to the two-dimensional parallel beam geometry. The application of our results to other geometries, such as helical cone-beam CT, is straightforward.**

I. Introduction

The differentiated backprojection (DBP) with a subsequent inverse Hilbert transform (HT) [1]–[4] has recently received attention as a theoretically-exact and stable (TES) reconstruction method for helical cone-beam computed tomography (CT) [5]–[8]. The method appears especially attractive because of its capability to perform a TES reconstruction using all measured data at arbitrary pitch [9], [10].

In contrast to standard FBP-type algorithms, the DBP does not yield a reconstruction of the object function $f(\underline{x})$ directly, but rather the HT of $f(\underline{x})$ along M-lines, where an M-line is any line that connects a source position with a point on the detector. To achieve the reconstruction of $f(\underline{x})$, an inverse HT has to be applied after the DBP is performed. In this paper, we focus on this second part of the two-step DBP-HT method.

In all implementations of the DBP-HT method published so far, the direction of the M-lines had to be aligned with either the rows or the columns of the image for the inverse HT to be performed. For helical cone-beam CT, this means that at least one interpolation from the M-line geometry to Cartesian coordinates has to be performed after application of the inverse HT to obtain a reconstruction result in Cartesian coordinates [6], [9]; however, this interpolation comes at the cost of a possible loss of resolution. Furthermore, depending on the exact formulation of the DBP chosen for backprojection, an arrangement of the backprojection geometry such that the image rows and columns are perpendicular to the x and y axis of the Cartesian coordinates, respectively, may be advantageous [6]. In this case two interpolation steps are necessary to obtain a reconstruction in Cartesian coordinates, one before and one after the inverse HT.

Manuscript received February 1st, 2010. This work was supported in part by the U.S. National Institutes of Health (NIH) under grant R21 EB009168, and in part by Siemens Healthcare. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

*Siemens Healthcare, Computed Tomography, Forchheim, Germany
†Department of Radiology, University of Utah, Salt Lake City, USA
Contact: harald.schoendube@siemens.com, noo@ucair.med.utah.edu

In this paper, we examine the use of a new approach of performing the inverse Hilbert transform. This approach is based on a two-dimensional filtering in Fourier domain and does not require the image to be aligned with the direction of the M-lines. The paper is organized as follows: In section II a short description of the DBP-HT method and the one-dimensional inverse HT employed in [6], [9], [10] will be given. Section III introduces our new two-dimensional approach. In section IV an evaluation of the new method and a comparison with the classical one-dimensional approach will be presented. A discussion of our results in section V concludes this paper.

Since our focus is specifically on evaluating the properties of the inverse Hilbert transform, which are independent from the data acquisition geometry, we limit our considerations and evaluations in this paper to the two-dimensional parallel beam geometry. The application of our results to other geometries, such as helical cone-beam CT, is straightforward.

II. The DBP-HT method

A. The differentiated backprojection

Let $g(\vartheta, r)$ denote the sinogram of the two-dimensional object function $f(\underline{x})$, i.e.,

$$g(\vartheta, r) = (\mathcal{R}f)(\vartheta, r) = \int_{-\infty}^{\infty} f(r\underline{\theta} + t\underline{\theta}^{\perp})\,\mathrm{d}t, \qquad (1)$$

where $(\mathcal{R}f)(\vartheta, r)$ denotes the Radon transform of $f(\underline{x})$ and where $\vartheta \in [0, 2\pi)$ is an angle that defines the two unit vectors $\underline{\theta}(\vartheta) = [\cos(\vartheta), \sin(\vartheta)]^T$ and $\underline{\theta}^{\perp}(\vartheta) = [-\sin(\vartheta), \cos(\vartheta)]^T$. Note that both of those vectors depend on ϑ as indicated, but we usually drop this explicit dependence to enhance readability of the equations. Furthermore, r denotes a scalar with $r \in [-R_{\mathrm{FOV}}, R_{\mathrm{FOV}}]$, where R_{FOV} denotes the radius of the CT field-of-view.

It can then be shown that the following expression for the two-dimensional DBP in parallel beam geometry holds [3]:

$$\int_{\vartheta_1}^{\vartheta_1+\pi} g'(\vartheta, \underline{x}\cdot\underline{\theta})\,\mathrm{d}\vartheta = -2\!\!\!\!\!\!\int_{-\infty}^{\infty} \frac{1}{t} f(\underline{x} - t\underline{\theta}_1^{\perp})\,\mathrm{d}t, \qquad (2)$$

where

$$g'(\vartheta, r) := \frac{\partial}{\partial r}\, g(\vartheta, r) \qquad (3)$$

and $\underline{\theta}_1^{\perp}$ is a short-hand notation for $\underline{\theta}^{\perp}(\vartheta_1)$. Analogously we will later also use $\underline{\theta}_1$ as a short-hand notation for $\underline{\theta}(\vartheta_1)$. Note that ϑ_1 can assume any value in equation 2, since $g(\vartheta, r)$ is 2π-periodic in ϑ. The integral on the right hand side of equation 2 can be interpreted as applying a Hilbert transform along direction $\underline{\theta}_1^{\perp}$ to $f(\underline{x})$ as can be seen when considering a rotated (s, τ) coordinate system, i.e., performing the change

of variables $\underline{x} = s\underline{\theta}_1 + \tau\underline{\theta}_1^\perp$:

$$-\!\!\!\!\!\!\int_{-\infty}^{\infty} \frac{1}{t} f(\underline{x} - t\underline{\theta}_1^\perp)\,\mathrm{d}t = \pi -\!\!\!\!\!\!\int_{-\infty}^{\infty} \frac{1}{\pi t} f\left(s\underline{\theta}_1 + (\tau - t)\underline{\theta}_1^\perp\right)\,\mathrm{d}t \tag{4}$$

$$= \pi(\mathcal{H}f)(s\underline{\theta}_1 + \tau\underline{\theta}_1^\perp, \underline{\theta}_1^\perp) \tag{5}$$

$$= \pi(\mathcal{H}f)(\underline{x}, \underline{\theta}_1^\perp), \tag{6}$$

where the integral sign with a dash denotes an integral that has to be interpreted in the sense of a Cauchy principal value [11].

While this application of the Hilbert transform, which is a one-dimensional transform, to a two-dimensional image may seem counterintuitive, the concept can be understood when interpreting the image as consisting of one-dimensional functions which are combined together to form the two-dimensional image. In other words, the Hilbert transform of a two-dimensional image $f(\underline{x})$ along the direction of $\underline{\theta}_1^\perp$ can be interpreted as a union of independent one-dimensional Hilbert transforms. In that sense, $(\mathcal{H}f)(\underline{x}, \underline{\theta}_1^\perp)$ denotes the value at point $\underline{x}$ of the Hilbert transform of $f(\underline{x})$ along the line of direction $\underline{\theta}_1^\perp$ and containing $\underline{x}$. Equation 2 therefore yields

$$\int_{\vartheta_1}^{\vartheta_1+\pi} g'(\vartheta, \underline{x}\cdot\underline{\theta})\,\mathrm{d}\vartheta = -2\pi \cdot (\mathcal{H}f)(\underline{x}, \underline{\theta}_1^\perp). \tag{7}$$

B. The finite inverse Hilbert transform

As described by equation 7 the DBP does not yield the object function $f(\underline{x})$ directly, but rather its Hilbert transform along a unit vector $\underline{\theta}_1^\perp$. To achieve a reconstruction of $f(\underline{x})$ we thus need to find a way to compute $f(\underline{x})$ from $(\mathcal{H}f)(\underline{x}, \underline{\theta}_1^\perp)$. While the HT of a function $k(t)$ is its own inverse up to a factor of -1, this direct approach can only be applied if $(\mathcal{H}k)(t)$ is known for all $t \in (-\infty, \infty)$, since the HT of a bounded function is generally unbounded. In our DBP-HT implementations we use a modified, more practical, approach called *finite inverse Hilbert transform*, which was first published by Söhngen in 1937 [12] and which yields $f(\underline{x})$ from $(\mathcal{H}f)(\underline{x}, \underline{\theta}_1^\perp)$ in a theoretically-exact and stable way if $f(\underline{x})$ is strictly contained within the FOV, i.e., if $f(\underline{x}) = 0\ \forall\ \underline{x} \in \{\underline{x}\,|\|\underline{x}\| \geq R_{\mathrm{FOV}}\}$.

Using the rotated coordinate system with change of variables $\underline{x} = s\underline{\theta}_1 + \tau\underline{\theta}_1^\perp$ Söhngen's formula reads

$$f(s\underline{\theta}_1 + \tau\underline{\theta}_1^\perp) = \frac{1}{\sqrt{t_{\mathrm{R}}(s)^2 - \tau^2}}\left[C(s) - -\!\!\!\!\!\!\int_{-t_{\mathrm{R}}(s)}^{t_{\mathrm{R}}(s)} \frac{(\mathcal{H}f)(s\underline{\theta}_1 + t\underline{\theta}_1^\perp, \underline{\theta}_1^\perp)}{\pi(\tau - t)} \sqrt{t_{\mathrm{R}}(s)^2 - t^2}\,\mathrm{d}t\right], \tag{8}$$

where $t_{\mathrm{R}}(s) = \sqrt{R_{\mathrm{FOV}}^2 - s^2}$. The function $C(s)$ can be computed directly from the sinogram data according to the equation

$$C(s) = \frac{\int_{-\infty}^{\infty} f(s\underline{\theta}_1 + \tau\underline{\theta}_1^\perp)\,\mathrm{d}\tau}{\pi}. \tag{9}$$

Note that equation 8 only requires knowledge of $(\mathcal{H}f)(\underline{x}, \underline{\theta}_1^\perp)$ over that part of the line which goes through $\underline{x}$ along the direction of $\underline{\theta}_1^\perp$ that is inside the FOV. It thus allows to compute $f(\underline{x})$ from $(\mathcal{H}f)(\underline{x}, \underline{\theta}_1^\perp)$ as obtained by the DBP according to equation 7, namely by performing a one-dimensional finite inverse Hilbert transform along the line of direction $\underline{\theta}_1^\perp$ which contains $\underline{x}$. If the application of a 1D inverse HT is desired, the best choice for ϑ_1 is thus such that the resulting filtering direction $\underline{\theta}_1^\perp$ is along one of the axes of the desired reconstruction grid, as in this case the inverse HT can be performed directly along either the rows or the columns of the backprojection result.

C. Implementing the inverse HT in 1D

Apart from some weighting functions, for which the implementation for discrete data is straightforward, the finite inverse Hilbert transform according to equation 8 essentially consists of a convolution with $\frac{1}{\pi t}$ and of finding the value of $C(s)$. As shown in equation 9, the latter can directly be computed from the sinogram (or fanogram) data; in practice this step is implemented using a (bi-) linear interpolation. The convolution is performed as a discrete convolution of the band-limited Hilbert kernel

$$\hat{h}_{\mathrm{hilb}}(t) = -\int_{-\nu_g}^{\nu_g} (-i\,\mathrm{sign}(\nu))e^{i2\pi\nu t}\,\mathrm{d}\nu \tag{10}$$

with $(\mathcal{H}f)(s\underline{\theta}_1 + t\underline{\theta}_1^\perp, \underline{\theta}_1^\perp)$.

One important property of this implementation is that if $\Delta t = \Delta r$, the straightforward application of equation 10 results in severe *Gibb's* or *ringing* artifacts in the resulting image (see figure 1), whereas those artifacts disappear if a half-pixel shift is employed in the discrete convolution (i.e., using $\hat{h}_{\mathrm{hilb}}(t + \Delta t/2)$ as a convolution kernel instead of $\hat{h}_{\mathrm{hilb}}(t)$) [3]. An explanation of this somewhat surprising behavior has been given in [14]: the authors examined the behavior of $\hat{h}_{\mathrm{hilb}}(t)$ in Fourier domain and found that, apart from the jump discontinuity at $\nu = 0$, which is intrinsic to the Hilbert transform, this kernel also features jump discontinuities in Fourier domain at $\nu = \pm\nu_g$, which induce ringing artifacts. In contrast, these discontinuities can be avoided if a half-pixel shift is included into the Hilbert kernel, thus avoiding the buildup of ringing artifacts. In earlier investigations we found that another way of removing these artifacts is to increase the sampling of the backprojected image such that $\Delta r = 2\Delta t$ (see again figure 1) [15]. In this case $(\mathcal{H}f)(s\underline{\theta}_1 + t\underline{\theta}_1^\perp, \underline{\theta}_1^\perp)$ does not contain any frequency component near $\pm\nu_g = \pm\frac{1}{2\Delta t}$ and the artifacts disappear[1].

III. A 2D INVERSE HT FOR THE DBP-HT METHOD

As discussed in the introduction, in some cases a filtering direction $\underline{\theta}_1^\perp$ which is neither aligned with the rows now with the columns of the backprojected image is preferred. We thus have evaluated the use of a new two-dimensional inverse Hilbert transform formula. To perform the inverse HT using this new formula, the one-dimensional convolution with

[1]Note that in the example in figure 1 we actually changed the sampling of the sinogram instead of the backprojection grid to obtain images of equal size, therefore causing a difference in resolution.

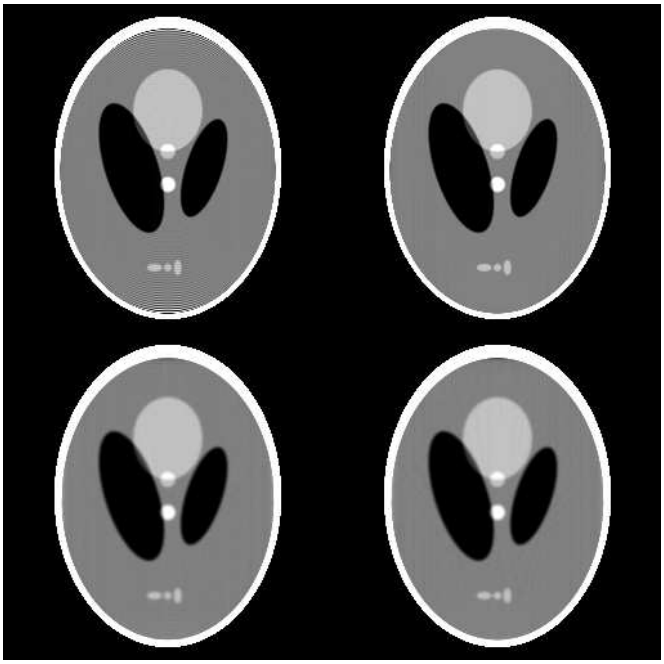


Fig. 1: DBP-HT reconstructions of the Shepp-Logan phantom [13] from sinogram data such that $\Delta r = \Delta t$ (upper row) and such that $\Delta r = 2\Delta t$ (lower row), each without (left column) and with (right column) a half-pixel shift applied during the inverse HT.

$h_{\mathrm{hilb}}(t)$ is replaced by a two-dimensional filtering step using the relation

$$\int_{-\infty}^{\infty} \frac{(\mathcal{H}f)(s\underline{\theta}_1 + t\underline{\theta}_1^{\perp}, \underline{\theta}_1^{\perp})}{\pi(\tau - t)}\,\mathrm{d}t = \iint_{\mathbb{R}^2} (\mathcal{F}_2\mathcal{H}f)(\underline{\nu}, \underline{\theta}_1^{\perp})(-i\,\mathrm{sign}(\underline{\nu}\cdot\underline{\theta}_1^{\perp}))e^{i2\pi\underline{\mathrm{x}}\cdot\underline{\nu}}\,\mathrm{d}\underline{\nu}, \quad (11)$$

where $(\mathcal{F}_2\mathcal{H}f)(\underline{\nu}, \underline{\theta}_1^{\perp})$ denotes the two-dimensional Fourier transform of $(\mathcal{H}f)(\underline{\mathrm{x}}, \underline{\theta}_1^{\perp})$. A proof of this equation is given in the appendix.

IV. Evaluations

The 2D filtering approach can be implemented as a two-dimensional discrete convolution with a filtering kernel $\hat{h}_{\mathrm{hilb},2}(\underline{\mathrm{x}})$ following similar steps to the ones in the 1D case discussed above. A reconstruction result using such a direct implementation of the 2D filtering with parameters $\vartheta_1 = 45°$ and $\Delta x = \Delta y = \Delta r$ is displayed in figure 2a. As could be expected from the 1D inverse HT results discussed in section II-C, strong ringing artifacts appear in the reconstructed image. These ringing artifacts can be eliminated when incorporating a shift by half a pixel in both x and y directions into the two-dimensional discrete convolution (see figure 2b). It is, however, not possible to find a formulation of the discrete convolution kernel that ensures that the continuation of the kernel's periodically repeated 2D Fourier spectrum is smooth everywhere for arbitrary filtering directions $\underline{\theta}_1^{\perp}$. Thus, there will always exist some values of ϑ_1 for which ringing artifacts can not be suppressed by incorporating a shift into the discrete convolution or a similar measure. This is exemplified by the reconstruction result displayed in figure 2c, for which a filtering direction corresponding to $\vartheta_1 = 15°$ was chosen.

One way to work around this problem is to use a backprojection grid with a higher pixel density such that $\Delta x = \Delta y = 0.5\Delta r$ holds. As in the 1D case, this measure effectively removes the ringing artifacts from the reconstructed image (see figures 3a and 3b). However, backprojection time increases significantly, making this solution not very attractive. One possible alternative to performing a backprojection on a denser pixel grid that might come to mind is performing the backprojection on the original grid and then interpolating to the denser grid. However, while this approach can reduce ringing artifacts compared to the direct non-shifted approach, the resulting image quality is not as good as when performing the backprojection directly onto the dense grid (see figure 3c for an example using a spline interpolation).

Another way of suppressing ringing artifacts is to enforce a smooth continuation of the periodically repeated 2D Fourier kernel spectrum by applying a frequency apodization to the kernel. In our implementation we achieve this apodization by using a two-dimensional generalization of the Hamming window (figure 4a). While this method works well in terms of eliminating the ringing artifacts, the associated resolution loss is quite high; the resulting resolution is comparable with the resolution that one can obtain by backprojecting onto a rotated grid, then performing a 1D inverse HT and finally interpolating to the Cartesian coordinates (cf. figures 4b and 4c). However, as discussed in the introduction, a backprojection onto a rotated grid is not always practical, and one can still obtain an increase in resolution by using the apodized 2D inverse HT instead of having to perform the interpolation step two times.

V. Discussion

In this paper, we have presented a new approach for implementing the inverse Hilbert transform in the context of the DBP-HT reconstruction method for CT. Our approach involves replacing the one-dimensional discrete convolution that is usually used to implement Söhngen's formula for a finite inverse Hilbert transform by a two-dimensional filtering step. For certain cases, in which performing the differentiated backprojection onto a rotating grid is impractical, using our new method constitutes a significant implementation advantage. However, as our evaluations have shown, the avoidance of ringing artifacts is more difficult when using the two-dimensional filtering approach than with the 1D discrete convolution, as there is no formulation of the discrete convolution kernel that ensures that the continuation of the kernel's periodically repeated 2D Fourier spectrum is smooth everywhere for arbitrary filtering directions $\underline{\theta}_1^{\perp}$.

The most stable approaches to circumvent the problem of ringing artifacts in our method are to either use a dense backprojection grid such that $\Delta x = \Delta y = 0.5\Delta r$ or to apply a two-dimensional generalization of the Hamming window as a frequency apodization. Both of those approaches might not always be satisfactory solutions, as they either involve a higher computational load for backprojection or a substantial resolution loss. However, the resulting resolution when using the

(a) $\vartheta_1 = 45°$, no shift (b) $\vartheta_1 = 45°$, half-pixel shift (c) $\vartheta_1 = 15°$, half-pixel shift

Fig. 2: Reconstructions of the Shepp-Logan phantom using the 2D inverse Hilbert transform approach.

(a) $\vartheta_1 = 45°$, double size backprojection grid (b) $\vartheta_1 = 15°$, double size backprojection grid (c) $\vartheta_1 = 15°$, spline stretching

Fig. 3: Reconstructions of the Shepp-Logan phantom using the 2D inverse Hilbert transform approach with decreased grid spacing.

(a) $\vartheta_1 = 15°$, Hamming apodization (b) $\vartheta_1 = 15°$, BP onto a rotated grid (c) difference between (a) and (b)

Fig. 4: Reconstructions of the Shepp-Logan phantom using the 2D inverse HT with apodization and using a backprojection onto a rotated grid with a 1D inverse HT.

Hamming apodization is still comparable with the one that can be obtained by performing the backprojection onto a rotated grid and then interpolating to Cartesian coordinates after a 1D inverse HT. Furthermore, in cases where a backprojection onto a rotating grid is not practical, one can still obtain an increase in resolution by applying the 2D inverse HT with Hamming apodization instead of using an algorithm which involves two interpolation steps.

APPENDIX

Proof of equation 11: Using a generic two-dimensional function $k(\underline{\mathrm{x}})$, we have

$$\begin{aligned}
&(\mathcal{H}k)(s\underline{\theta}_1 + \tau\underline{\theta}_1^{\perp}, \underline{\theta}_1^{\perp}) \\
&\quad = \int_{-\infty}^{\infty} \frac{k(s\underline{\theta}_1 + t\underline{\theta}_1^{\perp})}{\pi(\tau - t)}\,\mathrm{d}t \\
&\quad = \int_{-\infty}^{\infty} \frac{1}{\pi(\tau - t)} \iint_{\mathbb{R}^2} (\mathcal{F}_2 k)(\underline{\nu}) e^{i2\pi(s\underline{\nu}\cdot\underline{\theta}_1 + t\underline{\nu}\cdot\underline{\theta}_1^{\perp})}\,\mathrm{d}\underline{\nu}\,\mathrm{d}t \\
&\quad = \iint_{\mathbb{R}^2} (\mathcal{F}_2 k)(\underline{\nu}) e^{i2\pi s\underline{\nu}\cdot\underline{\theta}_1} \\
&\qquad \cdot \int_{-\infty}^{\infty} \frac{1}{\pi(\tau - t)} e^{i2\pi(t-\tau)\underline{\nu}\cdot\underline{\theta}_1^{\perp}} e^{i2\pi\tau\underline{\nu}\cdot\underline{\theta}_1^{\perp}}\,\mathrm{d}t\,\mathrm{d}\underline{\nu},
\end{aligned}$$

which using $\underline{\mathrm{x}} = s\underline{\theta}_1 + \tau\underline{\theta}_1^{\perp}$ finally results in

$$\begin{aligned}
&(\mathcal{H}k)(\underline{\mathrm{x}}, \underline{\theta}_1^{\perp}) \\
&\qquad = \iint_{\mathbb{R}^2} (\mathcal{F}_2 k)(\underline{\nu})(-i\,\mathrm{sign}(\underline{\nu}\cdot\underline{\theta}_1^{\perp})) e^{i2\pi\underline{\mathrm{x}}\cdot\underline{\nu}}\,\mathrm{d}\underline{\nu}.
\end{aligned}$$

REFERENCES

[1] Y. Zou and X. Pan, "Exact image reconstruction on PI-line from minimum data in helical cone-beam CT," *Phys. Med. Biol.*, vol. 49, pp. 941–959, 2004.

[2] ——, "An extended data function and its backprojection onto PI-lines in helical cone-beam CT," *Phys. Med. Biol.*, vol. 49, pp. N383–N387, 2004.

[3] F. Noo, R. Clackdoyle, and J. Pack, "A two-step Hilbert transform method for 2D image reconstruction," *Phys. Med. Biol.*, vol. 49, pp. 3903–3923, 2004.

[4] J. Pack, F. Noo, and R. Clackdoyle, "Cone-beam reconstruction using the backprojection of locally filtered projections," *IEEE Trans. Med. Imag.*, vol. 24, no. 1, pp. 70–85, Jan. 2005.

[5] L. Yu, D. Xia, Y. Zou, E. Sidky, J. Bian, and X. Pan, "A rebinned backprojection-filtration algorithm for image reconstruction in helical cone-beam CT," *Phys. Med. Biol.*, vol. 52, pp. 5497–5508, 2007.

[6] H. Schöndube, K. Stierstorfer, F. Dennerlein, T. White, and F. Noo, "Towards an efficient two-step Hilbert algorithm for helical cone-beam CT," in *Proc. 2007 Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine (Lindau, Germany)*, F. Beekman and M. Kachelrieß, Eds., 2007, pp. 120–123.

[7] H. Schöndube, K. Stierstorfer, F. Dennerlein, and F. Noo, "Comparative evaluation of two analytical methods for helical cone-beam tomography," in *IEEE Nuclear Science Symposium Conference Record NSS '07*, vol. 6, 2007, pp. 4467–4471.

[8] F. Sureau and M. Defrise, "Discretization of the two-step hilbert transform for truncated data reconstruction," in *Proc. 2009 Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine (Beijing, China)*, Sep. 2009, pp. 182–185.

[9] H. Schöndube, K. Stierstorfer, and F. Noo, "Accurate helical cone-beam CT reconstruction with redundant data," *Phys. Med. Biol.*, vol. 54, no. 15, pp. 4625–4644, Aug. 2009.

[10] ——, "Exact efficient handling of interrupted illumination in helical cone-beam computed tomography with arbitrary pitch," 2010, to appear in Tsinghua Science and Technology (Fully3D 2009 Special Issue).

[11] P. K. Kythe and P. Puri, *Computational Methods for Linear Integral Equations*. Boston, Basel, Berlin: Birkhäuser, 2002.

[12] H. Söhngen, "Die Lösungen der Integralgleichung $g(x) = \frac{1}{2\pi}\int_{-a}^{a} \frac{f(\xi)}{x-\xi} d\xi$ und deren Anwendung in der Tragflügeltheorie," *Math. Z.*, vol. 45, pp. 245–264, 1937.

[13] L. A. Shepp and B. F. Logan, "Reconstructing interior head tissue from x-ray transmissions," *Nuclear Science, IEEE Transactions on*, vol. 21, no. 1, pp. 228–236, Feb. 1974.

[14] A. Faridani, R. Hass, and D. C. Solmon, "Numerical and theoretical explorations in helical and fan-beam tomography," *Journal of Physics: Conference Series*, vol. 124, p. 012024 (20pp), 2008. [Online]. Available: http://stacks.iop.org/1742-6596/124/012024

[15] F. Dennerlein, F. Noo, H. Schöndube, G. Lauritsch, and J. Hornegger, "A factorization approach for cone-beam reconstruction on a circular short-scan," *IEEE Transactions on Medical Imaging*, vol. 27, no. 7, pp. 887–896, Jul. 2008.

X-ray Imaging with 50nm Spatial Resolution for Nanotomography

Peter Bruyndonckx, Alexander Sasov, Xuan Liu and Bart Pauwels

***Abstract*—We have demonstrated that structures down to ~50 nm can be visualized in x-ray projection images using a nano-focus x-ray source. This was achieved using a novel target structure to produce a very small but stable x-ray focal spot which can still emit sufficient x-rays. Due to their unlimited depth of focus, nano-focus x-ray sources don't pose a limit on the specimen size that can be imaged in 3D x-ray nanotomography.**

I. Introduction

X-ray optics such as Fresnel zone plates are often used to obtain a spatial resolution better than a few 100 nm in x-ray projection images. Despite the fact that they can achieve extreme high resolutions [1], Fresnel zone plates are not always suited for tomographic imaging due to their very limited depth of focus which restricts the size of specimen to be imaged to a few microns [2]. Alternatively, we looked into the possibility to use a set-up based on a nano-focus x-ray source as a way to prevail over this limitation while still retaining a nanometer scale resolution.

Resolving extremely small details in an object requires both a good image resolution and a high signal-to-noise ratio. The former depends on the size of the x-ray focal spot while the latter is given by the x-ray flux produced in the target of the x-ray generator. These two requirements are usually conflicting, i.e. to achieve the smallest possible diameter of the focal spot size, the electron beam current in the electron gun has to be reduced. To maintain the same number of detected x-rays, the exposure time has to be increased. This makes the systems more prone to instabilities of the electron beam positioning.

Also, when structures with dimensions below a few hundred nm are being imaged using x-rays of a few keV, Fresnel diffraction effects start to play a role. These can cause image contrast to be destroyed and render structures much larger than the focal spot size invisible.

To overcome these two limitations, both the target structure to generate the x-rays and the imaging configuration were optimized.

II. Experimental setup

Figure 1 shows a schematic drawing how a JEOL JSM-7000F scanning electron microscope (SEM) with a Schottky field emission electron gun was used for x-ray imaging. A specimen holder is mounted on a rotation stage. The distance from the specimen to the target can be varied. The target is mounted at a 45 degree angle relative to the electron beam. The produced x-rays penetrating through the sample are acquired by a cooled CCD camera with direct photon detection (PIXIS-XO 512, Princeton Instruments, USA). This camera has 512x512 pixels and a pixel size of 24 µm. The electron gun is operated at 30 kV. The target can be coated with a thin substrate of a specific material to generate characteristic x-rays with appropriate energies.

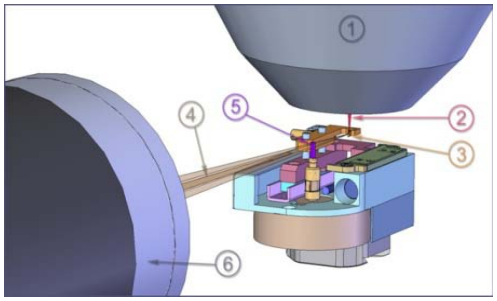


Fig. 1: Schematic drawing of the SEM attachment that converts a SEM into a X-ray tomography system. The electron optics (1) focuses a narrow beam (2) onto a 45° target (3), producing a x-ray beam (4) passing though a specimen on a rotation stage at a variable distance from the target (5). The x-rays are detected by a cooled CCD camera (6).

III. Diffraction limited imaging

To measure the resolution we placed an XRadia resolution pattern X50-30-2 (XRadia, USA) in the specimen holder. The resolution pattern has 180 nm thick Au structures varying in size down to 50 nm. The target consisted of a 200 nm silver substrate layer deposited on 50 µm beryllium. Figure 2 shows an x-ray projection image of a part of the resolution pattern. The enlarged portion of the resolution pattern shows that the image contrast completely disappears at spatial frequencies corresponding to details of 300 nm. At even higher spatial frequencies the image contrast is inversed.

The loss and inversion of image contrast at high spatial frequencies seems to indicate that Fresnel diffraction effects become important when x-rays pass through the resolution pattern. To verify this, the electromagnetic field strength in the detector plane was computed when an x-ray cone beam passes through a 1D grid pattern. To compute the field strength U(x,y) we use the following equation [3]:

$$U(x,y) = A\frac{e^{ik(z_1+z_2)}}{i\lambda z_1 z_2} e^{i\frac{\pi}{\lambda z_2}(x^2+y^2)} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (Obj(\xi,\eta)\, e^{i\frac{\pi(z_1+z_2)}{z_1 z_2}(\xi^2+\eta^2)}) e^{-i\frac{2\pi}{\lambda z_2}(x\xi+y\eta)} d\xi d\eta \quad (1)$$

where Obj(ξ,η) is the object transfer function (i.e. the

absorption and phase shift of the X-rays when passing through the object), z_1 is the target to specimen distance and z_2 is the specimen to detector distance. The influence of the x-ray focal spot size and the 24 µm CCD pixel size is taken into account by convolving U(x,y) with a Gaussian x-ray spot distribution and a 24 µm wide block function.

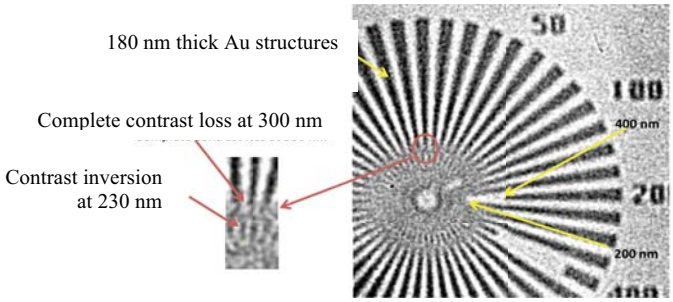


Fig. 2: X-ray projection image of an XRadia X50-30-2 resolution pattern. The enlarged portion shows a complete loss of contrast around 300 nm.

The energy spectrum emitted by the Ag thin film target, as seen by the detector, was computed using WinXray software package [4] and subsequently corrected for the energy depended detection efficiency curve (fig 4 top). The obtained spectrum is rather mono-energetic around 3 keV. Figure 3 shows the computed response to a regular 1D grid patterns with different periods. The object transfer function assumes 180 nm thick Au walls. The red lines show the ideal response while the blue lines correspond to the expected detector response taking Fresnel diffraction, the X-ray spot size and detector pixel size into account. The complete loss of image contrast around 290 nm corresponds with what we observed in the measured x-ray projection image.

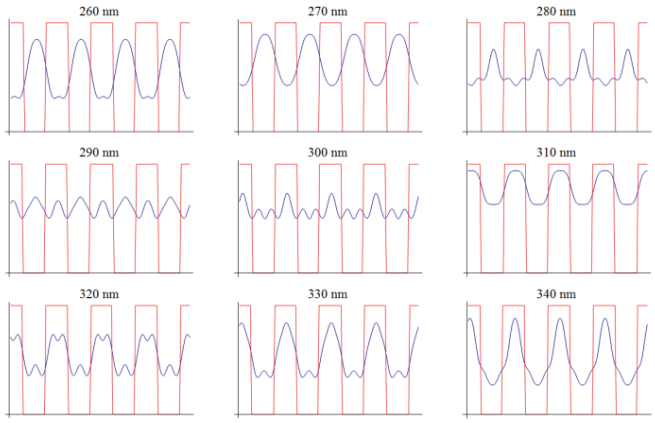


Fig. 3: Ideal (red) and predicted (blue) detector response to a regular 1D grid as a function of the structure width.

IV. TARGET MATERIAL AND DEFOCUSING DISTANCE

From equation (1) one finds there are two instrumental ways to push diffraction limiting effects to higher spatial frequencies: increase the mean x-ray energy and reduce the defocusing distance $z=(z_1*z_2)/(z_1+z_2)$.

The former can be achieved by choosing a different target material which produces characteristic x-rays at higher energies, e.g. replace Ag (L_α: 2.98 keV, L_β : 3.15 keV) coating with Au (L_α: 9,71 keV, L_β: 11.44 keV, M_α: 2.2 keV). The Au spectrum as seen by the camera is less mono-energetic than the Ag spectrum because of the reduce detection efficiency at higher x-ray energies, i.e. ~15% at 10 keV compared to ~85% at 3 keV (fig 4 bottom). Hence, the average energy of the detected x-rays is only 6.4 keV.

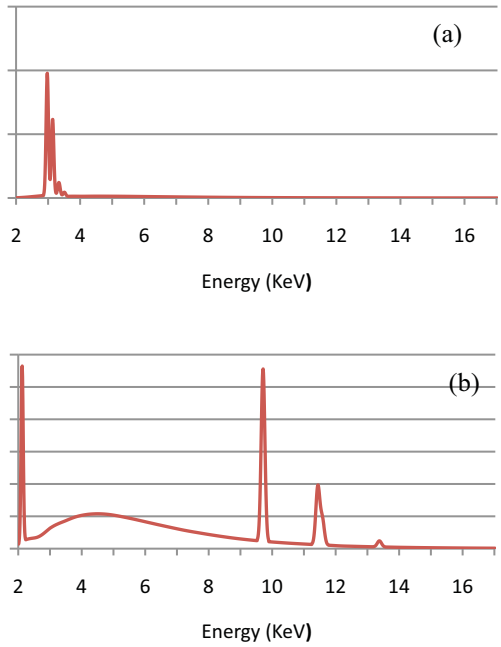


Fig. 4: Simulated x-ray spectrum [5] as measured by the CCD when either an Ag (top) or an Au (bottom) target substrate is hit by 30 keV electrons. The Ag spectrum is rather mono-energetic around 3 keV while the Au exhibits two strong peaks at 2.2 keV and 9.7 keV. The average energy of the Au spectrum is 6.4 keV.

Since $z_1 << z_2$ in our set-up, it follows that z can be approximated by z_1, i.e. the target to specimen distance. The defocusing distance was reduced by tilting the target from a 45° angle to a 75° angle such that the distance to the resolution pattern can be lowered from 0.7 mm to about 0.3 mm. To improve contrast at these higher energies the XRadia resolution pattern was replaced with an ATN/XRESO-50HC (NTT-AT Nanofabrication Co., Japan) resolution pattern. The latter has 500 nm thick Ta structures. Figure 5 shows a drawing of the central part of the NTTAT resolution pattern (left) and the resulting x-ray projection image. The resolution in the horizontal direction is better than the vertical resolution due to the tilt of the target. Structures of 200 nm are now still clearly visible. The details in the resolution pattern get wiped out when the sizes are less than ~150 nm.

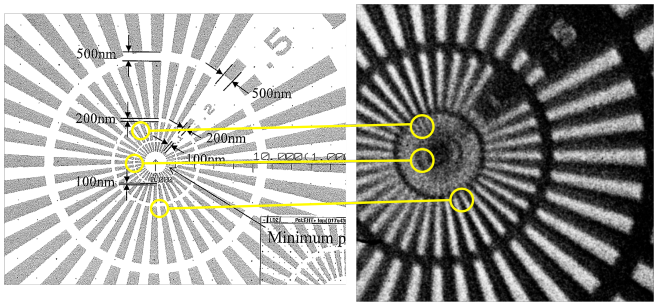


Fig. 5: Schematic drawing (left) and X-ray projection image (right) of the central portion of an ATN/XRESO-50HC (NTT-AT Nanofabrication Co.) resolution pattern.

V. Novel Target Structure

Using a thin target layer deposited on a flat substrate to produce x-rays prevents us from reaching very small (<100 nm) x-ray focal spot sizes with a high x-ray flux. This is due to the finite volume in the target material where x-rays are produced. In addition, the long scan times also have a detrimental effect due to possible thermal instability during the exposure. Therefore a novel target structure was invented. The planar target at 45° is replaced by a 250µm Pt(90%)/Ir(10%) rod with a sharp tip of ~50 nm (fig. 6). The axis of the rod is positioned perpendicular to the CCD surface. If an electron beam is positioned on the tip of the Pt/Ir rod, the CCD will see this as a ~50 nm x-ray spot. Because the x-ray focal spot size is now determined by the physical dimensions of the target tip, the electron beam current can be increased and small movements of the electron beam during the exposure can be tolerated without any influence to the spatial resolution. Using a rod tip as target also allows us to position the specimen closer to the x-ray focal spot, thereby reducing the defocusing distance.

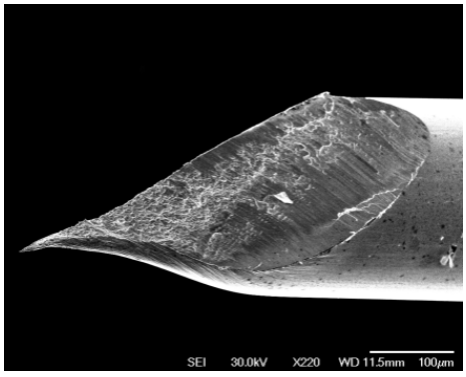


Fig. 6: SEM Image of the Pt/Ir target tip.

VI. Sub-100 nm Imaging

The new target structure described above was mounted under the electron beam of our SEM. The ATN/XRESO-50HC resolution pattern was placed at a distance of ~75 µm from the target tip. This distance was computed from the image magnification and the known target to detector distance. To further improve the detection quantum efficiency, the back-thinned CCD was replaced with a new fully depleted CCD (S10747-0909, Hamamatsu, Japan) which has a 250µm thick photosensitive structure. This camera has 512x512 pixels (24 µm pixel size) and an efficiency above 80% up to energies of 10 keV. The improved sensitivity at higher x-ray energies also results in a higher average energy of the detected x-rays and hence helps further reducing the Fresnel diffraction effects.

Figure 7 shows an image of the resolution pattern (top) and a close-up view of the central part (bottom). The ring with 100 nm structures is now perfectly visible. In some regions the 50 nm structures in the inner ring are also visible. The loss of resolution in certain radial directions is due to the fact that the home-made target tip is not perfectly cylindrical.

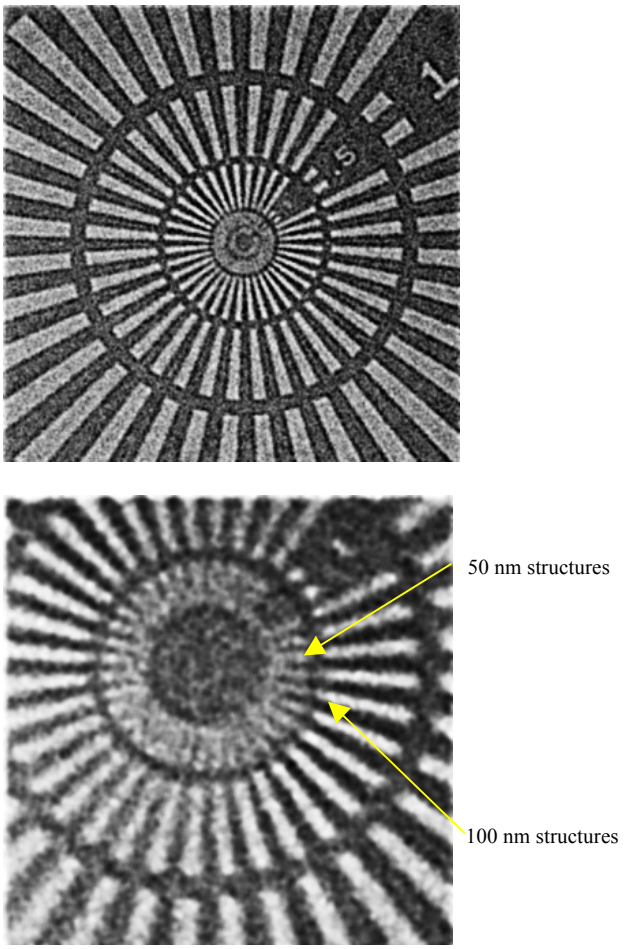


Fig. 7: (Top) x-ray projection image of the ATN/XRESO-50HC resolution pattern, (Bottom) Close up of the central part of the same resolution pattern.

VII. Nano-CT Implementation

A micro fossil was mounted on the rotation stage in the SEM for tomographic imaging. Figure 8a shows a projection image of the micro fossil acquired with an image pixel size of 138 nm. To obtain 3D tomographic images, projections were acquired at 0.9° intervals. Due to limited mechanical accuracies of the rotation stage, sample movements have been observed. These movements can be approximated as horizontal and vertical shifts in the projection space and can be corrected accordingly. These random shifts were estimated

by an iterative algorithm that fits measured projections to estimated projections from the reconstruction [4]. Once the magnitude of the X-Y shifts for each recorded projection is taken into account, an image can be reconstructed using the Feldkamp filtered back-projection algorithm. The axial and coronal slice through the 3D reconstructed space (figure 8b and 8c) clearly show internal channel structure. Figure 8d shows a 3D rendering of the micro fossil with the front-top corner virtually removed.

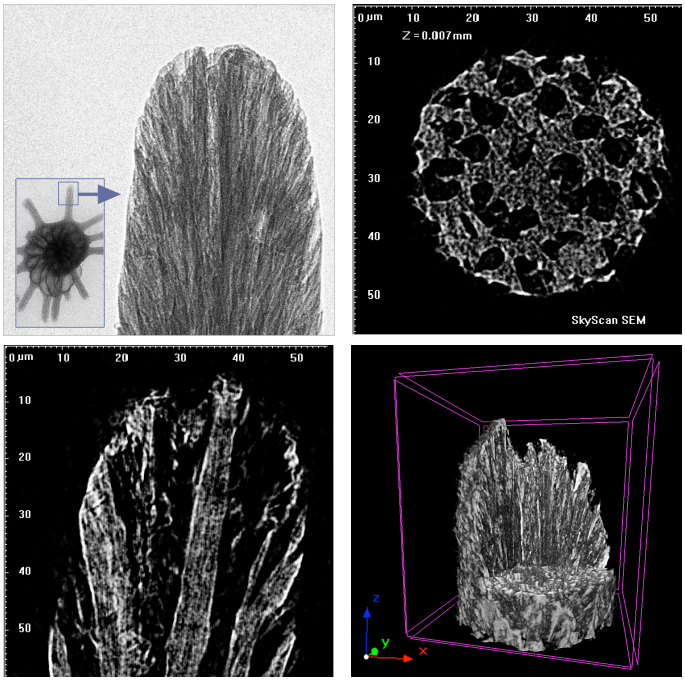


Fig. 8: A. (top-left): High resolution projection image showing a part of the micro fossil. The inset shows a low resolution projection image of the complete micro fossil. B. (top-right): Axial reconstructed slice, C. (bottom-left): Coronal reconstructed slice, D. (bottom-right): 3D rendering of the micro fossil with a virtually removed front-top corner.

VIII. Conclusion

Imaging bulk samples at sub-100 nm resolution poses a number of challenges. The usage of X-ray optics such as Fresnel zone plates limits the object to specimen with very small diameters (< few microns) in order to keep all projection images in focus. This prevents non-destructive imaging of a larger specimen.

To overcome this limitation we investigated the possibility of optimizing an imaging set-up based on a nano-focus x-ray source. The most important points of this development are:

1) Diffraction limited imaging can be overcome by choosing an appropriate target material which produces high energy characteristic x-rays and by reducing the distance between the x-ray focal spot and the specimen. This can be combined with the employment of a fully depleted CCD camera to optimize the sensitivity for the high energy x-rays and hence also further increase the average energy detected.

2) Using a nano-tip target structure allows creating a very small x-ray focal spot and become insensitive to thermal instabilities of the electron beam during long exposures. It also generates a larger x-ray flux compared to the conventional planar thin film target.

3) Small random mechanical movements during tomographic imaging are compensated in the image reconstruction process by estimating the magnitude of the shifts using an iterative procedure.

We have demonstrated that using this set-up 50 nm structures can be resolved in projection images. The infinite depth-of-focus allows any object to be completely in focus for all projections. This is essential for 3D tomographic imaging. The nano-scale 3D tomographic imaging capabilities were demonstrated on a microfosil using an isotropic voxel size of 138 nm.

References

[1] Chu, Y.S. (2008). "Hard x-ray microscopy with Fresnel zone plates reaches 40 nm Rayleigh resolution", Appl. Physics letters, 92, 103119

[2] Suzuki, Y. (2008). "Making hard x-ray micro-focus beam with Fresnel zone plate optics", Spring-8 summer school text

[3] Cowley, J.M. (1990), "Diffraction physics" (North-Holland, Amsterdam)

[4] Demers, H. (2004). Win X-ray, Version 1.2.1.13, http://montecarlomodeling.mcgill.ca/software/winxray/winxray.html

[5] Sasov, A., Liu X. and Salmon P. " Compensation of mechanical inaccuracies in micro-CT and nano-CT", Developments in x-ray tomography N°6 (San Diego 2008) , Proceedings of SPIE, vol. 7078, pp. 70781C.1-70781C.9

A compact CT geometry: theory and practice

Samit Basu and Jed D. Pack

I. Introduction

There are several practical factors that motivate the design and implementation of so-called compact CT geometries. The primary motivation is in space constrained applications (e.g., mobile or retrofit installations) where a conventional third-generation geometry simply requires too much space to achieve a given working field of view. Recall that in a third-generation geometry, the detector arc is focused at the focal spot, and thus achieving a given field of view requires that the outer diameter of the gantry must be significantly larger than the field of view itself. Traditional CT installations in medical or NDE applications are typically not highly space constrained, as other factors such as patient handling, the scan table, or the assembly line in the case of NDE, often dictate the space envelope required to site a CT system. In security application, however, CT systems generally are installed in one of two modes: as inline systems in new installations, and as retrofit systems in existing airports. In the former case, space is generally not at a premium, and conventional geometry CT systems have proliferated there. In the latter case, however, CT systems are being installed where no prior system existed (e.g., in the lobby of an airport, or in a checkpoint lane). In these cases, space is at a significant premium, and a conventional CT system based on a classic third-generation geometry may be difficult to site. Additional installation options (such as placing CT systems behind check-in counters) may not even be possible with a conventional third-generation system, as the required space simply does not exist, and reworking the check-in counters to make the space available may prove impractical or economically feasible.

By contrast, fourth-generation CT systems, which featured a complete ring of isocenter focused detector modules and a rotating source, could (in principle) be much more compact than their third-generation cousins. The outer diameter of a fourth-generation system is dictated only by the physical components themselves, and not by the geometry of the imaging system. In principle, a fourth-generation CT system can be made to be only slightly larger than the imaging field of view, allowing for the physical size of the X-ray source and the detectors. While we do not know of any fourth-generation CT systems still in production, many stationary-type CT systems are essentially fourth-generation in design, and feature an iso-center focused ring of detectors and a generally iso-center focused ring of sources as well. In addition to being compact, fourth-generation CT systems, have other advantages including the fact that ring artifacts do not generally occur in fourth-generation systems as no single detector contributes to a ring of pixels in the image). Unfortunately, these systems struggled with scatter, detector area, and ultimately cost.

As a compromise, compact CT systems are typically designed with detectors that are placed on some type of modified arc which reduces the required outer diameter of the system. In this paper, we outline such a compact geometry CT system which provides a significantly smaller external envelope than a corresponding conventional third-generation system. Our implementation features a highly uniform resolution across the field of view, and excellent image quality. We discuss the history of compact CT geometries, describe some of the attributes of a compact design, and ultimately show images from prototype systems that were built using a compact geometry.

II. Compact CT and Parallel Beam

With the exclusion of fourth-generation and stationary geometries (which are a topic in and of themselves), compact CT geometries have been described in various forms for the past 30 years. The earliest reference we could find to a compact CT geometry is the patent from Hounsfield [1]. In 1977, Hounsfield proposed a novel CT geometry, not because of a desire to create a more compact CT system, but rather as a means of optimizing the sampling of the data for parallel beam rebinning of the data. Recall that for a conventional third-generation detector geometry, the data are typically sampled at equal angles. When mapped to parallel beam samples using the classic equation:

$$s_i = D\sin(\alpha_i) \tag{1}$$

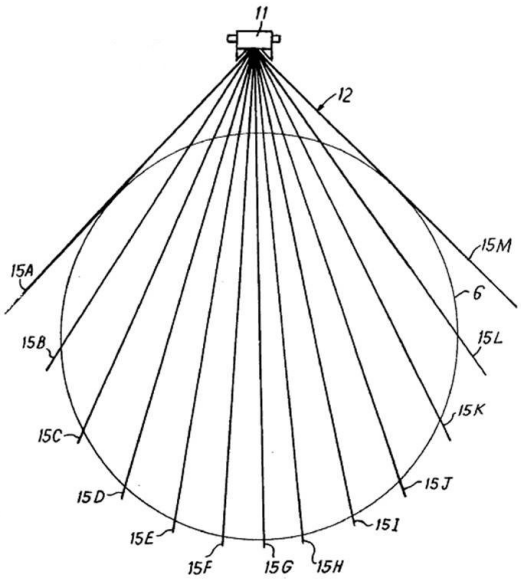
a uniform sampling in α_i leads to a decidedly non-uniform sampling in s_i.

The usual solution to this problem is to reinterpolate the set of s_i to a uniform set of samples using some interpolation scheme. The observation made by Hounsfield and rediscovered by ourselves was that by changing the placement of the detectors, one can minimize the deviation of the s_i from a uniform set. In mathematical

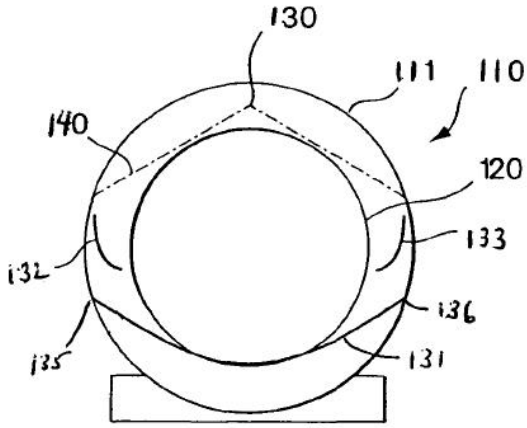
terms we are searching for a set of α_i that satisfy the constraints

1) The samples are adjacent (i.e., no gaps between the measured rays).
2) The set s_i are close to uniform.
3) The inner radius is fixed by the field of view.
4) The outer radius should be minimized (the compactness requirement).

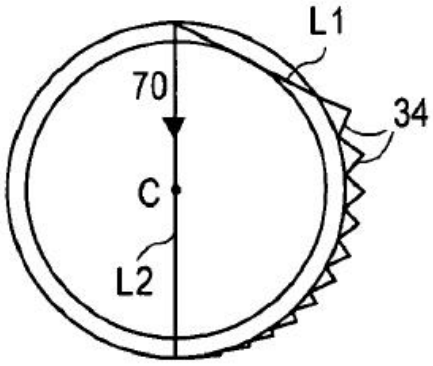
We are unaware if a closed form expression exists for the optimal set of α_i. Hounsfield did not explicitly provide a set of α_i other than in the form of a figure. However he was motivated by the desire to minimize the impact of interpolation from the non-uniform set of s_i to a uniform set for parallel beam reconstruction. As an aside, the same argument applies in 3D helical scanning (or else this would be a purely intellectual exercise) with algorithms that require some type of pseudo-parallel beam or cone-parallel rebinning in their processing. Of course item 2 requires a suitable definition of "close to uniform", and we used metrics in our own design process to select the positions and orientation of the detectors. The actual optimization is also subject to mechanical constraints and constraints on manufacturability, but those details are likely of little interest to this audience. The system developed by our team most closely resembles the one in [1].



Of interest is also the "folded" CT geometry used by others and disclosed in [2]. This geometry is focused primarily on the compactness problem, and less on the issue of parallel beam resampling. Note that in this geometry, the detector is at two different magnifications, and is thus split into three pieces: a center piece labeled 131 in the figure which is a conventional focally-aligned third-generation type detector, and two peripheral detectors 132 and 133 which capture fan angles not measured by the central detector 131. According to the patent literature, the detectors at the periphery of the gantry are chosen so as to present a uniform α_i spacing to the reconstruction algorithm. On the other hand, the outer diameter of the gantry, which is driven by the outermost set of detectors, is reduced in their geometry. Apart from fourth-generation CT systems, the systems described in [2] are the only other commercially available (and deployed) CT systems that use a compact CT geometry.



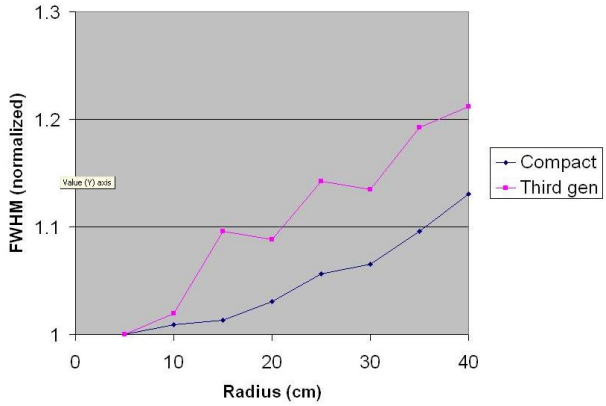
Another proposed compact geometry was presented in [3]. This geometry was effectively a half-fan geometry in which approximately half of the field of view was illuminated by the source, and a complete rotation provided data sufficient for reconstruction. It bears significant similarities to the concept in [1], and in one incarnation had two sources and two detectors for full FOV coverage. We include it for completeness.



III. Resolution Uniformity

One of the interesting consequences of the compact geometry is a potential increase in the uniformity of the system resolution across the field of view. Unlike medical systems, which tend to be highly optimized towards

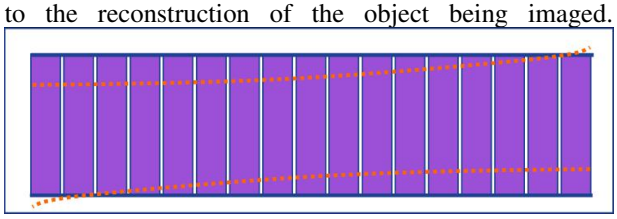
imaging near isocenter, security applications typically require high resolution uniformly throughout the field of view. Threat objects are equally likely to appear at the periphery of the field of view instead of the center, and when system resolution is critical (as in the detection of small or thin objects), a uniform resolution will result in better system performance. During the design of our own compact geometry system, we studied via simulation the in-plane resolution as a function of radius, and found that an optimized geometry lead to a slower degradation in system resolution than a conventional third-generation system with the same effective detector size and system resolution. While the exact numbers cannot be disclosed, the curves below are representative of the difference.



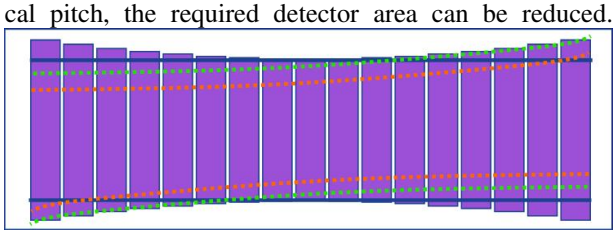
The difference in resolution between the conventional and compact geometries is almost entirely a practical issue. When rebinning to parallel beam, one must choose a Δs with which to resample the data. The conventional choice is to choose Δs so that the parallel beam rays match the fan beam rays at isocenter. This choice leads to the minimum resolution loss at isocenter (where we typically care most about the resolution). But far from isocenter, the parallel ray spacing and fan beam ray spacing begin to drift apart, and the sampling becomes suboptimal. On the other hand, if the fan angles are chosen so that $\sin(\alpha_i)$ are approximately uniform, the fan beam rays and the parallel beam rays remain aligned throughout the field of view. This choice results in a more uniform resolution.

IV. Helical Pitch

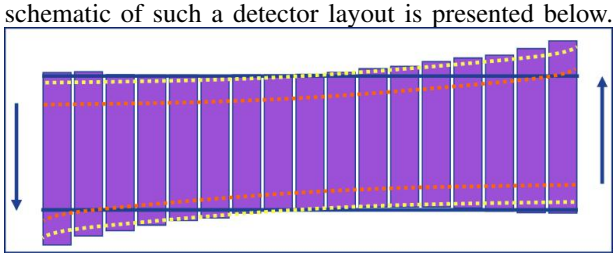
The second consequence of the compact geometry is in terms of helical pitch. Without describing the exact helical pitch used by any system, consider the problem of the mapping of the Tam window (required for exact helical cone beam reconstruction) onto a conventional third-generation volumetric (i.e., many-slice) detector. As can be seen in the following figure, the Tam boundaries, which are the projection of the source helix onto the detector, do not align with the physical boundaries of the detector. Thus to avoid missing data in the acquisition (and missing data artifacts in the reconstructed images), one must limit the helical pitch so that the Tam boundary is contained inside the detector. This limitation results in a significant portion of the detector being essentially "wasted", as the measurements from those detectors do not contribute materially to the reconstruction of the object being imaged.



For the compact geometry case, the detectors at the periphery of the field of view are closer to scan circle, resulting in a lower magnification to the focal spot. As a result, the Tam window projected onto the detector has a significantly better fit. A better fit implies that for the same physical detector area, a higher helical pitch can be used, or conversely, for a given helical pitch, the required detector area can be reduced.



In the extreme case of a helical-only scanner, one can also skew the detector along the translation direction. The skew can be chosen to maximize the intersection of the detector with the Tam window, and results in the highest possible pitch for a given detector area. A schematic of such a detector layout is presented below.



V. Results

Image results from our prototype systems will be presented at the conference.

References

1 Hounsfied, *US Patent 4160911.*
2 Ellenbogen, et al, *US Patent 7046761.*
3 Gaultier, et al, *US Patent 7027554.*

INTERIOR RECONSTRUCTION FOR MULTI-SCALE CT FACILITY

Ge Wang[1], Hengyong Yu[1], Kriti Sen Sharma[1], Chris Wyatt[2], Linbing Wang[3], Tea Andric[1], Joseph Freeman[1], Steve Wang[4], Michael Feser[4], S H Lau[4], Wenbing Yun[4]*

[1]School of Biomedical Engineering and Sciences, Virginia Tech, Blacksburg, VA 24060, USA
[2]Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA 24060, USA
[3]Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, VA 24060, USA
[4]Xradia Inc., 5052 Commercial Circle, Concord, CA 94520, USA

ABSTRACT

To quantify the structures and functions of organs, tissues, cells, and sub-cellular components, we are developing a unique multi-scale x-ray CT facility that performs image reconstruction over six orders of magnitude in length scale and object size. To handle large samples and reduce radiation dose, we have combined interior tomography and compressive sensing to improve the CT imaging capabilities. This approach does not need exact prior knowledge for precise reconstruction of an interior region-of-interest (ROI), and is demonstrated here to be valuable in an experimental CT study on a scaffold for bone regeneration.

Index Terms— Computed tomography (CT), multi-scale CT, interior tomography, compressive sampling (CS), regenerative medicine.

1. INTRODUCTION

Among modern μm/nm-scale imaging tools such as visible-light microscopy, electron microscopy, atomic force microscopy, and scanning tunneling microscopy, micro-/nano-CT fills an important gap in terms of image resolution, sample preparation, and application territory. More importantly, this capability gap is precisely what is most critical for the nano-scale studies that will be urgently needed in the near future. For example, MEMS and NEMS industries are producing components on the hundreds of nm scale, and behaviors of fuel cells are critically dependent on conditions of their nano-pores in the hundreds of nm mean size. More relevant to biomedical applications, specimens in nano-medication need x-ray nano-CT characterization.

Coincidentally, our new multi-scale CT facility is highly complementary and synergic with a variety of micro-/nano-scale imaging equipment integrated in the Nano-scale Characterization and Fabrication Lab at Virginia Tech. Hence, the noninvasive analytical capability brought by our facility will be indispensible for meeting comprehensive nano-imaging needs related research and education.

Recently, compressive sensing (CS) theory has emerged and proved that high-quality signals and images can be reconstructed from far fewer sampled data than what is usually considered necessary by the Nyquist sampling theory [1, 2]. The main idea of CS is that most signals are sparse in an orthonormal system when represented in a proper domain. Typically, CS-based algorithms recovers signal/images from a limited amount of data via the ℓ_1 norm minimization, which is also equivalent to the total variation (TV) minimization in several important cases [3].

Based on the CS theory, our group recently proved that if an object under reconstruction is piecewise constant or piecewise polynomial, a local region-of-interest (ROI) can be exactly reconstructed via the TV minimization in the CT field [4-6], which is complementary to our exact knowledge based interior tomography scheme [7-9]. Because many objects in CT/micro-CT/nano-CT applications can be approximately modeled as piecewise constant or piecewise polynomial, our approach is practically useful, and suggests a new research direction of x-ray CT.

The introduction of interior tomography overcomes the restriction that a subject/sample must stay within the field of view (FOV) of a CT system, relaxing the tight constraints on FOV and sample size [4-9], and leading to powerful, compatible and flexible multi-scale imaging with minimal distortion of the subject material.

In this paper, we will report our preliminary attempt to evaluate interior tomography for our multi-scale CT facility. In the next section, our facility and method are described. In the third section, experimental results are presented and analyzed. In the last section, related issues are discussed.

*The work is partially supported by the NSF/MRI program (CMMI-0923297), NIH/NCRR grant (RR025667) and NIH/NIBIB grants (EB002667, EB011785). Address correspondence to Ge Wang, email: ge-wang@ieee.org.

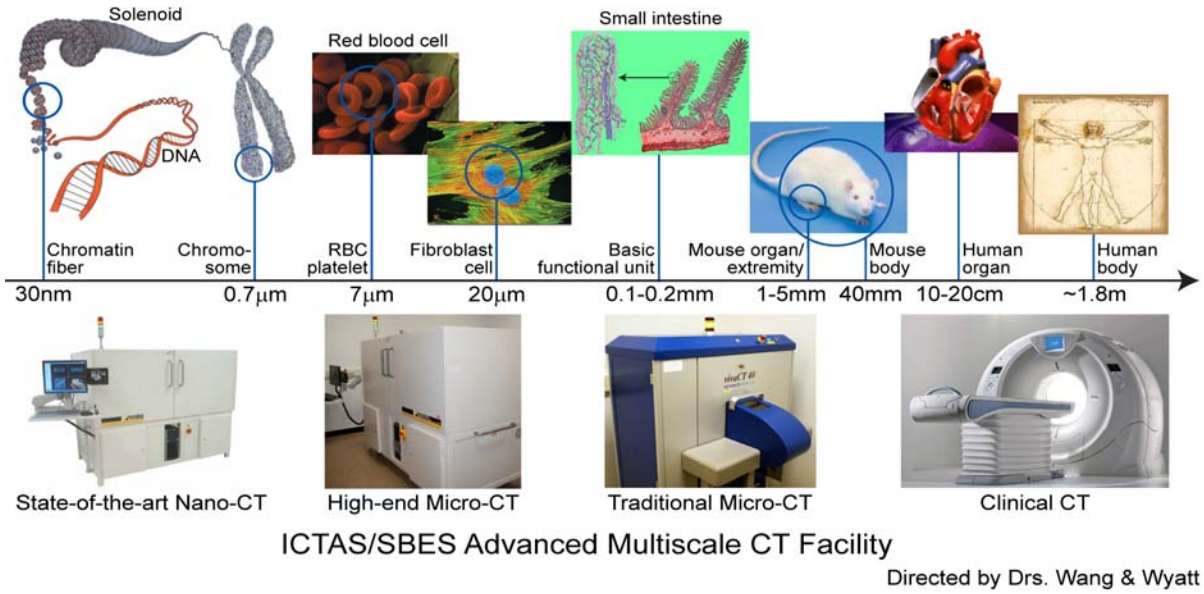


Figure 1. SBES (School of Biomedical Engineering & Sciences) Advanced Multi-scale CT Facility, comprising Xradia Nano-CT system, Xradia Micro-CT system, and a Scanco micro-CT system.

2. SYSTEM AND METHOD

2.1. Multi-scale CT Facility

Biology and medicine is a multi-scale science and demands systematic approaches. Quantifying the structures and functions of organs, tissues, cells, and sub-cellular components is challenging because the transition and fusion of information among different scales requires different systems and methods. Adopting x-ray CT as a common across-scale platform has the advantage of compatible staining and imaging techniques from mm to nm. Supported by an NIH SIG grant, in March 2009 we acquired an Xradia 500nm micro-CT system. Supported by an NSF MRI grant, in December 2009 we acquired a state-of-the-art 50nm nano-CT system from the same company. Combining with our Scanco micro-CT scanner and medical CT scanners in our collaborative hospitals, we are developing an advanced multi-scale x-ray CT facility, as illustrated in Fig. 1. To our best knowledge, there is no other imaging modality that could perform image reconstruction over six orders of magnitude in length scale and object size like x-ray CT.

Since Scanco micro-CT and clinical CT scanners are rather conventional, here we focus on Xradia micro- especially nano-CT systems, which are also referred to as x-ray microscopy. This nano-CT technology was first developed in the 1990s by two groups, one at Stony Brook University [10] and the other at Göttingen University[11]. Built on those development and other advances, Xradia made significant improvements and commercialized this technology since 2000.

X-ray nano-CT is a relatively new technique that offers unique capabilities stemming from its large penetration length in biological materials. The multi-keV x-rays used in the Xradia nanoXCT system has a 1/e attenuation length of > 0.25mm in soft tissues and over 100μm in hard tissues such as bone. As a consequence, thick tissue or bone sections can be studied with little sample preparation or modification. To date, up to 25μm spatial resolution has been demonstrated with sub-keV energy "soft" x-rays, and 50nm resolution achieved with multi-keV "hard" x-rays. Currently, the state-of-the-art nanoXCT™ platform made by Xradia is the world's only x-ray nano-CT system with tens of nm resolution that does not require a synchrotron radiation source. It resembles a conventional light microscope, consisting of an x-ray source, a condenser lens, an objective lens, and an area detector. The key features of the system include (1) penetration depths in the mm range for organic specimens of nano-medicine applications, (2) 50nm resolution that is uniform in 3D, (3) multiple imaging modalities including absorption or Zernike phase contrast modes to optimize image contrast, and (4) automated acquisition, reconstruction, and analysis.

In the Zernike phase-contrast imaging mode [12], a phase ring is inserted in the back focal plane of the objective lens in the nanoXCT™ system, which causes the affected radiation to interfere with the un-affected radiation on the image plane to produce an intensity pattern. This pattern can be expressed in terms of linear integrals. That is, for a phase object the recorded image intensity is proportional to the integrated phase shift through the object plus a constant offset. Then, we can extract the linear integral information from the measured intensity data via appropriate data preprocessing. Hence, both the absorption and Zernike phase contrast imaging can be achieved by inverting linear integral equations.

2.2. CS-based Interior Tomography

Based on the CS theory, our group recently found that a local ROI can be exactly reconstructed via the total variation minimization if an object under reconstruction is essentially piecewise constant [4, 5] or piecewise polynomial [6]. The gradient transform of $f(\rho,\theta)$ is expressed as

$$\mu(\rho,\theta)=\sqrt{\left(\frac{\partial f(\rho,\theta)}{\partial\rho}\right)^2+\left(\frac{\partial f(\rho,\theta)}{\rho\partial\theta}\right)^2} \qquad (1)$$

which is the gradient magnitude of the maximum directional derivative at (ρ,θ). Typically, the integration of $\mu(\rho,\theta)$ over the domain of interest is referred to as the total variation. We have the following theorem [4].

Theorem 1: *In the compressive sensing framework, an interior ROI of a general compactly supported function f can be exactly determined by minimizing the total variation defined if f can be decomposed into finitely many constant sub-regions.*

To perform the interior reconstruction from data collected on the aforementioned multi-scale CT facility, we implemented an alternating iterative algorithm in the CS framework [4]. Basically, this algorithm is similar to our previous version [4] with some minor modifications. Different from our previous interior tomography algorithms in terms of both projection onto convex sets (POCS) [8] and singular value decomposition (SVD) [9, 13] based on exact knowledge on a sub-region in the ROI, our CS based interior tomography approach only assumes a flexible imaging model, and is more practical and more powerful.

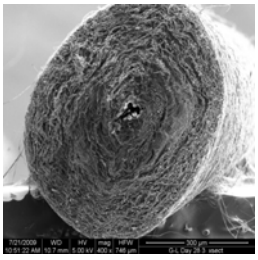
3. EXPERIMENT

3.1. Data Acquisition

Our general imaging protocol is as follows. First, an object/sample is scanned by a lower-resolution scanner to generate scout views or CT images and identify an ROI or volume-of-interest (VOI) in these views or images. Then, only the ROI/VOI is locally scanned by a higher-resolution scanner, and reconstructed by interior tomography. The procedure will be repeated until it is successful.

In this experiment, our subject is a scaffold for bone regeneration. It was fabricated using a new electro-spinning setup [14]. Briefly, poly (L-lactide) (PLLA) and poly (D-lactide) (PDLA) polymer solutions were electro-spun onto rotating poly (glycolide) (PGA) microfibers. The thickness of the nano-fiber layers around a PGA microfiber scaffold was varied by controlling the volume of the electro-spun solution. The osteon-like sections were then stacked together and wrapped with a sheet of the PLLA/PDLA electro-spun mat and placed into a 10mm cylindrical mold with 5mm diameter. These were then sintered to create the 3D scaffold.

Then, the scaffold was first scanned by the 500nm micro-CT scanner. It was placed in a 0.5ml PCR tube secured to a sample holder on the rotary stage of the system. The x-ray source voltage and current were set at 30kV and 100μA respectively. During the scan, 271 projections of 2048^2 pixels per projection were acquired between -135º and +135º. Acquisition time was ~5 hours for the specimen. Fig. 2 shows an SEM cross-section of the scaffold and an image slice reconstructed on the Xradia micro-CT scanner. Our next step is to scan this scaffold using the state-of-the-art 50nm nano-CT system to depict inner structures with finer resolution. Note that the nano-CT system will be available in November, 2009, and used immediately after its acceptance test at Virginia Tech.


(a)

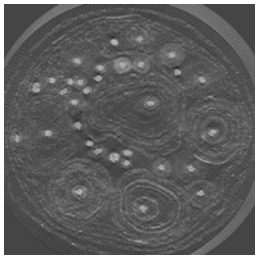

(b)

Figure 2. Osteon-like scaffold. (a) An SEM cross section of a microfiber scaffold, and (b) a reconstructed slice of the scaffold.

3.2. Reconstructed Images

Using our in-house interior tomography software, we repeated the reconstruction from the aforementioned dataset acquired by our Xradia micro-CT scanner. A typical reconstructed image slice is given in Fig. 3 (a). Because our 50nm CT system is not available at this moment, we truncated the micro-CT dataset to simulate a corresponding local nano-CT dataset of the ROI indicated in Fig. 3(a). In comparison of the magnified ROI Fig. 3 (b) from the global reconstruction Fig. 3 (a), the resultant CS-based interior reconstruction is presented in Fig. 3(c). It can be clearly observed that our interior reconstruction offers an image quality comparable to that of global reconstruction. Because the CS-based iterative reconstruction is capable of noise suppression, our CS-based interior reconstruction has a higher signal-noise-ratio (SNR) than that of the conventional reconstruction.

While a higher resolution CT scanner can always produce a better spatial resolution by irradiating the ROI and recording data on a finer detector array, the scanning time is typically increased to keep the same SNR. Because the CS theory is based on so-call sparsifying transform, the scan time can be reduced by specifying a less number of projections. However, in general, the smaller the projection number, the worse the reconstruction quality. To study how to maintain the image quality with a limited number of

projections in the CS framework, the first 180 projections were extracted from the above-mentioned dataset covering a range of 180°. Then, 90 and 60 projections were sequentially selected by discarding 1/2 and 2/3 projections respectively. For these reduced projection datasets, scan time can be accordingly shortened by 34%, 67% and 78% respectively. The results are shown in Fig. 3 (d)-(f). It can be seen that the image quality was still sufficiently good even if we only used 90 projections.

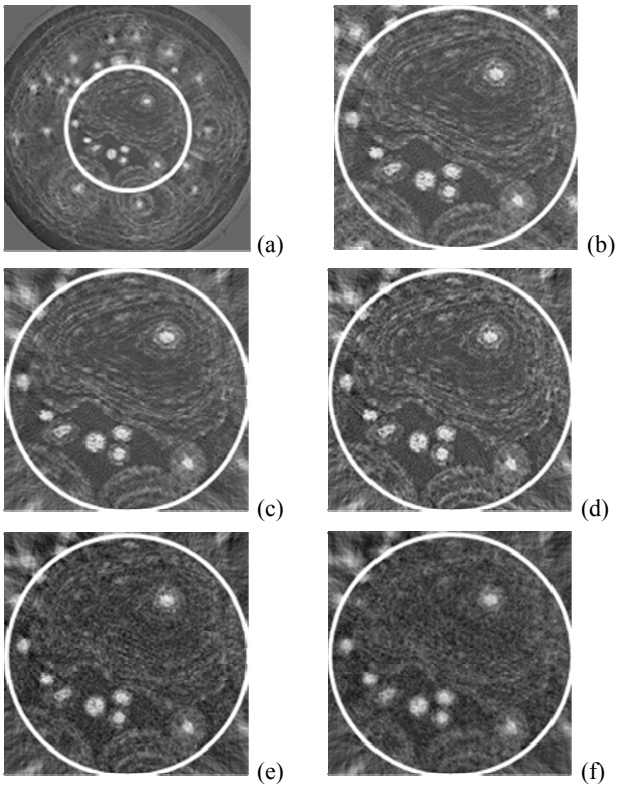


Figure 3. CS-based interior reconstructions for the slice of the microfiber scaffold shown in Fig. 2 (b). (a) The reconstruction from a global projection dataset of 270 views, (b) a magnification of the ROI indicated in (a) to serve as a benchmark, (c)-(f) CS-based interior reconstructions from truncated local datasets of 270, 180, 90 and 60 views respectively.

4. DISCUSSION AND CONCLUSION

Evaluating our results against that reconstructed using the Xradia commercial software, it is noticed that our image quality is slightly lower than the standard. This is most likely because of our in-house software in a preliminary development stage. First, it did not perform any correction to raw projections, such as scatter reduction, beam hardening correction, *etc*. Second, our codes and control parameters for the iterative procedure have not been optimized yet. Thus, we are still working to address these issues. Nevertheless, our proposed interior tomography and projection dataset reduction techniques have been successfully demonstrated in the experiment, which seem promising for real-world micro-/nano-CT applications.

In conclusion, we have developed our CS-based interior reconstruction technology and applied it for the advanced multi-scale CT facility. Our preliminary results have demonstrated the feasibility and merits of our methodology. After the nanoXCT™ system is installed in November, 2009, we will have real projection datasets of the 3D scaffold in 50nm resolution. Then, a comprehensive analysis will be performed to couple micro- and nano-CT results and derive critical information for bone regeneration research.

5. REFERENCES

[1] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory,* vol. 52, no. 4, pp. 1289-1306, Apr, 2006.

[2] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory,* vol. 52, no. 2, pp. 489-509, Feb, 2006.

[3] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear Total Variation Based Noise Removal Algorithms," *Physica D,* vol. 60, no. 1-4, pp. 259-268, Nov, 1992.

[4] H.Y. Yu, and G. Wang, "Compressed sensing based Interior tomography," *Phys Med Biol,* vol. 54, no. 9, pp. 2791-2805,Apr, 2009.

[5] H.Y. Yu, J.S. Yang, M. Jiang, and G. Wang, "Supplemental analysis on compressed sensing based interior tomography," *Phys Med Biol,* vol. 54, no. 18, pp. N425-N432, Sep, 2009.

[6] J.S. Yang, H.Y. Yu, M. Jiang, and G. Wang, "High order total variation minimization for interior tomography," *Inverse Problems*, pending revision, 2009.

[7] G. Wang, Y.B. Ye, and H.Y. Yu, *Interior Tomography and Instant Tomography by Reconstruction from Truncated Limited-Angle Projection Data,* USA Patent: 12/362979, 2007.

[8] Y.B. Ye, H.Y. Yu, Y.C. Wei and G. Wang, "A general local reconstruction approach based on a truncated Hilbert transform," *International Journal of Biomedical Imaging,* vol. 2007, Article ID: 63634, 8 pages, 2007.

[9] H.Y. Yu, Y.B. Ye, and G. Wang, "Local Reconstruction Using the Truncated Hilbert Transform via Singular Value Decomposition," *Journal of X-Ray Science and Technology,* vol. 16, no. 4, pp. 243-251, Dec, 2008.

[10] M. R. Howells, J. Kirz, and D. Sayre, "X-ray microscopes," *Scientific American,* vol. 264, no. 2, pp. 88-94, Feb, 1991.

[11] G. Schmahl, D. Rudolph, B. Niemann *et al.*, "X-ray microscopy," *Naturwissenschaften,* vol. 83, no. 2, pp. 61-70, Feb, 1996.

[12] F. Zernike, "Das Phasenkontrastverfahren bei der mikroskopischen Beobachtung," *Z. techn. Physik,* vol. 16, pp. 454-457, 1935.

[13] G. Wang, H.Y. Yu, and Y.B. Ye, "A scheme for multi-source interior tomography," *Med Phys*, vol. 36, no. 8, pp. 3575-3581, Aug, 2008.

[14] T. Andric, L. D. Wright, and J. W. Freeman, "Fabrication of Mineralized Osteon-like Scaffolds," BMES Annual Meeting, Pittsburgh, PA, USA, Oct, 2009.

Plenary Talk

Tuesday, June 8, 7:45am - 8:35am

Explosive Detection in Aviation Applications Using CT

Carl Crawford[1] & Harry Martz[2]

[1]U.S. Department of Homeland Security, Science and Technology Directorate
[2]Lawrence Livermore National Laboratory

CT scanners are deployed world-wide to detect explosives in checked and carry-on baggage. Though very similar to single- and dual-energy multi-slice CT scanners used today in medical imaging, some recently developed explosives detection scanners employ multiple sources and detector arrays to eliminate mechanical rotation of a gantry, photon counting detectors for spectral imaging, and limited number of views to reduce cost. For each bag scan, the resulting reconstructed images are first processed by threat detection algorithms to screen for explosives and other threats. Human operators review the images only when these automated algorithms report the presence of possible threats. The US Department of Homeland Security (DHS) has requirements for future scanners that include dealing with a larger number of threats, higher probability of detection, lower false alarm rates and lower operating costs. One tactic that DHS is pursuing to achieve these requirements is to augment the capabilities of the established security vendors with third-party algorithm developers. A third-party in this context refers to academics, national laboratories, and companies other than the established vendors. DHS is particularly interested in exploring the model that has been used very successfully by the medical imaging industry, in which university researchers develop algorithms that are eventually deployed in commercial medical imaging equipment. The purposes of this presentation are to review the presently deployed scanners and their concept of operations, and to discuss opportunities for third-parties to develop advanced reconstruction and threat detection algorithms.

Iterative reconstruction in micro-CT.

Michel Defrise[1], Christian Vanhove[1], Xuan Liu[2]

ABSTRACT

The paper analyses the performance of the algorithm of O'Sullivan and Benac for the maximum likelihood reconstruction from cone-beam micro-CT data, with the goal of reducing radiation dose in longitudinal studies of rats and mice. The algorithm is implemented both without regularization and with a regularizer based on a total variation penalty. Results with simulated 2D fan-beam data and with measured cone-beam data confirm the attractive properties of the TV penalty that have been illustrated by other authors, including a significant noise reduction and a good robustness to imperfectly corrected data defects, to reduced number of projections and to cone-beam artefacts caused by the incompleteness of the circular orbit. However, iterative reconstruction may introduce artefacts not observed in the FDK reconstructions.

I. INTRODUCTION

A major problem in micro-CT imaging of small animals is the potential impact of the radiation dose on the biological parameters that are estimated from the micro-CT data or from a correlated PET or SPECT study [1]. This problem is acute in longitudinal studies where an animal undergoes repeated scans. The problem is also related to the high spatial resolution if we remember that the dose should increase by a factor 16 when the isotropic resolution is improved by a factor 2 -at least with a linear reconstruction algorithm and if maintaining a constant voxel variance [2].

The potential benefits of iterative reconstruction belong to three categories: better physical model of the system, better modeling of the data noise, and possibility to incorporate prior knowledge on the object. In emission tomography these benefits have been extensively validated in the last twenty years and iterative algorithms are now exploited in almost all applications. In CT the validation of iterative algorithms is more recent and fragmentary but has already produced encouraging results [3]. The goal of this work is to determine whether penalized maximum likelihood reconstruction has a positive impact on the micro-CT studies of rats and mice performed in our laboratory. The central question is whether the relevant imaging tasks can still be fulfilled from iterative reconstructions of data acquired with a reduced number of projections, hence with a lower dose and reduced scan time. This preliminary paper describes the algorithm implemented and evaluates its performances on phantom data.

Fully exploiting the potential benefits of iterative reconstruction is more difficult in CB-CT than in PET and SPECT. Due to the large data size, it is impossible to pre-calculate the system matrix so that simplified geometric models must be used. In this work we combine a voxel-driven backprojector with an unmatched ray driven projector with linear interpolation [4]. These two operators are fairly accurate discretizations of the corresponding continuous operators, and this unmatched combination is known [5] to minimize the artefacts observed when the projector and backprojector are discretized using the same ray driven method [6]. The price to pay is that convergence of the iterative algorithm is not guaranteed and a better solution might be a matched distance-driven discretization [7].

The number of detected x-rays is large even in low-dose CT studies, but the dynamic range of the attenuation I_0/I can be large, resulting in a non-uniform data variance that should be taken into account during reconstruction. This can be done by assuming a gaussian distribution for the linearized data $\log(I_0/I)$, with the advantage that the resulting optimization problem is quadratic (if the regularizing penalty is quadratic). Accurately estimating the variance is difficult however if the data are noisy. This work takes an alternative approach: we adopt a Poisson model for the measured intensities I and, among other candidates [8]–[10], we implement the monochromatic maximum-likelihood (ML) algorithm of O'Sullivan and Benac [12] (see section 2), which is similar to the ML-TR algorithm [13] but with a proof of convergence. This Poisson model is an approximation because we apply it to data pre-corrected for a number of physical effects, as described in section 3. Though computation time is not our main concern at this point, we use two classical empirical tricks to accelerate the algorithm: the first is the ordered-subset processing of the data, the second one consists in using a FBP reconstruction as the initial image estimate for iterative reconstruction [14]. The second trick may introduce high frequency streaks, which disappear slowly -if ever- when iterating. We found however that this drawback is largely alleviated when a TV penalty is used and when the initial FBP reconstruction is obtained from interpolated data with a fine azimuthal sampling.

Recent works have generated interest for prior image models based on sparsity in some basis or frame [11]. Impressive reconstructions from a small number of azimuthal samples have been obtained with a TV penalty, which corresponds to assuming sparsity of the object gradient [21]. Most illustrations concern relatively simple objects such as Shepp's phantom for which as little as 16 azimuthal samples produce, with noise-free data, accurate results. As shown in [19] however, more complex objects require a larger number of projections. The efficacy of the TV penalty for non piece-wise constant, but somehow piece-wise homogeneous - objects remains an open question.

In this work we regularize the ML algorithm of O'Sullivan

[1]Dept. of Nuclear Medicine, Vrije Universiteit Brussel, Brussels, Belgium. e-mail: mdefrise@vub.ac.be, [2] Skyscan, Kontich, Belgium, e-mail: Xuan.Liu@skyscan.be

and Benac with a TV penalty, and illustrate the results obtained when reducing the number of projections by a factor 3 compared to the standard acquisition protocol and FDK reconstruction [15].

II. Maximum-likelihood algorithm of O'Sullivan and Benac.

We use the following notations.

- $y_j, j = 1, \ldots, D$: detected number of photons in line of response (LOR) j. There are D data bins (LORs),
- $I_j, j = 1, \ldots, D$: number of incident photons in LOR j. This is the "air" scan.
- $x_i, i = 1, \ldots, N$: linear attenuation coefficient in voxel i. There are N voxels.
- $s_{j,i}$: system matrix.

The average number of counts in LOR j is

$$< y_j >= I_j \, e^{-\sum_{i=1}^{N} s_{j,i} x_i} \qquad j = 1, \cdots, D \quad (1)$$

and the log-likelihood of a data set $\{y_j, j = 1, \ldots, D\}$ is

$$\Phi(x) = \sum_{j=1}^{D} y_j \log(< y_j >) - < y_j > - \log y_j! \quad (2)$$

Given the data set $\{y_j\}$ and the blank scan $\{I_j\}$, a ML estimator is an image x that maximizes $\Phi(x)$. The algorithm of O'Sullivan and Benac [12] converges to a ML estimator:

1) Pre-calculate the backprojection of the raw data, $b_i = \sum_{j=1}^{D} s_{j,i} y_j$. This needs to be done only once,
2) Determine factors Z_i satisfying the inequalities

$$\sum_{i=1}^{N} \frac{s_{j,i}}{Z_i} \leq 1 \quad j = 1, \ldots, D, \quad (3)$$

as closely as possible. We use a constant

$$Z_i = Z = \max_{1 \leq j \leq D} \sum_{l=1}^{N} s_{j,l} \quad i = 1, \ldots, N \quad (4)$$

equal to the maximum of the forward projection of a unit image.
3) Start with an initial image estimate x^0.
4) For each iteration $n = 0, 1, 2, \ldots$
   a) Forward project the previous image estimate x^n and calculate the number of counts $y_j^n = I_j \, e^{-\sum_{l=1}^{N} s_{j,l} x_l^n}$ that would be expected if x^n was the true image,
   b) Backproject these estimated counts $b_i^n = \sum_{j=1}^{D} s_{j,i} y_j^n$
   c) Calculate the next estimate as:

$$x_i^{n+1} = \left[x_i^n - \frac{1}{Z_i} \log(\frac{b_i}{b_i^n}) \right]_+ \qquad i = 1, \ldots, N \quad (5)$$

where $[u]_+ = u$ if $u \geq 0$ and $[u]_+ = 0$ if $u < 0$.

The definition of an ordered-subset version is straightforward.

III. Total-variation regularization

We use one of the usual forms of the TV penalty [16], [17]

$$TV(x) = \sqrt{||\nabla x||^2 + \epsilon^2} \quad (6)$$

with a two point discretization of each component of the gradient, and a small ϵ to obtain a differentiable functional. The regularized image estimate (ML-TV) is then defined as a maximizer of $\Psi(x) = \Phi(x) - \tilde{\beta} TV(x)$. Maximization is achieved using the one-step-late [18] modification of the algorithm of O'Sullivan and Benac, where equation (5) is replaced by

$$x_i^{n+1} = \left[x_i^n - \frac{1}{Z_i} \log \left(\frac{b_i + \tilde{\beta} \, (\frac{\partial TV(x)}{\partial x_i})_{x=x^n}}{b_i^n} \right) \right]_+ \quad (7)$$

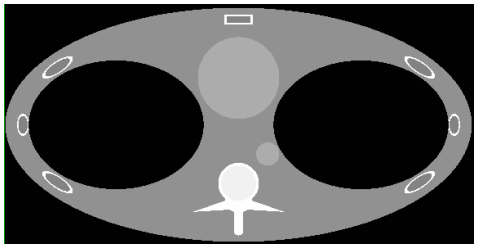
To normalize the regularization parameter we note that the gradient of $TV(x)$ is invariant (if $\epsilon = 0$) for a global scaling of x and we take $\tilde{\beta} = \beta \times \max_{i=1,\ldots,N}(b_i)$. This normalization ensures that the algorithm is invariant for a global scaling $I_j \rightarrow \alpha \times I_j$ of the incident intensities, hence a conversion of the measured intensities into number of quanta is not needed. Appropriate values of β are very small and for these small values we did not observe convergence problems with the OSL algorithm (7).

IV. 2D simulations

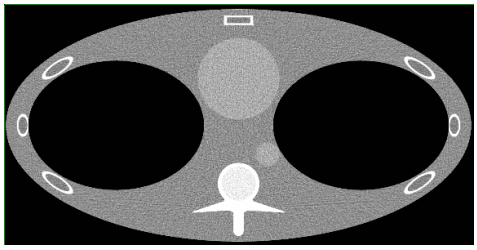
We simulated 2D parallel-beam CT data of one slice of a thorax phantom, with a maximum attenuation of 0.177 typical for a small animal study, with 600 projections and 600 radial samples, and with Poisson noise corresponding to $2\,10^5$ incident photons per ray. We use $\beta = 0.00025$ and 50 subsets.

The simulations confirm the noise reduction properties of the TV regularization described in previous works (see e.g. [3], [17], [19]). See figure 1. Additional observations are:

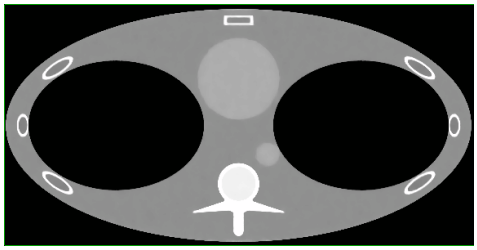
- Variance images obtained from 50 realizations of the noisy data confirm the result of Kohler et al [20] that the variance reduction achieved by edge-preserving penalties within homogeneous regions is accompanied by an increase of the variance around boundaries, when compared to FBP. One must therefore interpret with care the apparently precise localization of the sharp boundaries recovered by ML-TV reconstructions (results not shown).
- Despite the simplicity of this phantom, at least 100 to 200 projections are required to avoid distorsions, e.g. of the ribs. This might allow a reduction of the number of projections by a factor of 3 or more, but is far from the reduction allowed for a simple object such as Shepp's phantom for which accurate reconstructions are obtained from as little as 16 projections. Similar conclusions are found in [19].
- Figure 2 shows reconstructions from 100 projections. Initializing the iteration with the FBP reconstruction [14] from 600 projections obtained using cubic spline interpolation, and then performing ML-TV iterations using only the 100 original projections (and 20 subsets), significantly accelerate the convergence, though a detailed study of potential bias remains to be done.



(a) phantom

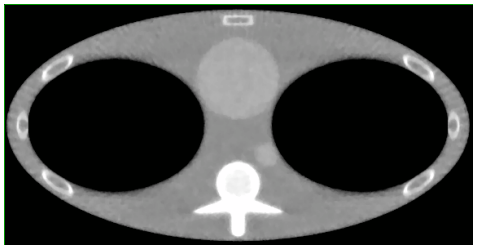


(b) FBP

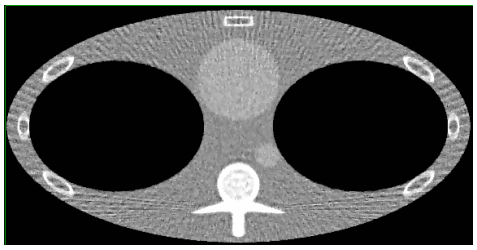


(c) ML-TV, $\beta = 0.00025$, 10 subsets, 250 iterations

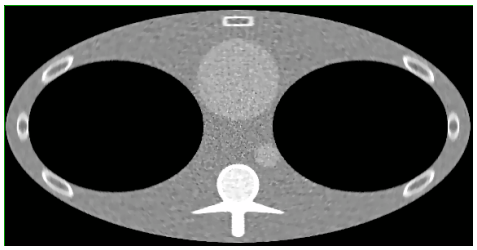
Fig. 1. 2D reconstructions from 600 noisy simulated projections.Grey scale $0.04, 0.06$ (true value is 0.096 in the spine and 0.070 in the ribs).



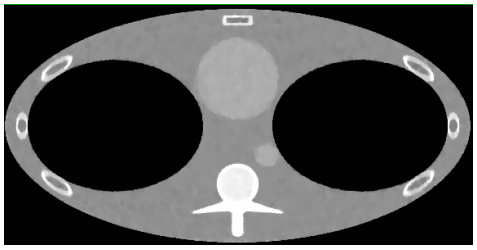
(a) Initial image: 0



(b) Initial image: FBP from 100 views



(c) Initial image: FBP from 600 interpolated views



(d) Initial image: 0, 250 iterations

Fig. 2. Grey scale $0.04, 0.06$. ML-TV reconstructions from 100 noisy projections, $\beta = 0.00025$, 10 subsets, 25 (first three rows) or 250 iterations.

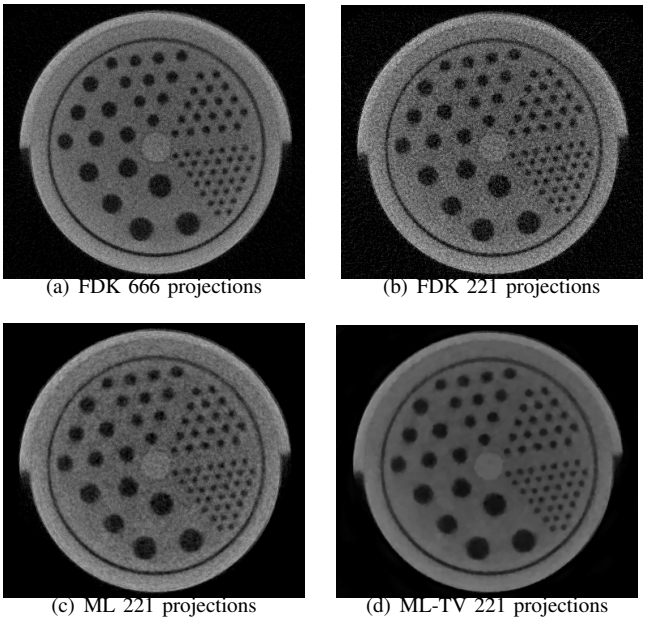
V. MICRO-CT DATA

The Micro Deluxe Phantom (Data Spectrum, Hillsborough, North-Carolina, USA) contains six sectors with rod diameters of 4.8 mm, 4.0 mm, 3.2 mm, 2.4 mm, 1.6 mm, and 1.2 mm. The rods are 34 mm in height and placed in a cylinder with 45 mm internal diameter. Micro-CT acquisitions were performed using a SkyScan 1178 (Kontich, Belgium) high-throughput in vivo system. X-ray projections were acquired into a 512x640 matrix at 50 kV and 615 μA, with a 0.5 mm Al filter. Projections (n=666 and n=221) were acquired by means of two x-ray detector-source pairs. The CT exposure time was set to 160 ms per projection. The data were corrected for geometric distorsions and flat field, a median filter is then applied to correct for individual defective pixels. The data were reconstructed into a $512 \times 512 \times 640$ matrix with 0.166 mm voxel size. The un-regularized ($\beta = 0$, ML) and regularized ($\beta = 0.0005$, ML-TV) reconstructions from 221 projections were done with 17 subsets and 18 iterations (un-optimized values). Figures 3 and 4 also show the FDK reconstructions from 221 and from 666 projections, where the latter corresponds to an approximately three-fold larger dose.

Several observations can be drawn from these results. As expected the TV penalty significantly suppresses noise for this piece-wise constant object. The iterative reconstructions remove a streak artefact that is observed in the FDK-221 transaxial slice (figure 3, top right) and is tentatively attributed to an imperfect projection pixel. It also suppresses an axial artefactual line in the FDK images. The TV penalty, however, does not blur out some very fine physical details of the phantom, such as the thin glue layer that links the two halves of the phantom. This layer is visible as a faint vertical line at the center of the coronal sections in figure 4: with 221 projections that layer is visible only in the iterative reconstructions. As observed previously, e.g. in cone-beam SPECT [6], maximum-likelihood reconstruction slightly improves the cone-beam artefacts caused by the non-completeness of the circular CB-CT data (see the external edges of the phantom). However, the iterative reconstructions show dark bands (vertical in figure 4 and best visible on the ML-TV reconstructions), the origin of which is still under investigation. Hypothesis include incomplete convergence or a specific reaction of the

ML estimation to data inconsistencies (beam hardening). The hypothesis that data inconsistencies are propagated differently is also supported by the artefactual spreading of the increased density of the o-ring (figure 4, centre right) within the plastic screw: this artefact, probably due to beam hardening, disappears in the two iterative reconstructions. Finally, although the ML-TV images have excellent spatial resolution (see e.g. the indentation of the various screws), the somewhat irregular edges of the bars (seen in the coronal slices) in the ML-TV image suggests as in the previous section that the variance is enhanced near edges [20].

We will present a detailed and quantitative analysis of these phantom studies, and show reconstructions from small animal CT data with the SkyScan 1178.



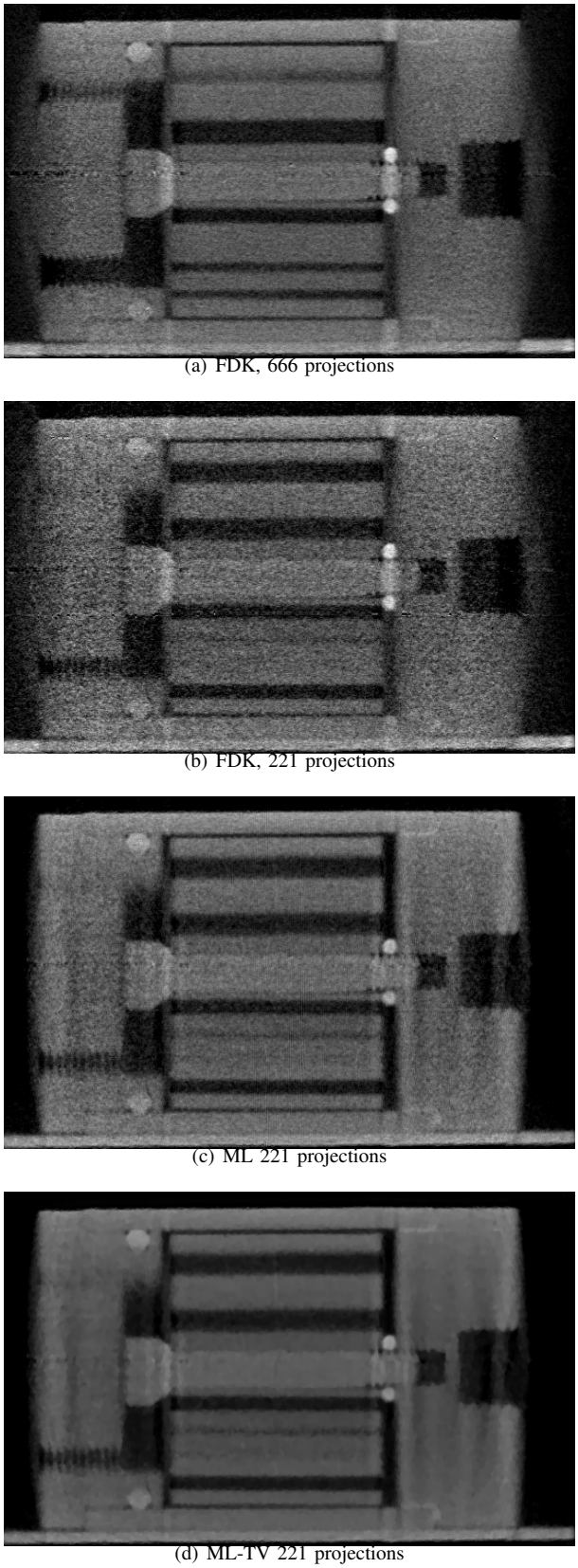
(a) FDK 666 projections (b) FDK 221 projections

(c) ML 221 projections (d) ML-TV 221 projections

Fig. 3. Central transaxial slice reconstructed from 666 projections with FDK, and from 221 projections with FDK, ML, ML-TV. The data with 666 projections are from an independent scan, hence the slightly different orientation of the slice.

REFERENCES

[1] S. R. Cherry, Phys Med Biol, 49, R13-R48, 2004.
[2] Huesman R H, Phys Med Biol, 22, 511-521, 1977.
[3] X. Pan et al,Inverse Problems 25, 123009 (36 pp), 2009.
[4] Joseph P, IEEE Trans. Med. Imaging 1 192-6, 1983.
[5] Zeng G and Gullberg G, IEEE Trans. Med. Imaging 19 548-55, 2000.
[6] Zeng G and Gullberg G, IEEE Trans. Nucl. Sc. 37, 759-767, 1990.
[7] De Man B and Basu S 2003, Phys. Med. Biol. 49, 2463-2475, 2004.
[8] Lange K and Carson R, J. Comput. Assist. Tomogr. 8 306-16, 1984.
[9] F. J. Beekman and C. Kamphuis, Phys. Med. Biol. 46, 1835-1844, 2001.
[10] H. Erdogan and J. A. Fessler, Phys. Med. Biol. 44, 2835-2851, 1999.
[11] L. Rudin et al, Physica D., vol.60, pp. 259-268, 1992.
[12] O'Sullivan and Benac, IEEE Trans Med Imag, vol 26, p. 283, 2007.
[13] Nuyts J et al, Phys. Med. Biol. 43 729-37, 1998.
[14] W. Zbijewski and F. J. Beekman, Med Phys 31, (1), 2004.
[15] Feldkamp L et al, J. Opt. Soc. Am. A 1 612-9, 1984.
[16] J.P. Oliveira et al, Signal Process. (2009), doi:10.1016/j.sigpro.2009.03.018 .
[17] V. Y. Panin et al, IEEE Trans. Nucl. Sc. 46, 2202-2210, 1999.
[18] P. Green, , IEEE Trans. Med. Imag., vol. 9, pp. 84-93, 1990.
[19] Jie Tang et al, Phys. Med. Biol. 54, 5781-5804, 2009.
[20] T. Kohler and R. Proksa, Proc. 10th Inter. Conf. Fully 3D Reconstruction, Beijing 2009, p. 263-266.
[21] E. Y. Sidky et al, J. X-Ray Sci. Tech., vol. 14, pp. 119-139, 2006.
[22] E. Y. Sidky and X. Pan, Phys. Med. Biol., vol. 53, pp. 4777-4807, 2008.



(a) FDK, 666 projections

(b) FDK, 221 projections

(c) ML 221 projections

(d) ML-TV 221 projections

Fig. 4. Coronal slice reconstructed from 666 projections with FDK, and from 221 projections with FDK , ML, ML-TV. The scanner axis is horizontal. The data with 666 projections are from an independent scan. The o-ring appears as the two brighter disks on both sides of the screw at the right edge of the central bar.

SOFT-THRESHOLD FILTERING APPROACH FOR RECONSTRUCTION FROM A LIMITED NUMBER OF PROJECTIONS

Hengyong Yu and Ge Wang

SBES Division and ICTAS Center for Biomedical Imaging, VT-WFU School of Biomedical Engineering and Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA, 24060, USA

hengyong-yu@ieee.org, ge-wang@ieee.org

ABSTRACT

In the medical imaging filed, the discrete gradient transform (DGT) is widely used as a sparsifying operator to define the total variation (TV). Recently, the TV minimization becomes a hot topic in image reconstruction, and usually implemented using the steepest descent method (SDM). Since the TV minimization with the SDM takes long computational time, here we construct a pseudo-inverse of the DGT and adapt a soft-threshold filtering algorithm, whose convergence and efficiency have been theoretically proved. Also, we construct a pseudo-inverse of the discrete difference transform (DDT) and design an algorithm for L1 minimization of the total difference (TD). These two methods are evaluated in numerical simulation. The results demonstrate the merits of the proposed techniques.

Index Terms— Compressive sampling (CS), soft-threshold filtering, discrete gradient transform (DGT), total variation (TV), total difference (TD)

1. INTRODUCTION

Because the x-ray attenuation coefficient often varies mildly within an anatomical component, and large changes are usually confined around borders of tissue structures, the discrete gradient transform (DGT) has been widely utilized as a sparsifying operator in compressive sampling (CS) inspired CT reconstruction [1-4], which is also referred to as the total variation (TV) minimization [5]. This kind of algorithms can be divided into two major steps. In the first step, an iteration formula (e.g., SART) is used to update a reconstructed image for data discrepancy reduction. In the second step, a search method (e.g., the standard steepest descent technique) is used in an iterative framework for TV minimization. These two steps need to be iteratively performed in an alternating manner. However, there are no standard stopping and parameter selection criteria for the second step. Usually, these practical issues are addressed in an ad hoc fashion. On the other hand, soft-threshold nonlinear filtering [6-8] was proved to be a convergent and efficient algorithm for the L1- norm minimization regularized by a sparsity constraint. Unfortunately, because the discrete gradient transform is not invertible, it does not satisfy the restricted isometry property (RIP) required by the CS theory [9-10] and soft-threshold algorithm [7-8]. In other words, the soft-threshold algorithm cannot be directly applied for TV minimization. Motivated by this challenge, here we construct two pseudo-inverse transforms and apply the soft-threshold filtering for image reconstruction from a limited number of projections.

2. SOFT-THRESHOLD FILTERING PRINCIPLE

Let $\mathbf{f} = [f_1, f_2, ..., f_N]^T \in \mathbb{R}^N$ be an object function and $\mathbf{g} = [g_1, g_2, ..., g_M]^T \in \mathbb{R}^M$ be a dataset. They are linked by

$$\mathbf{g} = \mathbf{A}\mathbf{f} + \mathbf{e}, \tag{2.1}$$

where $\mathbf{A} = (a_{m,n}) \in \mathbb{R}^M \times \mathbb{R}^N$ is the linear measurement matrix satisfying $\|\mathbf{A}^T\mathbf{A}\| < 1$, and $\mathbf{e} \in \mathbb{R}^M$ the measurement noise. When the system (2.1) is ill-posed, additional constraints are required to regularize the solution. Particularly, given a basis or frame $(\boldsymbol{\varphi}_\gamma)_{\gamma\in\Gamma}$ of the space $\mathbb{R}^N$ satisfying $\mathbf{f} = \sum_{\gamma\in\Gamma} \langle \mathbf{f}, \boldsymbol{\varphi}_\gamma \rangle \boldsymbol{\varphi}_\gamma$, and a sequence of strictly positive weights $\mathbf{w} = (w_\gamma)_{\gamma\in\Gamma}$, we define the functional $\Phi_{\mathbf{w},p}(\mathbf{f})$ by

$$\Phi_{\mathbf{w},p}(\mathbf{f}) = \|\mathbf{g} - \mathbf{A}\mathbf{f}\|^2 + \sum_{\gamma\in\Gamma} 2w_\gamma \left| \langle \mathbf{f}, \boldsymbol{\varphi}_\gamma \rangle \right|^p, \tag{2.2}$$

where $\langle \bullet, \bullet \rangle$ represents the inner product and $1 \le p \le 2$. The solution of Eq. (2.2) is the minimizer of $\Phi_{\mathbf{w},p}(\mathbf{f})$ that can be recursively determined by the soft-threshold algorithm [7-8]:

$$\mathbf{f}^k = \mathbb{S}_{\mathbf{w},p}(\mathbf{f}^{k-1} + \mathbf{A}^T(\mathbf{g} - \mathbf{A}\mathbf{f}^{k-1})), \tag{2.3}$$

where $k = 1, 2, ...$ is the iteration number, $\mathbf{f}^0$ the initial guess in $\mathbb{R}^N$, and

$$\mathbb{S}_{\mathbf{w},p}(\mathbf{f}) = \sum_{\gamma\in\Gamma} S_{w_\gamma,p}\left(\left\langle \mathbf{f},\boldsymbol{\varphi}_\gamma \right\rangle\right)\boldsymbol{\varphi}_\gamma \tag{2.4}$$

with

$$S_{w,1}(x) = \begin{cases} x-w & if\ x \ge w \\ 0 & if\ |x| < w \\ x+w & if\ x \le -w \end{cases}. \tag{2.5}$$

Eq.(2.5) is called soft-threshold filtering [7]. Daubechies *et al.* proved the convergence of Eq. (2.3) [7-8].

3. ALGORITHM DEVELOPMENT

3.1. Imaging Model

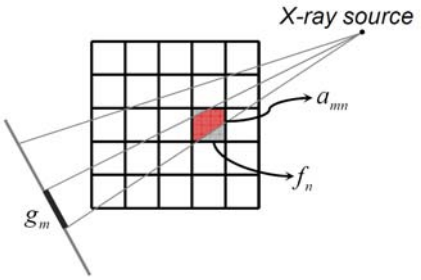


Figure 1. Projection model of a discrete image in fan-beam geometry.

In the context of CT reconstruction, a two-dimensional digital image can be expressed as $\mathbf{f} = (f_{i,j}) \in \mathbb{R}^I \times \mathbb{R}^J$, where the index $1 \le i \le I$ and $1 \le j \le J$ are integers. Define

$$f_n = f_{i,j},\ n = (i-1)\times J + j, \tag{3.1}$$

with $1 \le n \le N$ and $N = I \times J$, we can re-arrange the image into a vector for the measurement model Eq. (2.1) . In this paper, we will use both the signs $f_{i,j}$ and f_n for convenience. Each component of the function $\mathbf{g}$ in Eq. (2.1) is a measured datum with M being the product of the number of projections and the number of detector elements. In a typical fan-beam geometry, the n^{th} pixel can be viewed as a rectangular region with a constant value f_n, the m^{th} measured datum g_m as an integral of areas of pixels partially covered by a narrow beam from an x-ray source to a detector element and weighted by the corresponding x-ray linear attenuation coefficients respectively. Thus, the component $a_{m,n}$ in Eq. (2.1) denotes the interaction area between the n^{th} pixel and the m^{th} fan-beam path (Fig. 1). While the whole matrix $\mathbf{A}$ represents the forward projection, $\mathbf{A}^T$ implements the back projection. The SART-type solution to Eq. (2.1) can be written as [11]:

$$f_n^k = f_n^{k-1} + \lambda^k \frac{1}{a_{+n}} \sum_{m=1}^{M} \frac{a_{m,n}}{a_{m+}} \left(g_m - \mathbf{A}_m \mathbf{f}^{k-1}\right), \tag{3.2}$$

where $a_{+n} = \sum_{m=1}^{M} a_{m,n} > 0$, $a_{m+} = \sum_{n=1}^{N} a_{m,n} > 0$, $\mathbf{A}_m$ is the m^{th} row of $\mathbf{A}$, k the iteration index, and $0 < \lambda^k < 2$ a free relaxation parameter. The solution Eq. (3.2) exactly corresponds to "$\mathbf{f}^{k-1} + \mathbf{A}^T(\mathbf{g} - \mathbf{A}\mathbf{f}^{k-1})$" in Eq. (2.3) with a_{+n} and a_{m+} being weights for normalization, and can be used in the same place for a better converging behavior. To apply the soft-threshold algorithm subject to a sparsity constraint, we must find a suitable basis or frame $\boldsymbol{\varphi}_\gamma$ and $\mathbb{S}_{\mathbf{w},p}(\mathbf{f})$, which is the major contribution of this paper.

3.2. Pseudo-Inverse of the Discrete Gradient Transform

In the medical imaging field, the DGT has been widely used to define a sparsity constraint. Let us assume that a digital image satisfies the so-called Neumann conditions on the boundary:

$$f_{0,j} = f_{1,j} \text{ and } f_{I,j} = f_{I+1,j} \text{ for } 1 \le j \le J,$$
$$f_{i,0} = f_{i,1} \text{ and } f_{i,J} = f_{i,J+1} \text{ for } 1 \le i \le I. \tag{3.3}$$

Then, the standard isotropic discretization of TV can be expressed as

$$TV(\mathbf{f}) = \sum_{i=1}^{I}\sum_{j=1}^{J} d_{i,j},\ d_{i,j} = \sqrt{(f_{i,j} - f_{i+1,j})^2 + (f_{i,j} - f_{i,j+1})^2}. \tag{3.4}$$

Thus, Eq. (2.2) for CT becomes

$$\Phi_{w,1}(\mathbf{f}) = \|\mathbf{g} - \mathbf{A}\mathbf{f}\|^2 + 2wTV(\mathbf{f}). \tag{3.5}$$

Note that there does not exist a frame such that $d_{i,j} = \left\langle \mathbf{f}, \boldsymbol{\varphi}_{i,j} \right\rangle$, the solution Eq. (2.3) cannot be directly applied to minimize $\Phi_{w,1}(\mathbf{f})$ defined by Eq. (3.5). However, we can construct a pseudo-inverse of the discrete gradient transform as follows. Rewriting the result from Eq. (3.2) as $\tilde{f}_n^k$, we can compute

$$d_{i,j}^k = \sqrt{(\tilde{f}_{i,j}^k - \tilde{f}_{i+1,j}^k)^2 + (\tilde{f}_{i,j}^k - \tilde{f}_{i,j+1}^k)^2}. \tag{3.6}$$

According to the soft-threshold operation in Eq. (2.5), when $d_{i,j}^k < w$ we can adjust the values of $\tilde{f}_{i,j}^k$, $\tilde{f}_{i+1,j}^k$ and $\tilde{f}_{i,j+1}^k$ to make $d_{i,j}^k = 0$, and when $d_{i,j}^k \ge w$ we can reduce the values of $(\tilde{f}_{i,j}^k - \tilde{f}_{i+1,j}^k)^2$ and $(\tilde{f}_{i,j}^k - \tilde{f}_{i,j+1}^k)^2$ to perform the filtering. That is, we can construct the following pseudo-inverse:

$$f_{i,j}^k = \frac{1}{4}\left(2f_{i,j}^{k,a} + f_{i,j}^{k,b} + f_{i,j}^{k,c}\right), \tag{3.7}$$

$$f_{i,j}^{k,a} = \begin{cases} \dfrac{2\tilde{f}_{i,j}^k + \tilde{f}_{i+1,j}^k + \tilde{f}_{i,j+1}^k}{4}, & \text{if } d_{i,j}^k < w \\ \tilde{f}_{i,j}^k - \dfrac{w\left(2\tilde{f}_{i,j}^k - \tilde{f}_{i+1,j}^k - \tilde{f}_{i,j+1}^k\right)}{4d_{i,j}^k}, & \text{if } d_{i,j}^k \ge w \end{cases}, \tag{3.8}$$

$$f_{i,j}^{k,b}=\begin{cases}\dfrac{\tilde{f}_{i,j}^{k}+\tilde{f}_{i-1,j}^{k}}{2}, & \text{if } d_{i-1,j}^{k}<w\\ \tilde{f}_{i,j}^{k}-\dfrac{w\left(\tilde{f}_{i,j}^{k}-\tilde{f}_{i-1,j}^{k}\right)}{2d_{i-1,j}^{k}}, & \text{if } d_{i-1,j}^{k}\geq w\end{cases}, \quad (3.9)$$

$$f_{i,j}^{k,c}=\begin{cases}\dfrac{\tilde{f}_{i,j}^{k}+\tilde{f}_{i,j-1}^{k}}{2}, & \text{if } d_{i,j-1}^{k}<w\\ \tilde{f}_{i,j}^{k}-\dfrac{w\left(\tilde{f}_{i,j}^{k}-\tilde{f}_{i,j-1}^{k}\right)}{2d_{i,j-1}^{k}}, & \text{if } d_{i,j-1}^{k}\geq w\end{cases}. \quad (3.10)$$

In summary, we have a soft-threshold algorithm for TV minimization in the following pseudo-code:

S1: Initialize the parameters k, w;
S2: Update the current reconstruction using Eq. (3.2);
S3: Perform the non-linear filter using Eq. (3.7);
S4: Go to S2 until the stopping criterion is met.

3.3. Pseudo-Inverse of the Discrete Difference Transform

In addition to the discrete gradient transform, there are other possible sparse transforms. For example, we can define a total difference (TD) of $\mathbf{f}$ as

$$TD(\mathbf{f})=\sum_{i=1}^{I}\sum_{j=1}^{J}d_{i,j},\ d_{i,j}=\left|f_{i,j}-f_{i+1,j}\right|+\left|f_{i,j}-f_{i,j+1}\right|, \quad (3.11)$$

and rewrite Eq. (2.2) as

$$\Phi_{w,1}(\mathbf{f})=\left\|\mathbf{g}-\mathbf{A}\mathbf{f}\right\|^{2}+2wTD(\mathbf{f}). \quad (3.12)$$

We call $d_{i,j}$ in Eq. (3.11) a discrete difference transform. Similar to what we have done in Subsection 3.2, after the soft-threshold filtering, we can construct a pseudo-inverse of $f_{i,j}^{k}$ as

$$f_{i,j}^{k}=\frac{1}{4}\begin{pmatrix}q(w,\tilde{f}_{i,j}^{k},\tilde{f}_{i+1,j}^{k})+q(w,\tilde{f}_{i,j}^{k},\tilde{f}_{i,j+1}^{k})\\ +q(w,\tilde{f}_{i,j}^{k},\tilde{f}_{i,j-1}^{k})+q(w,\tilde{f}_{i,j}^{k},\tilde{f}_{i-1,j}^{k})\end{pmatrix}, \quad (3.13)$$

$$q(w,y,z)=\begin{cases}\dfrac{y+z}{2}, & \text{if } |y-z|<w\\ y-\dfrac{w}{2}, & \text{if } (y-z)\geq w\\ y+\dfrac{w}{2}, & \text{if } (y-z)\leq -w\end{cases}. \quad (3.14)$$

That is, we have a soft-threshold algorithm for TD minimization in the following pseudo-code:

S1: Initialize the parameters k, w;
S2: Update the current reconstruction using Eq. (3.2);
S3: Perform the non-linear filter using Eq. (3.13);
S4: Go to S2 until the stopping criterion is met.

4. NUMERICAL SIMULATION

To demonstrate the validity of the proposed algorithms, we implemented them in MatLab and performed numerical tests. We assumed a circular scanning locus of radius 57.0 cm and fan-beam geometry. The object was a modified Shepp-Logan phantom in a compact support with a radius of 10.0 cm. We used an equi-spatial virtual detector array of length 20.0 cm. The detector was centered at the system origin and made perpendicular to the direction from the origin to the x-ray source. The detector array consisted of 300 elements. For each of our selected numbers of projections over a full-scan range, we first equi-angularly acquired the corresponding projection dataset based on the discrete projection model shown in Fig. 1. Then, we reconstructed the images using the following four methods: (1) the classical SART iteration method without the regularization of sparsity, (2) the TV minimization algorithm implemented in [2] using the steepest descent search method, (3) the TV minimization algorithm with soft-threshold filtering proposed in Subsection 3.2, and (4) the TD minimization method with soft-threshold filtering proposed in Subsection 3.3.

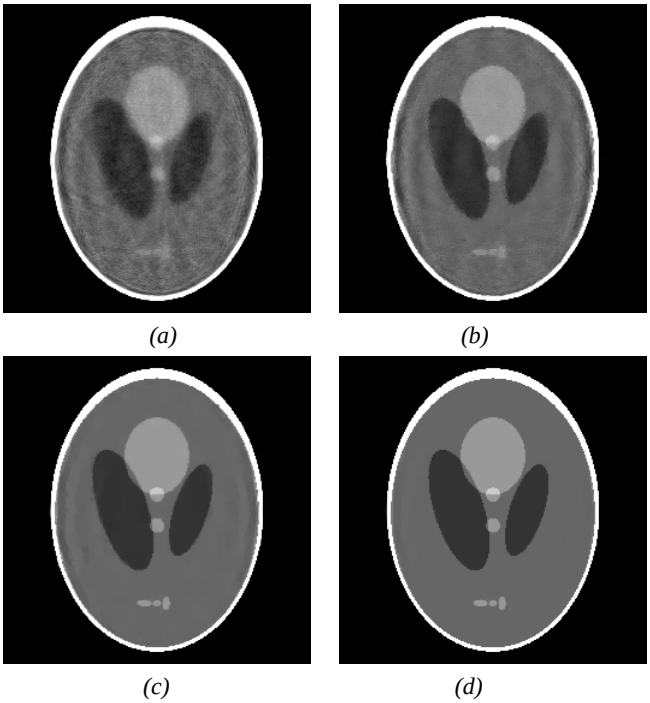
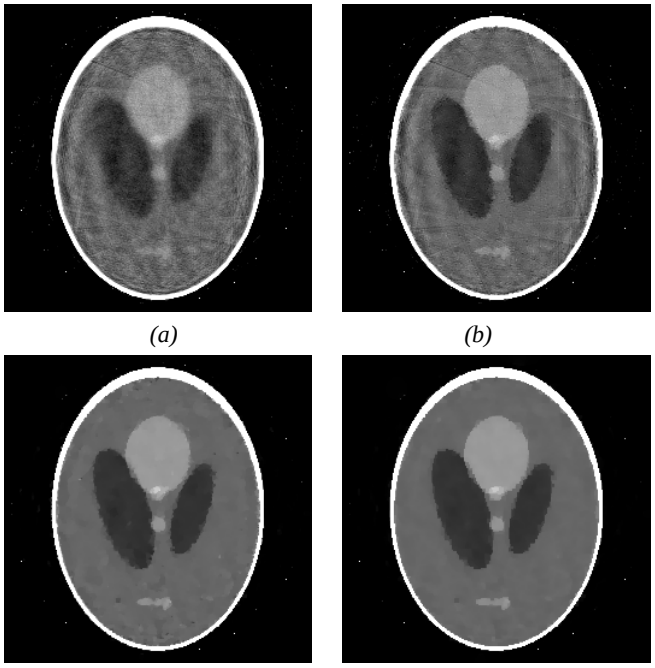


Figure 2. Reconstruction from 21 noise-free projections. (a) The reconstruction using the SART method without the TV minimization, (b) using the steepest descent method for TV minimization, (c) and (d) using the soft-threshold filtering methods for TV and TD, respectively. The display window is [0,0.5].

To accelerate the convergence, we employed the projected gradient method [8] to determine an optimal threshold w for each filtering, which is similar to what we have done in [12]. For all the above methods, the parameter λ^{k} in the SART iteration formula in Eq.(3.2) was set to be the constant 1.0, and the stopping criterion was defined as reaching the maximum iteration number 5000. Figs. 2 and 3 show the reconstructed 256x256 images from 21 noise-free

and noisy projections, respectively. For the given finite iteration number 5000, the results in Figures 2-3 shows that the proposed TV and TD minimization methods outperform the traditional steepest descent search method. Meanwhile, the TD minimization is not only simpler for practical implementation but also performs slight better than TV. Therefore, we believe that the proposed TD minimization method in this paper can attract many applications in the near future



(a) *(b)* *(c)* *(d)*

Figure 3. The counterparts of Figure 2 from projections corrupted by Poisson noise, assuming $5X10^4$ photons per detector element.

5. DISCUSSIONS AND CONCLUSION

Clearly, the pseudo-inverse transforms constructed in this paper can also be regarded as non-linear filters. However, we prefer to call them pseudo-inverse transforms, because they serve as inverse transforms of DGT and DDT respectively after the soft-threshold filtration defined by Eq. (2.3) in the framework of soft-threshold filtering framework.

In conclusion, we have constructed a pseudo-inverse of the discrete gradient transform and a pseudo-inverse of the discrete difference transform to apply the soft-threshold filtering principle for image reconstruction subject to a sparsity constraint. In the spirit of the above-described strategy for algorithm construction, we can derive a variety of algorithms of this type. In the near future, we will investigate and evaluate other reconstruction schemes with numerical, physical, preclinical and clinical datasets.

6. ACKNOWLEDGEMENT

This work is partially supported by NIH/NIBIB grants (EB002667, and EB011785).

7. REFERENCES

1. G.H. Chen, J. Tang, and S. Leng, *Prior image constrained compressed sensing (PICCS): A method to accurately reconstruct dynamic CT images from highly undersampled projection data sets.* Medical Physics, 2008. 35(2): p. 660-663.
2. H.Y. Yu, and G. Wang, *Compressed sensing based interior tomography.* Phys Med Biol, 2009. 54(9): p. 2791-2805.
3. T. Tang, B.E. Nett, and G.H. Chen, *Performance comparison between total variation (TV)-based compressed sensing and statistical iterative reconstruction algorithms.* Phys Med Biol, 2009. 54(19): p. 5781-5804.
4. E.Y. Sidky, C.M. Kao, and X.H. Pan, *Accurate image reconstruction from few-views and limited-angle data in divergent-beam ct.* J. of X-Ray Science and Technology, 2006. 14(2): p. 119-139.
5. L.I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms.* Physica D, 1992. 60(1-4): p. 259-268.
6. M.A.T. Figueiredo, and R.D. Nowak, *An em algorithm for wavelet-based image restoration.* IEEE Transactions on Image Processing, 2003. 12(8): p. 906-916.
7. I. Daubechies, M. Defrise, and C. De Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint.* Communications on Pure and Applied Mathematics, 2004. 57(11): p. 1413-1457.
8. I. Daubechies, M. Fornasier, and I. Loris, *Accelerated projected gradient method for linear inverse problems with sparsity constraints.* Journal of Fourier Analysis and Applications, 2008. 14(5-6): p. 764-792.
9. E.J. Candes, J. Romberg, and T. Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information.* IEEETransactions on Information Theory, 2006. 52(2): p. 489-509.
10. D.L. Donoho, *Compressed sensing.* IEEE Transactions on Information Theory, 2006. 52(4): p. 1289-1306.
11. G. Wang, and M. Jiang, *Ordered-subset simultaneous algebraic reconstruction techniques (OS-SART).* J. of X-ray Science and Technology, 2004. 12(3): p. 169-177.
12. H.Y. Yu, and G. Wang, *SART-type image reconstruction from a limited number of projections with the sparsity constraint.* International Journal of Biomedical Imaging, Article ID:934847, 2010.

Investigation of Image Reconstruction in CT with a Limited Number of Stationary Sources

D. Xia[1], J. Bian[1], X. Han[1], E. Y. Sidky[1], J. Lu[2], O. Zhou[2], and X. Pan[1]
[1]Department of Radiology, The University of Chicago
[2]Department of Physics and Astronomy, University of North Carolina at Chapel Hill

***Abstract*— As carbon-nanotube (CNT) X-ray sources with performance properties comparable to that of X-ray tubes in conventional CT are becoming available, interest exists in developing innovative CT imaging systems employing stationary arrays of CNT sources because they possess a number of potential advantages over conventional CT systems. However, various physical and cost constraints can limit the number of CNT sources within the arrays, or, equivalently, the number of projection "views", in a stationary CNT-based CT system. In this work, we investigate image reconstruction for several CNT-based CT systems of practical interest that use different configurations of CNT-source arrays. In addition, we investigate the effect of linear and curved arrays of CNT sources on image reconstruction. Results of the study demonstrate that images with quality comparable to that of conventional CT can be reconstructed for CNT-based CT systems of different configurations with either linear or curved CNT source arrays.**

I Introduction

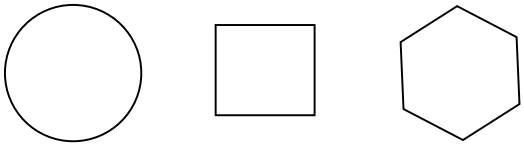
Conventional computed tomography (CT) systems generally use single or dual sources for illuminating the imaged subject and collect projection data at multiple views by moving either the source(s)/detector or imaged subject. The movement of sources/detector at high speed with high precision imposes a difficult challenge on the design, calibration, and maintenance of higher throughout and higher resolution CT systems. In recent years, there have been significant advances in the development of carbon-nanotube (CNT) X-ray source [1, 2]. In contrast to conventional X-ray tubes with thermionic cathodes, CNT sources may have a number of potential advantageous features, including ultra-fast switch, compact size, and facile electronic control. Effort has been devoted to developing radiography [3], micro-CT [4], and digital breast tomosynthesis (DBT) [5] systems that employ CNT X-ray sources. An unique feature of CNT sources is their compactness, which allows their integration to form a compact array of CNT X-ray sources. CT imaging system uses stationary arrays of CNT sources can achieve faster imaging than conventional CT, because it replaces the mechanical rotation of the imaged subject or source(s)/detector in conventional CT imaging with fast electronic switch of CNT sources.

In a typical scan of conventional CT, projection data are typically acquired at a large number ($400 \sim 1000$) of views, so that images with no significant aliasing artifacts can be reconstructed by use of analytic algorithms such as filtered-backprojection (FBP) algorithm. Arrays of integrated CNT sources offer the opportunity for developing innovative CT systems with stationary sources tailored to specific applications. In the CT systems, different CNT sources can be interpreted as different projection views of a single source in conventional CT imaging. Despite the fact that an array of sources may consist of a considerable number of CNT sources, however, as compared to the number of projection views in conventional CT imaging, the number of CNT sources is substantially low due to various physical and cost constraints. When images are reconstructed from data collected at such a number of "views" (or, equivalently, CNT sources) by use of analytic algorithms, images generally suffer from significant aliasing artifacts.

A purpose of the work is to investigate whether, by use of optimization-based algorithms, images with reduced artifacts can be reconstructed from data collected at a small number of "views" (or, equivalently, CNT sources). Also, CNT sources can be integrated, in general, into an array of different shapes, including linear and curved shapes. However, levels of costs, calibration, and maintenance for arrays of different shapes can be substantially different. In fact, it is believed that a linear array of CNT sources is much easier to make, calibrate, and maintain than is a curved array. Therefore, the second purpose of the work is to investigate image reconstruction from data collected CT systems that use stationary, linear or curved arrays. Results of the work can provide potentially valuable information and guidance for the design and manufacture of cost-effective CNT-source arrays and innovative CT imaging configurations.

II Imaging Configurations

We investigate several potentially useful imaging configurations, as shown in Figs. 1 and 2, each of which use stationary, linear or curved arrays of CNT sources. In particular, linear and curved arrays are used for forming square-/hexagon-shaped and circle-shaped imaging configurations. A leading difference between the sets of configurations shown in Figs. 1 and 2 is that the configura-



a b c

Fig. 1. Imaging configurations without array gaps: (a) circular, (b) square-, and (c) hexagon-shaped imaging configurations.

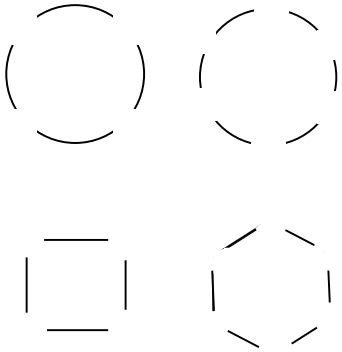


Fig. 2. Imaging configurations with array gaps. Top row: circle-shaped configurations with 4 (left) and 6 (right) curved arrays. Bottom row: square-shaped configuration with 4 (left) arrays and hexagon-shaped configuration with 6 (right) curved arrays.

tions in Figs. 1 contains gaps between arrays. Gaps between source arrays of the configurations in Fig. 2 are introduced for accommodating practical hardware and cost constraints that are likely to be encountered in the assembly of the configurations. For the configurations studied, we use a total of 60 CNT sources. For example, in the configuration with 4 or 6 source arrays, there are 15 or 10 CNT sources uniformly distributed on each of, or a portion of each, of arrays. For the circular imaging configuration, the 60 CNT sources are uniformly distributed over 2π.

III Reconstruction algorithms

The FBP algorithm is a widely used analytical algorithm for image reconstruction in practical CT imaging. It can yield images of acceptable utility when a circular trajectory is used and when data are acquired at a large number of projection views. For non-circular configurations, such as square- and hexagon-shaped configurations, one may, in principle, rebin the collected data into the form so that the FBP algorithm can be applied. However, as discussed above, CT systems with stationary, source arrays are likely to have only a small number of sources, as compared to the typical number of views in conventional

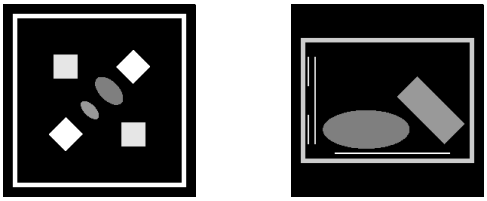


Fig. 3. Phantoms 1 (left column) and 2 (right column) used in the numerical studies. Display window: [0.0,2.0].

CT imaging. In this case, the data-rebinning process is likely to introduce significant data error and thus lead to severe artifacts in the reconstructed images. Therefore, the study thus will focus on whether optimization-based algorithms can reconstruct images with reduced artifacts directly from data collected with imaging configurations employing stationary arrays of CNT sources.

The problem of image reconstruction discussed above entails image reconstruction from highly sparse data. A constrained total-variation (TV)-minimization algorithm, which we simply refer to as the TV algorithm, has been developed recently for image reconstruction from highly sparse projection data [6, 7]. It reconstructs an image through minimizing the image TV subject to the data and other constraints. In this work, we modify and apply the TV algorithm to reconstructing images from data acquired with imaging configurations displayed in Figs. 1 and 2, which use only 60 CNT sources (or equivalently, 60 "views"). While the TV algorithm and its implementation details have been described in Refs. [6, 7], additional details concerning its modifications and application involved in the study will be reported at the meeting. Moreover, for evaluation of algorithm performance, we have also implemented some existing algorithms such as projection-onto-convex-set (POCS) [8] and expectation-maximization (EM) algorithms [9, 10] and applied them to reconstructing from the same data sets.

IV Results

We have performed computer-simulation studies to investigate and evaluate image reconstructions for the imaging configurations described above.

A Study materials

As displayed in Fig. 3, we use two objects, which simulate suitcases containing objects of different shapes at different orientations, to generate projection data. For the imaging configurations with curved arrays on a circle of a radius $R = 40$ cm, we use a virtual 2D flat-panel detector placed at the center of the rotation and assume that the line connecting the detector center and a CNT source is perpendicular to the detector panel. For the square- and hexagon-shaped imaging configurations without array gaps, the radii of their circumcircles are identical to

that of the circular configuration. As such, the lengths of each side for the square- and hexagon-shaped configurations are 56.6 cm and 40 cm. For each of the CNT sources on an array, the corresponding detector panel is parallel to the array. The detector array consists of 512×512 elements each of which has a size of 1.17 mm.

For the imaging configurations without array gaps in Fig. 1, an equal number of CNT sources 60 CNT sources are distributed over the source array, with an equal angular interval of 6° for the circular configuration, and with equal distance intervals of 4.0 cm and 3.8 cm, respectively, for the square- and hexagon-shaped configurations. For the imaging configurations with array gaps depicted in Fig. 2, an equal number of CNT sources are uniformly distributed in the middle region of an array, with an equal angular interval of 4° for the circular configuration, and with equal distance intervals of 1.3 cm and 2.0 cm for the square- and hexagon-shaped configurations. This leads to an effective gap of 20° between any two source arrays in each of the configurations in Fig. 2.

We have computed projection data from each of the two phantoms using each of the imaging configurations described above. In realistic CT imaging, it is likely that the underlying objects are continuous. Therefore, in an attempt to include realistically the continuous distribution of the objects, we have used an analytic method to compute the line integrals through the objects. Because the imaging model based upon which the TV and other algorithms developed is typically a discrete model, there will be some inconsistencies between the analytically generated data and the discrete imaging model used. Such intentionally introduced inconsistencies serve as a test of algorithms' robustness when applied to data containing inconsistencies.

In addition to visual inspection of the reconstructed images, we also use a number of metrics to evaluate quantitatively image properties reconstructed from data collected with different configurations under study by use of different algorithms. We describe two of the metrics used here: (a) root mean square error (RMSE) and (b) the universal quality index (UQI). RMSE is widely used for measuring differences between two images and is given by

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(f_i - f_i^r)^2}, \qquad (1)$$

where f and f^r denote the reconstructed image and the reference image, respectively, and N is the number of pixels in each image. On the other hand, the index UQI describes the structural similarity between the reconstructed and reference images and is defined as [11]

$$UQI(f, f^r) = \left(\frac{\mu\mu_r}{\mu^2 + \mu_r^2}\right)\left(\frac{2\sigma_{f,f^r}}{\sigma^2 + \sigma_r^2}\right), \qquad (2)$$

where $\mu = \frac{1}{N}\sum_{i=1}^{N} f_i$ and $\mu_r = \frac{1}{N}\sum_{i=1}^{N} f_i^r$ are the mean values of the reconstructed and reference images, $\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(f_i - \mu)^2$ and $\sigma_r^2 = \frac{1}{N-1}\sum_{i=1}^{N}(f_i^r - \mu_r)^2$ denote the variances of the reconstructed and reference images, and

$$\sigma_{f,f^r} = \frac{1}{N-1}\sum_{i=1}^{N}(f_i - \mu)(f_i^r - \mu_r)$$

indicates the covariance of the two images. The higher the UQI , the closer the reconstructed image is to the reference image. It can be shown that $UQI = 1.0$ for two identical images.

B Image reconstruction for imaging configurations without array gaps

We first reconstruct images from data acquired with each of the three imaging configurations shown in Fig. 1 by use of EM, POCS, and TV algorithms, and display the reconstructions in Figs. 4 and 5. The results suggest that images without significant artifacts may be reconstructed for the non-conventional configurations. Furthermore, it can be observed that the TV algorithm yields images with less artifacts than those in images reconstructed by use of other algorithms. In particular, the TV algorithm appears to be able to recover a thin sheet more accurately than other algorithms. Using the true phantom images in Fig. 3 as the reference images and the corresponding reconstructed images in Eqs. (1) and (2), we have also calculated the RMSEs and UQIs, which are displayed in Fig. 6. These quantitative results also indicate that the TV algorithm generally yield images of high quality, in terms of RMSE and UQI metrics, than do the other algorithms.

C Image reconstruction for imaging configurations with array gaps

We have also reconstructed images from data acquired with each of the four imaging configurations shown in Fig. 2 by use of EM, POCS, and TV algorithms. For the two configurations with four arrays, as shown in column 1 of Fig. 2, we display reconstruction results in Figs. 7 and 8, whereas for the two configurations with six arrays, as shown in column 2 of Fig. 2, we display reconstruction results in Figs. 9 and 10. Again, results in the study suggest that images without significant artifacts may be reconstructed for the non-conventional configurations with array gaps by use of the TV algorithm. On the other hand, other algorithms appear to be less effective than does the TV algorithm for imaging configurations with array gaps, as their images contain artifacts that seem to be more prominent than those in the TV images. Also, using the true phantom images in Fig. 3 as the reference images and the corresponding reconstructed images in Eqs. (1) and (2), we calculated the RMSEs and UQIs which are

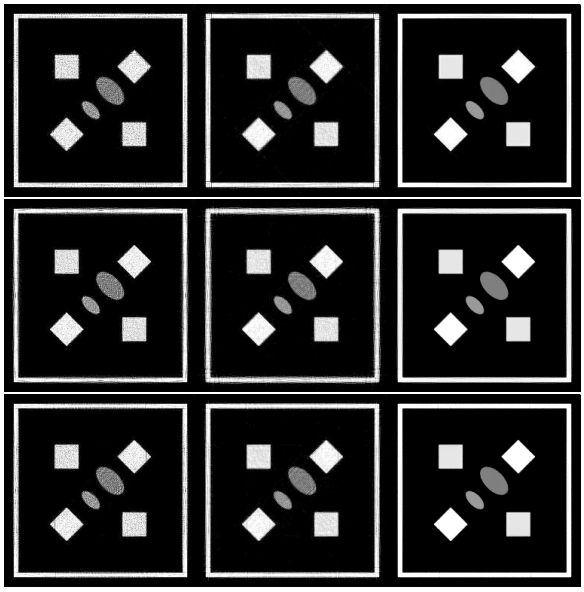


Fig. 4. Images reconstructed for the circular (top row), square-shaped (middle row), and hexagon-shaped (bottom row) configurations shown in Fig. 1 by use of the EM (column 1), POCS (column 2), and TV (column 3) algorithms. Display window: [0.0,2.0].

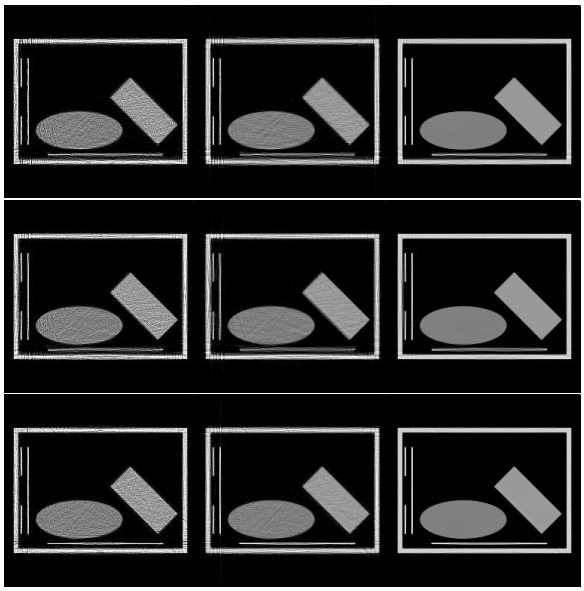


Fig. 5. Images reconstructed for the circular (top row), square-shaped (middle row), and hexagon-shaped (bottom row) configurations shown in Fig. 1 by use of the EM (column 1), POCS (column 2), and TV (column 3) algorithms. Display window: [0.0,2.0].

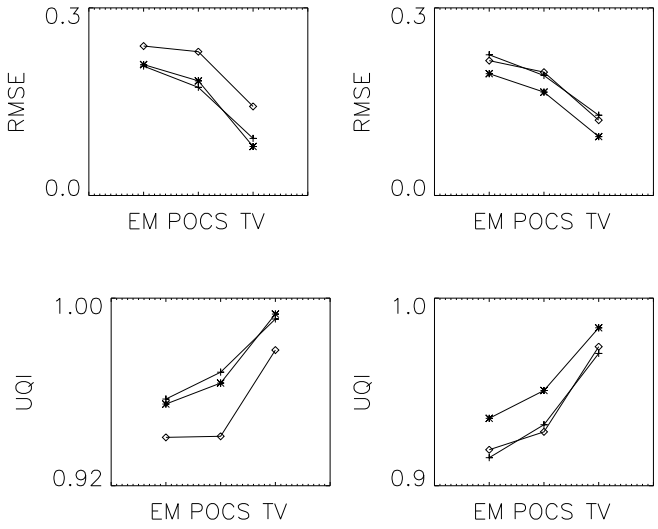


Fig. 6. RMSE and UQI results for the circular ("+"), square-shaped ("$\diamond$") and hexagon-shaped ("$\star$") configurations for phantoms 1 (left column) and 2 (right column), respectively. The three data points (from left to right) denoted with the same symbols are obtained with the EM, POCS, and TV algorithms.

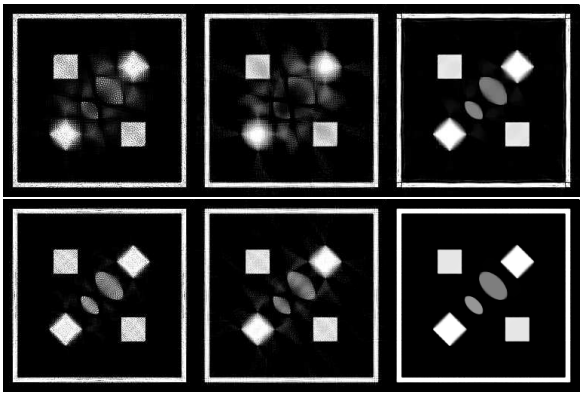


Fig. 7. Images reconstructed for the circular (upper row) and square-shaped (lower row) configurations shown in the left column of Fig. 2 by use of the EM (column 1), POCS (column 2), and TV (column 3) algorithms. Display window: [0.0,2.0].

displayed in Figs. 11 and 12 for the configurations in Fig. 2. These quantitative results also indicate that the TV algorithm generally yield images of high quality, in terms of RMSE and UQI metrics, than do the other algorithms.

V Conclusions

In this work, we have investigated image reconstruction from data acquired in CT systems consisting of arrays of stationary CNT sources by using recently developed TV algorithms. In particular, we have studied whether images without significant artifacts can be reconstructed for such systems with a number of CNT sources much lower than the number of projection views in conventional CT.

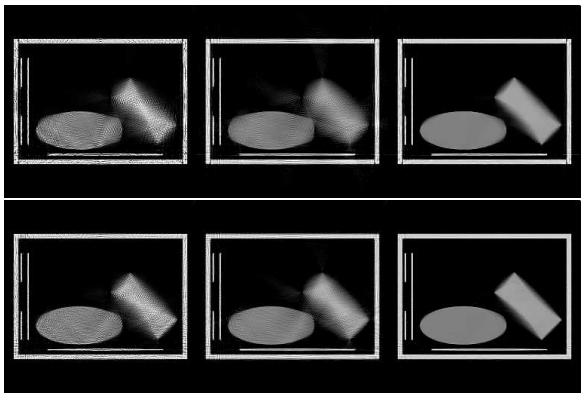


Fig. 8. Images reconstructed for the circular (upper row) and square-shaped (lower row) configurations shown in the left column of Fig. 2 by use of the EM (column 1), POCS (column 2), and TV (column 3) algorithms. Display window: [0.0,2.0].

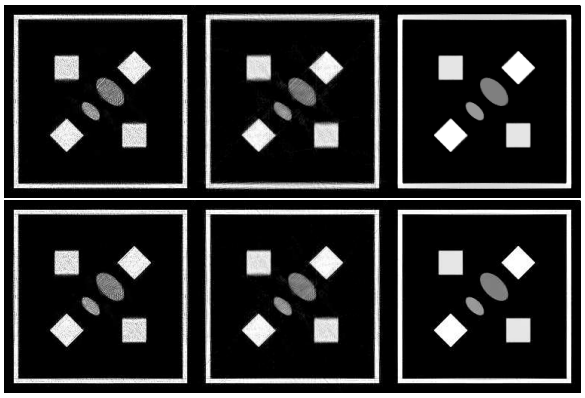


Fig. 9. Images reconstructed for the circular (upper row) and hexagon-shaped (lower row) configurations shown in the right column of Fig. 2 by use of the EM (column 1), POCS (column 2), and TV (column 3) algorithms. Display window: [0.0,2.0].

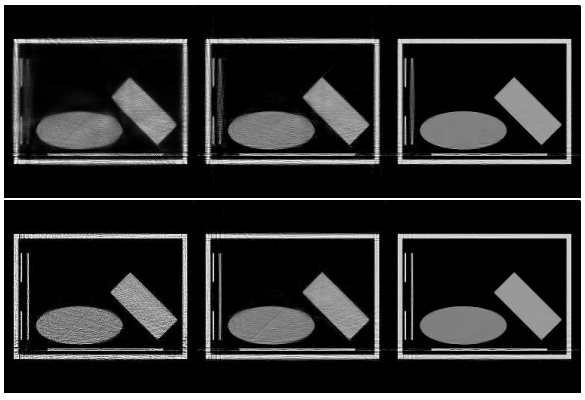


Fig. 10. Images reconstructed for the circular (upper row) and hexagon-shaped (lower row) configurations shown in the right column of Fig. 2 by use of the EM (column 1), POCS (column 2), and TV (column 3) algorithms. Display window: [0.0,2.0].

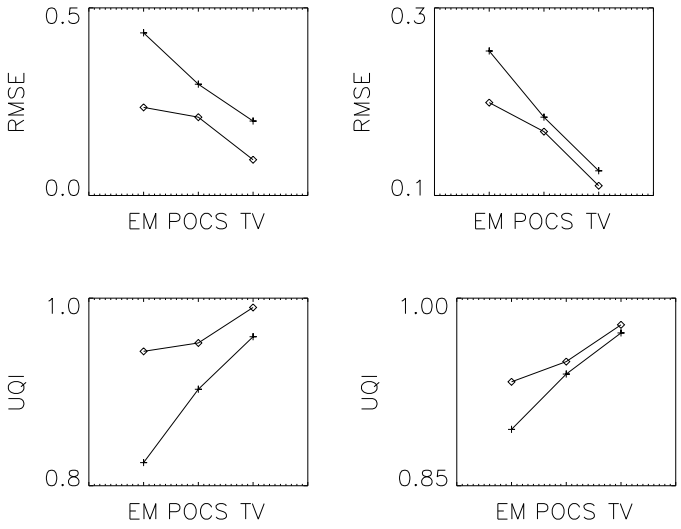


Fig. 11. RMSE and UQI results for the circular ("+") and square-shaped ("◇") configurations for phantoms 1 (left column) and 2 (right column), respectively. The three data points (from left to right) denoted with the same symbols are obtained with the EM, POCS, and TV algorithms.

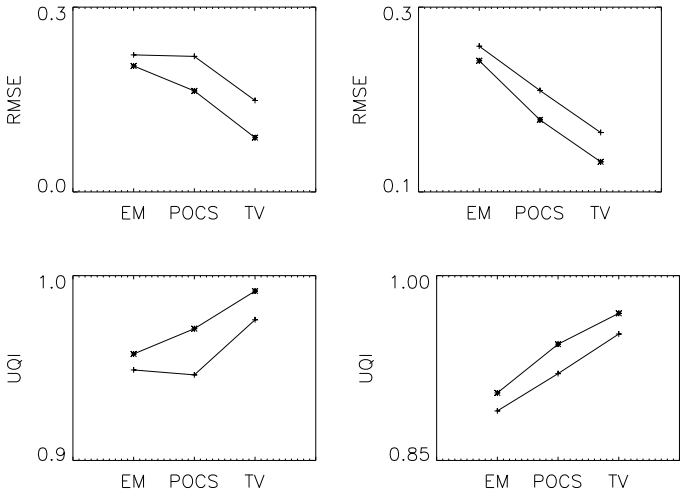


Fig. 12. RMSE and UQI results for the circular ("+") and square-shaped ("⋆") configurations for phantoms 1 (left column) and 2 (right column), respectively. The three data points (from left to right) denoted with the same symbols are obtained with the EM, POCS, and TV algorithms.

Also, because the cost for a linear array of CNT sources is considerably lower than that for a curved array, we have investigated image reconstructions for imaging configurations with linear and curved arrays of CNT sources. In our studies, we have carried out quantitative comparison of the performance of the TV algorithm and some well-known, existing algorithms. Results of the comparative studies demonstrate that the TV algorithm can reconstruct images better than those obtained with the existing

algorithms for imaging systems with a number of stationary CNT sources that is much smaller than the number of projections in conventional CT imaging. Based upon the results, observations can also be made that images with virtually identical quality can be reconstructed for CT systems with linear and curved source arrays and thus that array shapes result little difference in reconstructed images. The work seems to indicate that a low-cost CT system of flexible configuration with linear arrays of CNT sources can yield images comparable to that of conventional CT. It is likely that such low-cost, flexible CT systems will find significant applications in biomedical imaging, security scan, non-destructive inspection, and many other areas.

References

[1] G. Yue, Q. Qiu, B. Gao, Y. Cheng, J. Zhang, H. Shimoda, S. Chang, J. Lu, and O. Zhou, "Generation of continuous and pulsed diagnostic imaging x-ray radiation using a carbon-nanotube-based field-emission cathode," *Applied Physics Letters*, vol. 81, pp. 355–357, 2002.

[2] J. Zhang, G. Yang, Y. Cheng, , B. Gao, Q. Qiu, Y. Z. Lee, J. Lu, and O. Zhou, "Stationary scanning x-ray source based on carbon nanotube field emitters," *Applied Physics Letters*, vol. 86, pp. 184104, 2005.

[3] J. Zhang, G. Yang, Y. Z. Lee, S. Chang, J. Lu, and O. Zhou, "Multiplexing radiography using a carbon nanotube based x-ray source," *Applied Physics Letters*, vol. 89, pp. 064106, 2006.

[4] Z. Liu, Y. Guang, Y. Z. Lee, D. Bordelon, J. Lu, and O. Zhou, "Carbon nanotube based microfocus field emission x-ray source for microcomputed tomography," *Applied Physics Letters*, vol. 89, pp. 103111, 2006.

[5] X. Qian, R. Rajaram, X. Calderon-Colon, G. Yang, T. Phan, D. S. Lalush, J. Lu, and O. Zhou, "Design and characterization of a spatially distributed multibeam field emission x-ray source for stationary digital breast tomosynthesis," *Med. Phys.*, vol. 10, pp. 4389–4399, 2009.

[6] E. Y. Sidky, K.-M. Kao, and X. Pan, "Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT," *Journal of X-Ray Science and Technology*, vol. 14, pp. 119–139, 2006.

[7] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Physics in Medicine and Biology*, vol. 53, pp. 4777–4807, 2008.

[8] R. Gordon, R. Bender, and G. T. Herman, "Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography," *J. Theor. Biol.*, vol. 29, pp. 471–481, 1970.

[9] A.P. Dempster, N.M. Laird, D.B. Rubin, et al., "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.

[10] K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography," *J. Comput. Assisted Tomogr.*, vol. 8, pp. 306–316, 1984.

[11] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE. Trans. Image Proc.*, vol. 13, pp. 600–612, 2004.

Reconstruction from Sparse Data in Offset-Detector CBCT

[1]J. Bian, [2]J. Wang, [1]X Han, [1]E. Y. Sidky, [2]J. Ye, [2]L. Shao, and [1]X. Pan
[1]The University of Chicago, [2]Philips Healthcare

***Abstract*—Computed tomography (CT) unit has been added to advanced single-photon emission computed tomography (SPECT) for acquiring anatomic information about the imaged subject. CT images obtained can be used for attenuation correction in SPECT images and localization. Hardware and cost considerations have led to the use of a CT-imaging configuration in which the CT field of view (FOV) is extended through the offset of the X-ray detector. However, radiation dose in CT scans is of a concern. In this work, we investigate and develop algorithms, including the total-variation (TV) algorithm, for image reconstruction from data acquired at a considerably reduced number of projection views by use of the offset-detector CT unit of a commercially available SPECT/CT system. Results of the study suggest that the TV algorithm can reconstruct images of potential practical utility from projection views significantly fewer than what are used currently. A practical implication of the work may lie in that the use of well designed algorithms may allow for a significant reduction of CT scanning dose and time in SPECT/CT imaging through lowering the number of projection views.**

I Introduction

Single-photon emission computed tomography (SPECT) offers functional information within a subject through determining the spatial distribution of radiotracers within the subject. In general, SPECT images carry little anatomic information about the subject and also suffer from the attenuation effect of the radiotracer-emitted gamma rays experienced prior to their arrival at the detector. On the other hand, computed tomography (CT) can yield anatomic information of high quality about the subject. Therefore, combined SPECT/CT systems, which can yield information about functional process within, and anatomic structure of, the imaged subject, have been developed and made available commercially in recent years[1]. Furthermore, CT images with appropriately scaled CT numbers can be used to correct for the attenuation effect in SPECT images for improving their visual quality and quantitative accuracy.

Recently, a few flat panel based x-ray CT systems have been developed for isotropic volume imaging, which is used as a secondary imaging modality of a overall imaging or therapy device to improve the diagnostic or therapy outcomes[?]. Cost and hardware considerations have led to the use of a detector that can only offer a CT field of view (FOV) smaller than the cross section of a typical adult patient. However, an offset detector configuration can be used for expanding the CT FOV [3]. In this case, although data acquired at each of the projection views are truncated, it can be shown that, as long as the detector offset is less than the detector size, CT images without suffering truncation artifacts can be reconstructed from projection data collected over an angular range of 2π. Despite the fact that the inclusion of CT scans promises added-on values to SPECT imaging by providing information about patient anatomy and attenuation correction, radiation dose involved in CT scans constitutes a t concern. An approach to reducing CT-scan dose is to reduce the number of projection views at which CT data are acquired while maintaining the practical utility of CT images. Indeed, a byproduct benefit of using a small number of projection views in CT imaging is that the imaging time can also be reduced, thus decreasing the potential impact of patient-motion on CT imaging.

In this work, we investigate, develop, and implement a number of algorithms, including the recently developed TV algorithm, for CT-image reconstruction of potential, practical utility from data acquired at a number of projection views substantially lower than that used in current CT scans of SPECT/CT imaging. In addition to a computer-simulation study, which are not included in the paper, we have focused the study on image reconstruction from real cone-beam data collected with an offset-detector CT unit of a commercially available SPECT/CT system. For the cases studied, we demonstrate that CT images without significant artifacts can be reconstructed by use of carefully designed algorithms from data acquired at a considerably reduced number of projection views in CT scans with an offset-detector configuration.

II SPECT/CT System and CT Imaging Unit

In our study, the Philips BrightView XCT (SPECT/CT) system was used to collect data from a number of physical phantoms.

The XCT system, whose geometry is shown in Fig. 1, includes a CT unit composing an X-ray source and a flat panel detector, which, along with the SPECT unit, is mounted on a rotation gantry. In the CT unit, the distances of the X-ray source to detector and to the center of rotation are 133.2 cm and 88.1 cm, and the detector

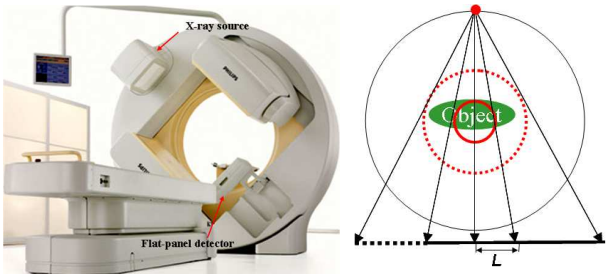


Fig. 1. The Philips BrightView XCT (SPECT/CT) system (left) and the schematic of the offset-detector configuration used in the XCT system (right). The red dot indicates the x-ray source and the horizontal thick line segment denote the detector. L indicates the offset of the detector.

panel consists of 1024×768 detection elements of sizes 0.388×0.388 mm^2. The detector panel is offseted horizontally by $L = 17.7$ cm, as indicated in Fig. 1, allowing a FOV of 47 cm(transaxial) $\times$ 14.4 cm(axial).

III Study Materials

A *Projection data and image array*

In the right panel of Fig. 1, we display schematically the offset configuration of XCT unit used in the study. In the configuration, the detector is offset horizontally by 17.7 cm with respect to the middle line connecting the X-ray source and center of rotation. As such, only the portion of the detector on the left side of the middle line will collect (quasi) redundant information when the detector is rotated over 2π.

We have also performed real data studies involving different physical phantoms. we report in the paper some of the results obtained in the pelvic phantom study (CIRS virtually human pelvis phantom). The pelvic phantom, as shown in 2, is made from proprietary epoxy materials that mimic the density and radiation attenuation properties of human tissue . It contains anatomically precise bone, cartilage, spinal cord, vertebral disks, muscle, intestines, bladder, prostate, rectum and interstitial fat.

In the study, we have collected cone-beam data at 720 projection views evenly distributed over 2π of a circular trajectory from a number of physical phantoms, including the pelvic phantom. Throughout the work, we refer to a 720-view data set as *the full data set*. The pelvis phantom data are acquired using 2x4 binning, resulting a projection size of 1024×384 for each view, and a detector pixel size of 0.388×0.776 mm^2. Projection data corrected for background, uniformity, detector gain mode, and defect pixels. Low-frequency drop and off-focal radiation are both compensated for by deconvolution methods and the scatter is corrected by a kernel-based correction method.

Because a focus of the study is image reconstruction from data sets containing a small number of views, we thus extract form three such data sets by selecting, from the full data set, data at 72, 120, and 180 views, respectively, uniformly distributed over 0 to 2π.

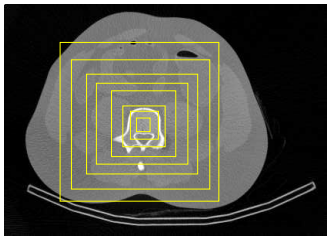


Fig. 2. The region enclosed by each of the eight squares indicates a region of interest (ROI) within which the Pearson correlation coefficient is calculated for evaluating the similarity of a reconstructed image to the reference image.

In the study, we reconstruct an image on a discrete array consisting of $300 \times 420 \times 189$ voxels of 1 mm^3 size, representing 189 transverse slices of dimension 300×420.

B *Reconstruction algorithms*

We have recently developed an algorithm for image reconstruction from sparse data. The algorithm reconstructs an image by the minimization of the image's total variation (TV) subject to the data and other constraint conditions [4, 5]. The focus of the work is to investigate and evaluate the performance of the TV algorithm in image reconstruction from these data sets. However, for the purpose of comparison, we have also investigated existing algorithms such as the FDK and projection-onto-convex-set (POCS) for image reconstruction from the same data sets. Some modifications to the reconstruction algorithms are necessary so that they can accommodate adequately the partially (quasi) redundant information contained in these data sets collected with an offset-detector geometry. Modified FDK algorithms exist for dealing with offset data sets. In the work, we have adjusted the POCS and TV algorithms so that they can adequately deal with the partially (quasi) redundant data information. Details of the modifications to the POCS and TV algorithms will be reported at the conference.

C *Evaluation metrics*

In addition to visual comparison of images reconstructed by use of the algorithms, we have also used a number of metrics described below for quantitative evaluation of reconstructions. In the calculation of the metrics, a reference image (i.e., the "ground truth") is needed. For real data studies, there is no ground truth available. As such, we will use the FDK reconstruction from the full data set as *the reference image* because the FDK reconstruction from full data is used widely current SPECT/CT imaging practice.

Pearson correlation coefficient The Pearson correlation coefficient can be used as an index measuring the similarity of a reconstructed image to a reference image

within a selected region of interest (ROI), and it is defined as

$$\rho(f, f^r) = \frac{Cov(f, f^r)}{\sigma(f)\sigma(f^r)}, \quad (1)$$

where f and f^r denote the reconstructed and reference images, $\mu = \frac{1}{N}\sum_i f_i$ and $\mu_r = \frac{1}{N}\sum_i f_i^r$ the mean values of the reconstructed and reference images, $\sigma^2(f) = \frac{1}{N-1}\sum_i (f_i - \mu)^2$ and $\sigma^2(f^r) = \frac{1}{N-1}\sum_i (f_i^r - \mu_r)^2$ the "variance" of the reconstructed and reference images, and $Cov(f, f^r) = \frac{1}{N-1}\sum_{ij}(f_i - \mu)(f_j^r - \mu_r)$ the covariance between the reconstructed and reference images. It should be noted that the summations above are all carried out over the voxels within a selected ROI of the images. The higher the correlation coefficient, the more similar the reconstructed image to the reference image. Clearly, two identical images lead to a perfect correlation $\rho = 1.0$.

Attenuation factors An important application of CT images in SPECT imaging is to use them for estimating the attenuation factors that can be incorporated in the reconstruction of SPECT images for potentially improving their quantitative accuracy. In general, a more accurate estimate of attenuation factors is likely to lead to an improved attenuation correction in SPECT imaging. The attenuation factor is defined as

$$A(\vec{r}, \hat{\alpha}) = \mathrm{e}^{-\int_{\vec{r}}^{\infty} dl\, f(\vec{r})}, \quad (2)$$

where $f(\vec{r})$ denotes a CT estimate of the attenuation distribution at location $\vec{r}$, and $\hat{\alpha}$ the direction of the line along which the integral is carried out. In the study, we calculate attenuation factors using images reconstructed from data sets containing small numbers of views by use of the reconstruction algorithms described above and compare them with those obtained from the reference image.

IV Results

Using the FDK, POCS, and TV algorithms modified to handle the partially (quasi) redundant data information, we have reconstructed images from data sets containing 72, 120, and 180 projection views, respectively, which, as described above, were selected from the full data set acquired at 720 projection views.

A Image reconstruction

We first reconstructed images by use of the FDK, POCS, and TV algorithms from the 72-view data set and display them in Fig. 3. For comparison, we also display the image reconstructed from the full data set by use of the FDK algorithm. As expected, the FDK reconstruction from 72-view data suffers from aliasing artifacts due to the lack of sufficient angular samples. Although the POCS algorithm can reduce somewhat such artifacts, some artifacts from

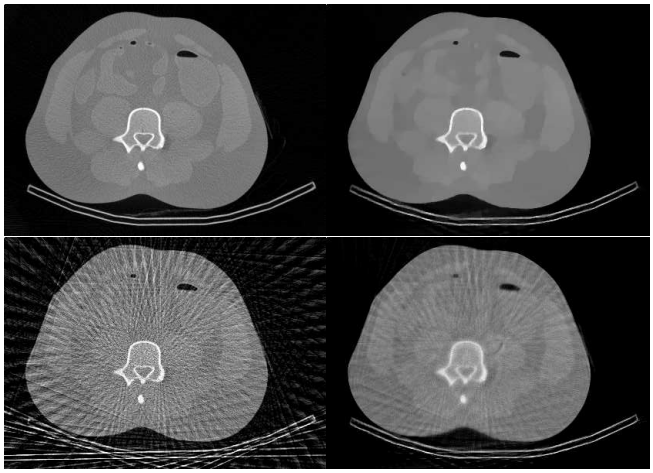


Fig. 3. Images reconstructed from the 72-view data set by use of the TV (top right), FDK (bottom left), POCS (bottom right) algorithms. For comparison, the FDK reconstruction from the full data set is also displayed in the top-left panel as the reference image. Display window: $[0.0, 0.4]$ cm^{-1}

sampling aliasing and data noise remain in the reconstructions. However, it can be observed that the TV image is virtually free of such aliasing artifacts and that it appears to comparable to the FDK reference image reconstructed from the full data set.

We then reconstructed images by use of the FDK, POCS, and TV algorithms from the 120- and 180-view data sets and display them in Figs. 4 and 5, respectively. Again, we also display the image reconstructed from the full data set by use of the FDK algorithm for the purpose of comparison. In these cases, despite the fact that as the numbers of projection views increase, the aliasing artifacts in FDK and POCS reconstructions diminish considerably, the effect of the lack of angular samples remains visible. On the other hand, it can be observed that the quality of TV reconstructions, in terms of spatial and contrast resolution, appears to be enhanced as the number of views increases. Also, the TV reconstruction appears to be less noisy than the FDK reference image.

B Quantitative evaluation

For an image reconstructed by use of one of the three algorithms from one of the selected data sets, we have also calculated by using Eq. (1) its Pearson correlation coefficients with respect to the reference image, and the results are displayed in Figs. 6, 7, and 8, respectively. Based upon the results, it can be observed that the TV reconstructions have consistently higher Pearson correlations than other reconstructions, suggesting that the TV algorithm yields images closer to the reference image than other algorithms.

A leading utility of CT images in SPECT/CT imaging is for attenuation correction in SPECT. In an attempt to further evaluate this utility of CT images reconstructed

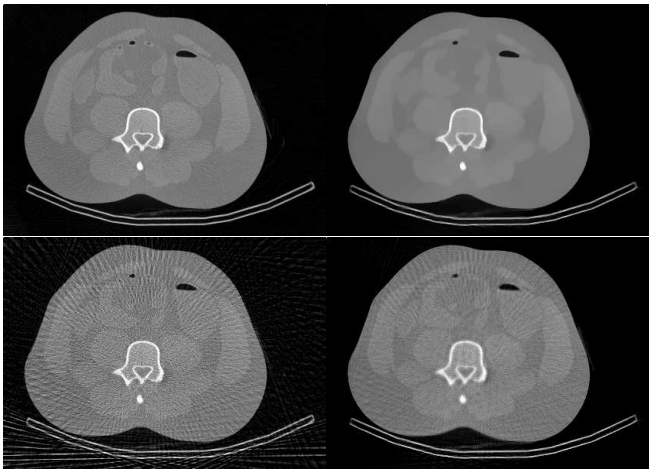


Fig. 4. Images reconstructed from the 120-view data set by use of the TV (top right), FDK (bottom left), POCS (bottom right) algorithms. For comparison, the FDK reconstruction from the full data set is also displayed in the top-left panel as the reference image. Display window: $[0.0, 0.4]$ cm^{-1}

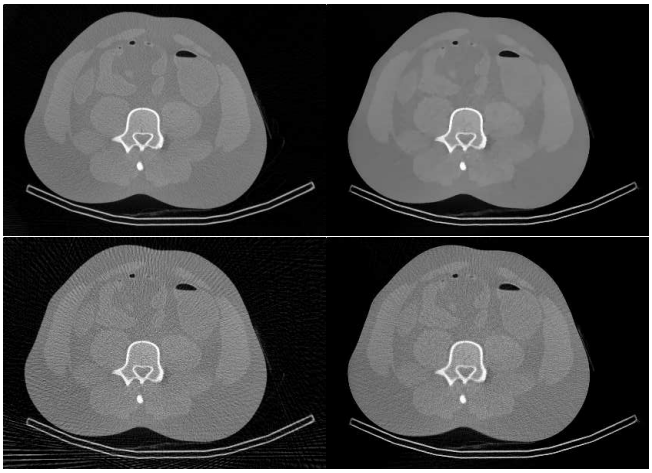


Fig. 5. Images reconstructed from the 180-view data set by use of the TV (top right), FDK (bottom left), POCS (bottom right) algorithms. For comparison, the FDK reconstruction from the full data set is also displayed in the top-left panel as the reference image. Display window: $[0.0, 0.4]$ cm^{-1}

by use of the algorithms for data containing relative small number of projections, we have also calculated the attenuation factors defined in Eq. (2) based upon the reconstructed images. In Fig. 9, we show the attenuation factors computed for a group of selected points within each of the images reconstructed by use of the FDK, POCS, and TV algorithms from the 72-view data set. For comparison, the attenuation factors computed for the same group of points within the reference image are also displayed in Fig. 9. Furthermore, the 72-view attenuation-factor result for each algorithm and the reference attenuation factor shown in Fig. 9 can be interpreted as an "image" and a "reference image". Therefore, using the results in Eq. (1), we also calculated the Pearson correlation coeffi-

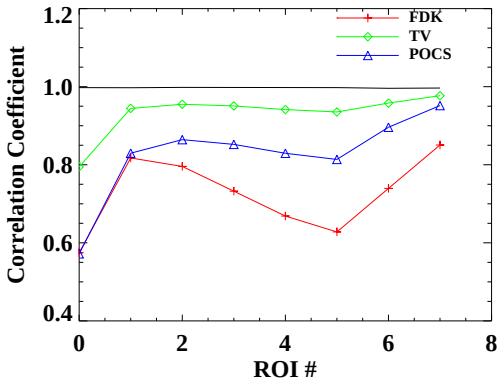


Fig. 6. Pearson correlation coefficients of images reconstructed from the 72-view data by use of the FDK (+), POCS (△), and TV (⋄) algorithms for the eight ROIs indicated in Fig. 2.

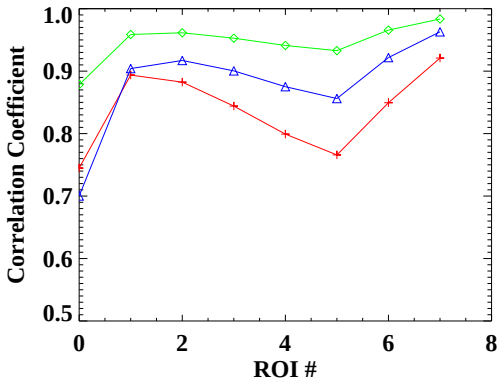


Fig. 7. Pearson correlation coefficients of images reconstructed from the 120-view data by use of the FDK (+), POCS (△), and TV (⋄) algorithms for the eight ROIs indicated in Fig. 2.

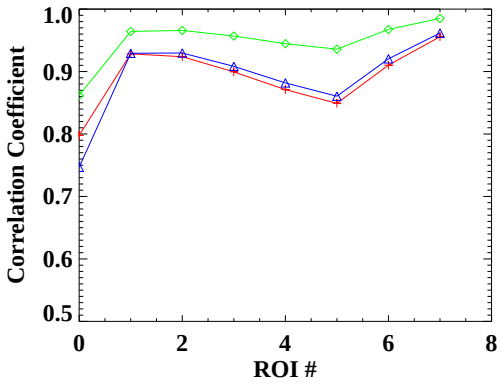


Fig. 8. Pearson correlation coefficients of images reconstructed from the 180-view data by use of the FDK (+), POCS (△), and TV (⋄) algorithms for the eight ROIs indicated in Fig. 2.

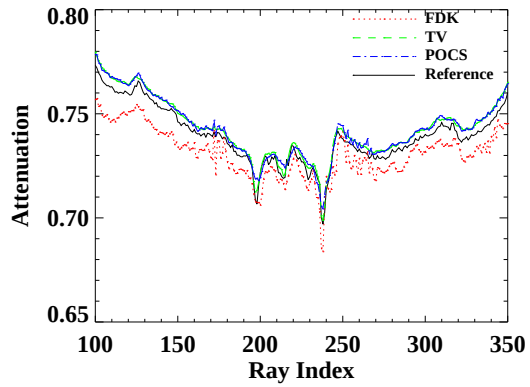


Fig. 9. Attenuation factors computed for a group of selected points within each of the images reconstructed by use of the FDK (dotted), POCS (dash-dotted), and TV (dashed) algorithms from the 72-view data set. For comparison, the attenuation factors computed for the same group of points within the reference image are also displayed (solid).

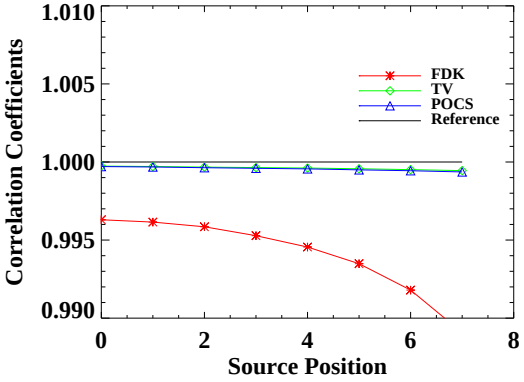


Fig. 10. Pearson correlation coefficients of the attenuation factors in Fig. 9, computed from images reconstructed by use of the FDK ($*$), POCS ($\triangle$), and TV ($\diamond$) algorithms, with respect to that computed from the reference image.

cients of the "image" to the "reference image" (i.e., for the entire group of the selected points,) and show the coefficient results in Fig. 10. Based upon the results in Figs. 9 and 10, it can be observed that, as far as the reference image is concerned, the TV algorithm yields more accurate attenuation factors than other algorithms, although the extent of accuracy improvement depends upon the algorithms, data quality, and the amount of data (such as the number of views) to be considered.

V Discussions

The general concern of radiation dose in CT imaging motivates the study on whether it is possible to lower radiation dose in CT imaging without sacrificing the CT-image utility for specific practical tasks. An approach to reducing the imaging dose is to lower the number of projection views. In this work, specifically, we have investigated and demonstrated accurate reconstruction of CT images from data collected with an offset detector at a number of projection views substantially smaller than that currently used in SPECT/CT imaging. The TV and POCS algorithms have been modified so that they can be applied to reconstructing images from data collected with an offset-detector CT unit of SPECT/CT system. In addition to visual inspection of reconstructed images, we have also used several metrics for quantitative evaluation of images reconstructed and attenuation factors estimated. We have carried out extensive studies in which simulation and real data were used, although only a small portion of the study results is presented due to the page constraint. We plan to report additional, detailed results at the conference. Results of our study demonstrate that the TV algorithm can reconstruct images from data containing projection views significantly fewer than the full data set currently acquired in a typical SPECT/CT scan and yield attenuation factors comparable to those obtained by use of the FDK algorithm from the full data set. A practical implication of the work may lie in that the use of a well designed algorithm such as the TV algorithm may allow for a significant reduction of CT imaging dose in SPECT/CT through lowering the number of projection views.

References

[1] B. Hasegawa, K. Wong, K. Iwata, W. Barber, A. Hwang, A. Sakdinawat, M. Ramaswamy, D. Price, and R. Hawkins, "Dual-modality imaging of cancer with SPECT/CT," *Technology in Cancer Research and Treatment*, vol. 1, no. 6, pp. 449–458, 2002.

[2] M. O'Connor and B. Kemp, "Single-photon emission computed tomography/computed tomography: basic instrumentation and innovations," in *Seminars in nuclear medicine*, vol. 36, no. 4. Elsevier, 2006, pp. 258–266.

[3] W. Chang, S. Loncaric, G. Huang, and P. Sanpitak, "Asymmetric fan transmission CT on SPECT systems," *Physics in Medicine and Biology*, vol. 40, pp. 913–928, 1995.

[4] E. Y. Sidky, K.-M. Kao, and X. Pan, "Accurate image reconstruction from few-views and limited-angle data in divergent-beam ct," *Journal of X-Ray Science and Technology*, vol. 14, pp. 119–139, 2006.

[5] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Physics in Medicine and Biology*, vol. 53, pp. 4777–4807, 2008.

Compensator Approaches for Intensity Modulated Computed Tomography

Steven Bartolac, Sean Graham, Jeff Siewerdsen and David Jaffray

Abstract --- **In computed tomography (CT), the tradeoff between image quality and imaging dose is an inherent factor in selecting the appropriate imaging technique. Modulation of the fluence intensity, spatially across the X-ray beam, and independently for each projection, has the potential to significantly improve or maintain regional image quality, while reducing radiation dose to the patient due to imaging. This approach is referred to as Intensity Modulated Computed Tomography (IMCT). In this work several compensator approaches for intensity modulation in CT are evaluated. Modulation profiles were generated using a simulated annealing optimization method that attempts to achieve prescribed spatially-dependent SNR criteria in a simulated water phantom. The resulting SNRs for each compensator approach were analysed in terms of regional SNR. This study suggests that the addition of constraints may increase feasibility by reducing the number of physical variables while still maintaining improved regional SNR performance.**

INTRODUCTION

Computed tomography continues to advance in terms of increased speed, resolution, image quality and volume coverage capabilities. These advancements make CT an attractive candidate for a growing number of applications. However, more extensive use of the modality is hindered by the radiation cost to the patient [1]. Cone-beam computed tomography (CBCT), for example, has received considerable attention for its benefits to patients in the fields of image-guided radiotherapy and surgery; despite these benefits, current practice generally limits the number of CT scans a patient can receive and/or restricts images to lower quality in order to minimize radiation dose to the patient due to imaging.

Steven Bartolac is with the Department of Medical Biophysics, University of Toronto, 610 University Ave., Toronto, Ontario M5G 2M9 CANADA. (Tel: (416) 946-4501. Email: steve.bartolac@rmp.uhn.on.ca)

Sean Graham was with the Department of Medical Biophysics, University of Toronto, 610 University Ave., Toronto, Ontario M5G 2M9 CANADA. He is now with Bubble Technology Industries, Chalk River, Ontario, K0J 1J0 CANADA. (Email: sean.graham@rmp.uhn.on.ca)

Jeff Siewerdsen was with the Radiation Medicine Program, Princess Margaret Hospital/Ontario Cancer Insitute, 610 University Ave., Toronto, Ontario, M5G 2M9 CANADA. He is now with the Department of Biomedical Engineering, Johns Hopkins University, 720 Rutland Ave., Baltimore, Maryland 21205. (Email: jeff.siewerdsen@jhu.edu)

David Jaffray is with the Radiation Medicine Program, Princess Margaret Hospital/Ontario Cancer Insitute, 610 University Ave., Toronto, Ontario, M5G 2M9 CANADA. (Email: david.jaffray@rmp.uhn.on.ca)

Much research has been devoted to decreasing dose to the patient while maintaining or bettering image quality. Some approaches, including automatic exposure control [2] and the addition of bow-tie filters [3-5], have made strides towards more efficient management of the X-ray fluence. However, the ideal allocation of X-ray fluence would take into account both patient-specific anatomy of the patient as well as the specific imaging task. Previous work [5,6] has shown that modulation of the fluence intensity across the X-ray beam, independently for each projection, has the potential to improve or maintain image quality where it is required, and allow for poorer image quality elsewhere. Such an approach could potentially yield great dose sparing to the patient without cost in the utility of the images. Benefits would extend across both diagnostic and image-guided techniques. The authors refer to this approach as Intensity Modulated Computed Tomography (IMCT).

A principal challenge in the design of an IMCT system is choosing an appropriate compensator approach for the modulation. It is likeley that a feasible approach will include physical constraints that reduce the number of variables by limiting the number of allowable modulation patterns. This work evaluates different approaches for their ability to deliver prescribed signal-to-noise ratio (SNR) values.

THEORY AND METHODS

Implicit to the following approach is that *a priori* information about the object is available. This *a priori* knowledge may come from a previous CT scan, or a population based model for example. Various analytic methods are then available for predicting the noise outcome of different fluence patterns. The approach taken is to pose a desired SNR distribution and to attempt to solve the inverse problem of finding a modulation profile that generates the desired SNR criteria via an optimization routine.

In this study we make the assumptions of Poisson distributed noise based on photon counting statistics, a parallel ray geometry, and a filtered backprojection reconstruction routine for analysis. Under these conditions, the expected variance at each voxel position (x,y) can be described by [7]

$$\mathrm{var}\{f(x,y)\} = \left(\frac{\pi\tau}{M}\right)^2 \times \sum_{i=1}^{M}\sum_{t}\frac{h^2(x\cos\theta_i + y\sin\theta_i - t)}{m_{\theta_i}(t)\bar{I}_{\theta_i}(t)}, \qquad (1)$$

where $f(x,y)$ represents the object function, (θ_i , t)are coordinates that parametrize the projection (and backprojection) ray paths as shown in Fig. 1, τ is the detector pixel width, M is the number of projections, $\bar{I}_{\theta_i}(t)$ is the output intensity of the X-ray fluence before modulation, $m_{\theta_i}(t)$ is the applied modulation factor, and h is the convolution kernel of the filtered backprojection operation.

All experiments were carried out in Matlab. A simulated, cylindrical, water phantom with a diameter of 28 cm was used in all analyses, as shown in Fig. 2 (a). Projections were taken over 180^o with 192 intervals. In order to achieve faster optimization times, modulation profiles were generated on low resolution images (0.54×0.54×0.5 cm voxel size). The prescribed, target SNR values were defined for varied regions of interest in the object, and stored in the matrix SNR_d (where the subscript denotes distinction from other matrices, and not index number), as illustrated in Fig. 2 (b). The target high SNR value was chosen such that it would result in a target standard deviation of noise of 1% with respect to water at 1 mm resolution in x and y.

In general, arbitrary SNR values cannot be achieved because the filtered backprojection method introduces correlations between the noise at different image locations. Moreover, various fluence profiles may produce the same image quality result but have different dose outcomes. Therefore, an iterative method was used to find a modulation profile that minimizes the cost function

$$C_n = \frac{\left(\sum_{x,y}(W_s(SNR_n - SNR_d))^2\right)}{\left(\sum_{x,y}(W_s(SNR_o - SNR_d))^2\right)} + w\frac{D_n}{D_o}, \qquad (2)$$

where W_s is a matrix of weights on the desired image qualities, subscripts *0* and *n* denote initial and current iteration number respectively, *D* is proportional to the total dose, and *w* is a separate weight that defines the relative importance of the dose term. The weightings can be set depending on task-specific priorities. In this study, *w* was arbitrarily set to 1, and the weights in W_s were chosen to

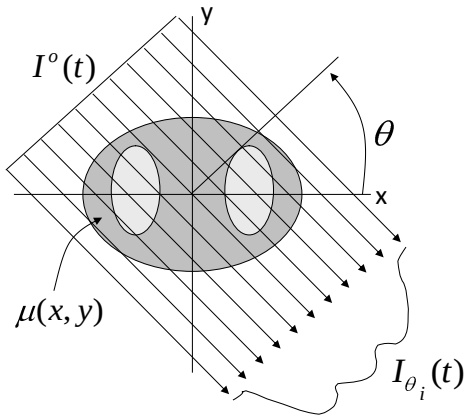


FIGURE 1: Schematic illustrating the parallel ray geometry and associated coordinates.

boost the priority of the high SNR region, and give intermediate priority to the low SNR region. Note that the first term in (2) requires updating the SNR values at each iteration of the optimization algorithm, which can be achieved by application of (1).

Optimization can be performed using any choice of feasible algorithm. In this case, we chose a simulated annealing approach. Advantages of simulated annealing algorithms are that they allow 'uphill' moves which allow the solution to escape local minima. Long iteration times were used for the following studies in order to allow for good convergence. Further details on simulated annealing can be found elsewhere [5,8].

Constraints were applied by limiting the number of modulation factors allowed, and/or the number of projections over which the factors can vary. The following cases were considered:

(i) no constraint on number of modulation factors
(ii) three (3) modulation factors per projection
(iii) one (1) modulation profile for all projections
(iv) eight (8) modulation factors for every 16 projections and at fixed locations

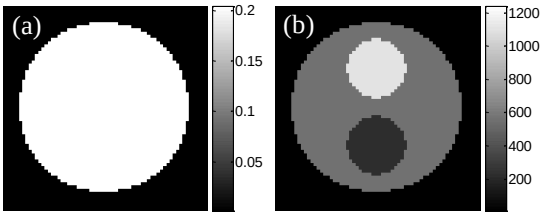


FIGURE 2: (a) simulated water phantom and (b) prescribed SNR regions within water phantom. SNR values appear high because the images are at low resolution.

As an example of how these constraints may be interpreted physically, (ii) could conceptually be achieved by a system of two moving collimator leaves; (iii) would require a fixed compensator, preset before the scan begins; and (iv) might involve a series of stationary filters that could be changed at large angular intervals. Case (i) is used as a reference scenario and can be considered to be that of an ideal, unconstrained modulator. Analysis of the data was performed by comparing the predicted SNR values of (2) against the prescribed values for each case.

RESULTS AND DISCUSSION

Modulation profiles generated by the optimization routine are shown in Fig. 3. In the unconstrained case, the light and dark regions indicate the regions of greatest and lowest pass of fluence intensity to the object, respectively. Spatially, these regions correspond to the prescribed high and low SNR regions in the object. Similarly, the constrained cases show similar trends but with reduced degrees of freedom. It is noted that in case (ii) there are small deviations that appear in the arc shaped light region instead of the smooth curve seen in case (i). While they appear to be artefacts of the algorithm, these abrupt deviations probably substitute the more subtle drop in fluence seen in case (i) near the same regions, and are likely important for compensating the thinner areas of the phantom, in order to more closely match the uniform prescribed SNR regions.

The resulting SNR values from the generated modulation profiles are also shown in Fig. 3. In all cases, the high SNR region shows greatest agreement with the prescribed values, as was expected since the cost function was weighted to give priority to this region. Overall, the unconstrained case shows the best performance, also as expected. In the remaining cases, the optimal method based on SNR outcomes is the second case, which generated values that agreed most closely to the prescribed ones. For the case of a single modulation profile for all projections, there is also success in achieving the high SNR region, but this region extends well into the prescribed low and intermediate SNR regions. Case (iv) shows more distinction between high, low and intermediate SNR values but has a more irregular pattern, with rectangular edges and less smooth transitions. In general, all cases succeeded in generating SNR distributions that provided regionally high SNR performance while limiting fluence to external regions. Reconstructions of the cylinder are also shown in Fig. 4, where random Poisson noise was added to the projections prior to reconstruction according to each modulation profile. For better visual comparison between high and low SNR regions, three inserts were added to each region, where the inserts deviate from the signal of water by 2%. Observations of these images reinforce those of the SNR distributions: the image generated using a modulation profile constrained to three modulation factors per projection achieves results closest to the ideal unconstrained scenario.

These outcomes suggest that the introduction of constraints to the modulation scheme may lead to substantial decreases in exposure to the patient due to X-rays while still achieving regional, high performance SNRs where prescribed by the user. The reduction in variables due to the physical constraints further suggests that these methods can lead to a more feasible implementation of IMCT than an unconstrained,

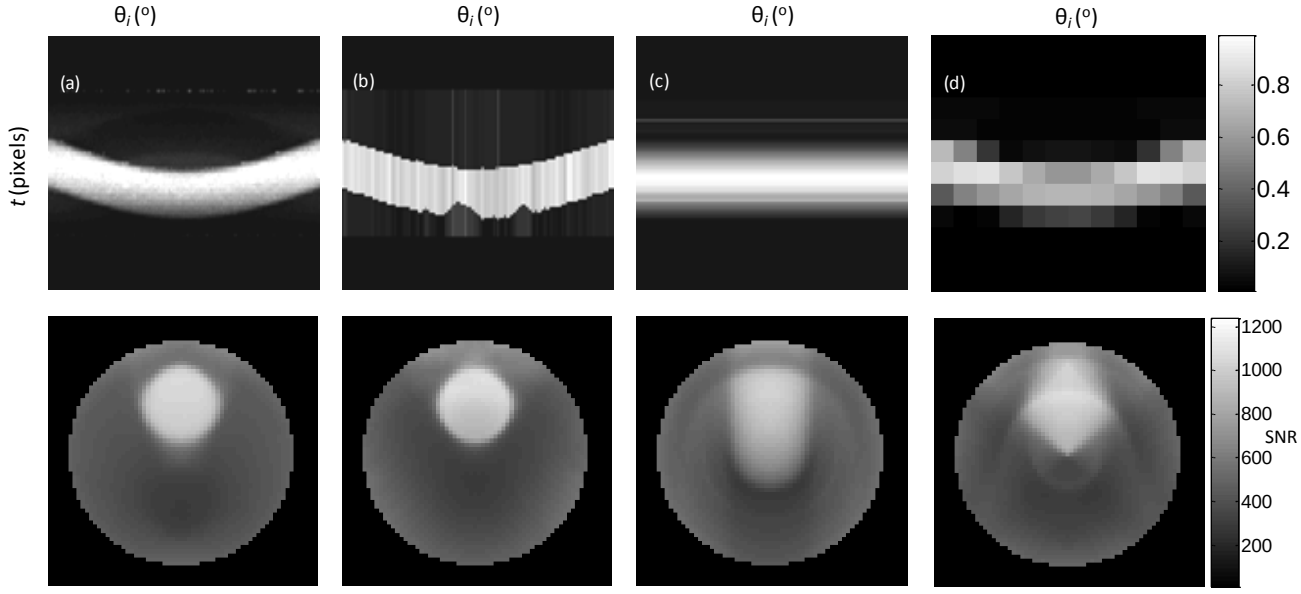


Figure 3 : Modulation profiles generated to meet specified SNR criteria for the case of (a) unconstrained modulation, (b) modulation constrained to three factors per projection, (c) modulation at fixed intervals in detector position, t, and angular interval, θ_i, and (d) single modulation profile over all angles. The resulting SNR distributions for each case are shown below the profiles. Contrast in modulation profiles has been enhanced to allow for greater visibility of low intensity modulation factors. Values of SNR may appear high because resolution of images is low.

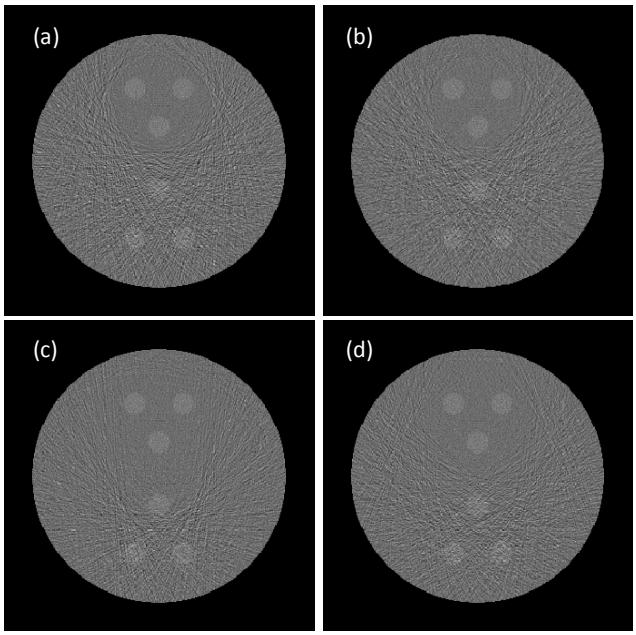


FIGURE 4: Reconstructed images with simulated noise for the case of (a) unconstrained modulation, (b) modulation constrained to three factors per projection, (c) modulation at fixed intervals in detector position, t, and angular interval, θ_i, and (d) single modulation profile over all angles. 'Inserts' have been added to the high and low SNR regions of the water cylinder for better visualization.

continuous modulation approach. The optimal choice suggested by this study is the approach where modulation factors are limited to few values per projection, but allowed to vary in extent and location. However, the best choice in compensator method will depend on additional factors, including mechanical design considerations. A more comprehensive comparison of techniques will also need to include scatter and beam hardening effects in the model, which were ignored in this simplified study, tolerances set by the user for desired dose and SNR preferences, and a more complete analysis of possible variations on the techniques shown. It is also recommended that more anatomic test objects, and alternative SNR distributions should also be explored.

CONCLUSION

The results of this limited study suggest that regional, high SNR performance can be achieved, with reduced fluence to surrounding regions, through the use of largely constrained modulation profiles. The optimal choice in compensator method will ultimately depend on further analysis of mechanical design factors, user-defined tolerances for SNR distributions, and the inclusion of more complicated X-ray physics (beam hardening, scatter).

REFERENCES

[1] D.J. Brenner and E.J. Hall, "Computed Tomography – An Increasing Source of Radiation Exposure," *NEJM* vol. 357, pp. 2277-2284, 2007.

[2] M. Kachelriess, W. Kalender, and S. Schaller, "Strategies for Dose Reduction and Improved Image Quality in MSCT" in *Multidetector-Row CT of the Thorax*, Berlin: Springer, 2007, pp 35-45

[3] N. Mail, D. Moseley,J.H. Siewerdsen and D.A. Jaffray, "The influence of bowtie filtration on cone-beam CT image quality," *Med Phys.* vol. 36 (1), pp. 22-32, 2009.

[4] F. Noo and A. Wunderlich, "Achieving uniform noise in direct fan-beam CT reconstruction through bowtie filter design," *Nuclear Science Symposium Conference Record* vol. 6, pp. 4379-4382, 2007.

[5] S.A. Graham, J.H. Siewerdsen, and D.A. Jaffray, "Compensators for dose and scatter management in cone-beam computed tomography," *Med. Phys.* 34 (7) 2691-2703, 2007.

[6] S.A. Graham, J.H. Siewerdsen, and D.A. Jaffray, "Intensity-modulated fluence patterns for task specific imaging in cone-beam CT" *Proc. SPIE* vol. 6510, pp. 651003, 2007.

[7] A.C. Kak and M. Slaney 1988, *Principles of computerized tomographic imaging*, IEEE Press, p 196, 1988.

[8] S. Kirkpatrick, C.D. Gelatt, Jr., and M.P. Vecchi, "Optimization by Simulated Annealing", *Science* vol. 220, pp. 671-680, 1983.

Spectral Shaping for Ultra-Low Dose CT Attenuation Correction in PET/CT

Ting Xia, Adam M Alessio, Paul E Kinahan

***Abstract*– We explore the bias and noise characteristics of attenuation corrected (AC) PET images formed from ultra-low dose (sub-millisievert) CT imaging. Under the constraint of using current clinical x-ray components, the overall goal of this work is to find the combination of methods that (i) minimize CT dose and (ii) suppress noise and bias artifacts. In this study we determine x-ray spectra that reduce radiation dose, while maintaining AC-PET image quality. We simulate CT and PET imaging with analytic simulation packages (CATSIM and ASIM) containing realistic models of x-ray tube spectra, beam conditioning, the bowtie filter, detector noise, and beam hardening. X-ray spectra were generated for a range of clinically viable filtration schemes with 80-140kVp generation, four filtration materials (Al, Cu, Mo, Sn), and filtration thicknesses of 0-1mm. The CT images, generated from each spectra across a range of tube currents, were used for attenuation correction of multiple realizations of PET data to evaluate noise and bias in the resulting AC-PET images. Results show that radiation absorbed doses of less than 0.11mGy can be achieved with less than a 5% increase in PET image bias. Furthermore, appropriate choice of tube potential and filtration can lead to lower doses at matched PET image quality, or vice versa. Specifically, spectra that are higher energy and more narrow are generally more dose efficient for equivalent PET image quality. Comparison of propagation of bias and noise into the PET images indicate that both bias and noise in the CT image determine quantitative accuracy in the PET image.**

I. INTRODUCTION

Current clinical CT components for PET/CT are essentially identical to standalone CT systems and were designed for high-detail anatomic imaging, acquired in short time frames. These CT components require high flux x-ray generators and high throughput detectors. The CT images in PET/CT scanners provide PET attenuation correction and can provide valuable side information for PET image enhancement, such as in 4D CT imaging for PET motion compensation or CT guided PET image reconstruction. In essentially all cases, the CT image has a higher signal to noise ratio than is necessary for PET image quality requirements. That is, the CT image could often be acquired at a dramatically lower technique, with a lower radiation dose to the patient, with no detriment to the attenuation corrected (AC) PET image. At the limit where low dose CT SNR and/or artifacts are detrimental to PET image quality, CT noise/artifact suppression techniques can be employed make low dose CT acquisitions acceptable for PET attenuation correction. The key to this approach is the recognition that since diagnostic quality CT image are not required for PET attenuation correction, significant reductions in patient radiation dose can be achieved by taking advantage of (1) lower SNR requirements, and (2) lower computational burden for CT image processing (e.g. respiratory gated and/or statistically-principled CT image reconstruction).

This effort explores the bias and noise characteristics of AC-PET images derived from ultra-low dose (sub-millisievert) CT imaging. There are numerous methods for reducing CT radiation dose such as reduction of mAs, x-ray tube pulsing, and beam conditioning. Our prior preliminary work explored the effect of adjusting kVp and reducing tube current below currently clinically available limits to determine the origin of the propagation of error into AC-PET images [1]. Through simulation studies, we showed that techniques can be used to significantly reduce the mAs needed for CT based attenuation correction if the CT is not used for diagnostic purposes. These results suggested that methods for data enhancement for ultra-low dose CTAC should focus on reducing bias and root mean square error (RSME) in CT images.

CT images acquired at lower doses (higher noise) may benefit from methods to reduce image errors, such as sinogram smoothing [2; 3], iterative reconstruction, or compressed sensing [4]. CT imaging for PET-only applications demands substantially less image quality requirements than current ‘diagnostic‘ CT [5; 6]. This fact can help govern the selection of appropriate CT acquisition and processing techniques.

The goal of this work is to find the x-ray spectrum (kVp and filtration scheme) that minimizes radiation dose, while maintaining AC-PET image quality. A constraint is to use small or no modifications of current clinical x-ray components ultra-low dose PET/CT imaging, In other words the CT component would have to be capable of still producing diagnostic-quality CT images as part of a standard PET/CT scanner.

II. METHODS

We evaluate the impact of different x-ray spectra on CT radiation dose and the reconstructed AC-PET image. The x-ray spectra are simulated with intrinsic tube spectra from current commercial CT systems and modified by varying filtration materials and thicknesses.

A. Measurement Simulation Model

We used Catsim [7] for the simulation of the CT images and estimation of the CT dose. The PET images were simulated with a simplified version of ASIM [8], where the output of Catsim was used as the attenuation correction data for PET simulations as illustrated in Figure 2. Both Catsim and ASIM are analytic simulation methods, which calculate

Manuscript received November 20, 2009. Revised May 20, 2010. This work was supported by NIH Grant R01-CA115870.

Authors are with the Departments of Bioengineering and Radiology at the University of Washington, Seattle, WA. Contact: kinahan@uw.edu.

noiseless line-integrals followed by appropriate addition of statistical noise. This approach allows for rapid generation of multiple i.i.d. realizations of sinogram data, but, unlike photon tracking simulations, are challenged in modeling more complex physics.

CT raw data y_i are formed from the combination of a photon counting process p_i and electronic noise, modeled as a zero-mean Gaussian random process, using following [3; 9; 10]:

$$p_i = \left[\sum_k E_k \cdot Poisson\left(DQE \cdot A_{ik} \cdot \sum_s \frac{1}{S} \exp(-\sum_o l_{iso}\mu_{ok}) + y_{ik}^{scatter} \right) \right] f_{CONV}$$

$$y_i = p_i + Normal(\sigma_{electronic})$$

The electronic noise was fixed for all simulations so that it was approximately 0.035% of the total noise for a CT technique of 50 mAs, which is on the order of 160 photons. Higher and lower electronic noise levels were also simulated but are not reported here. The CT simulation models the effects of the x-ray source focal spot, tube spectra, beam conditioning, detector noise, bow-tie filter, scatter, and beam hardening. The PET simulation includes the effect of attenuation and photon counting. All components were modeled to match the geometry and tube properties of the GE Discovery STE PET/CT scanner.

B. *Spectra Variations*

Different x-ray spectra were generated starting with the x-ray tube spectra from the manufacturer. The x-rays travel through the standard manufacturer's bowtie filter, then through varying flat filters. The post-filtration spectra is evaluated in terms of mean energy and transmission efficiency, which for this work we define as the sum of photons in spectra divided by sum of photons in spectra with no flat filter. Interactions in filters are evaluated in 1keV increments across all energies.

The CT spectra were varied with the following parameters: Tube voltages 80, 100, 120, and 140 kVp, plat filtration material Al (Z=13), Cu (Z=29), Sn (Z=50), Mo (Z=42), and filtration thickness 0 to 1mm in 0.05mm increments.

C. *Evaluation Methods*

The test object used was a uniform 20 x 30 cm elliptical cylinder containing FDG in water at a typical clinical concentration. All CT evaluations were simulated with rotation time 0.5 s, tube current ranging from 1000 mA down to 0.5 mA, slice thickness 3.125 mm. Data was corrected for beam hardening prior to reconstruction on 128x128 pixels over a 50 cm FOV (to match PET image dimensions). The CT images were converted to linear attenuation coefficients at 511 keV using a modification of the bilinear scaling method [6; 11]. The conversion parameters used for the CT-based attenuation correction (CTAC) are presented in Figure 1 and match those used in the GE DSTE PET/CT scanner. The PET images were reconstructed using filtered back projection on 128x128 pixels over a 50 cm FOV.

With 25 independent data realizations acquired with each x-ray spectra we evaluated the (1) total CT radiation absorbed dose, (2) CT image noise, bias, and RMSE, (3) PET image noise, bias, and RMSE.

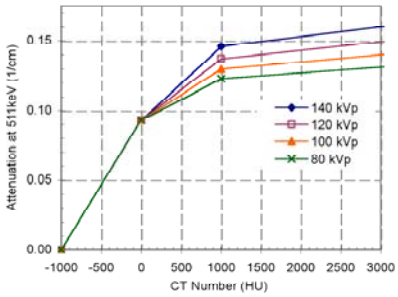


Figure 1. Conversion of CT image values to linear attenuation coefficients for CT-based attenuation correction (CTAC).

III. RESULTS

A. *Spectra Properties*

Figure 2 presents spectra for 120kVp and 140kVp with Cu filtration showing a small subset of the simulated filters and potentials. As expected, increased filtration thickness increases the mean energy and reduces transmission efficiency.

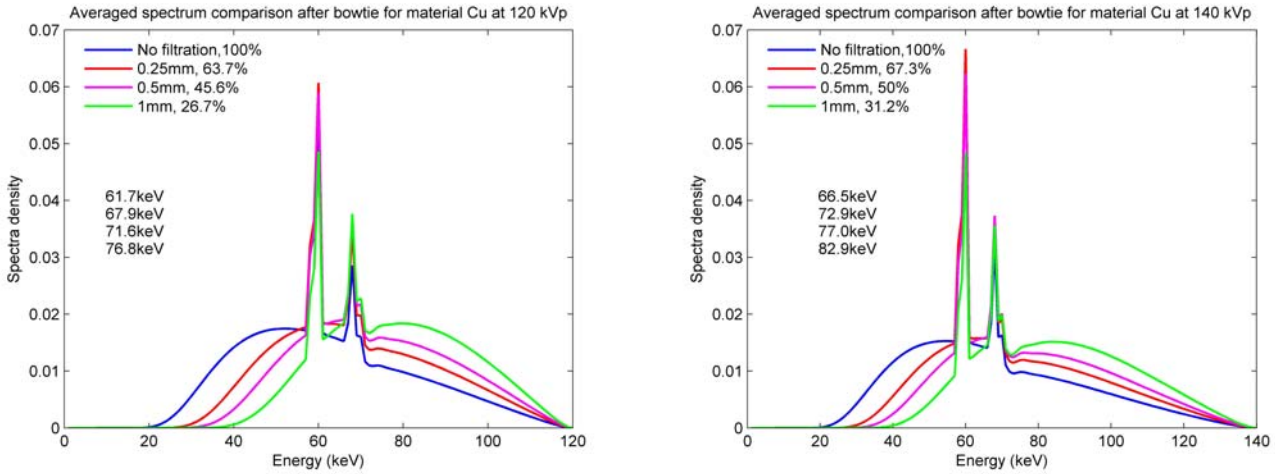


Fig 2. X-ray energy spectra incident on object after bowtie filter and varying beam filtration (0, 0.25, 0.5, 1 mm Cu) for 120 kVp (left) and 140 kVp (right). Legend contains thickness (mm) and transmission efficiency (%) for different filters. Inset text reports the mean energy (keV) of each spectrum from softest to hardest spectra. For each voltage, all spectra were generated with matched tube current.

B. Dose Efficiency

Spatial dose distributions and total dose as a function of kVp behaved as expected. Representative images of the relative dose distribution in the simulated object demonstrate fairly uniform dose distribution with an increased dose to the periphery of object (despite use of bowtie). The transmission efficiency and mean energy of all of the simulated spectra versus total absorbed dose were calculated for all 320 combinations (Figure 3).

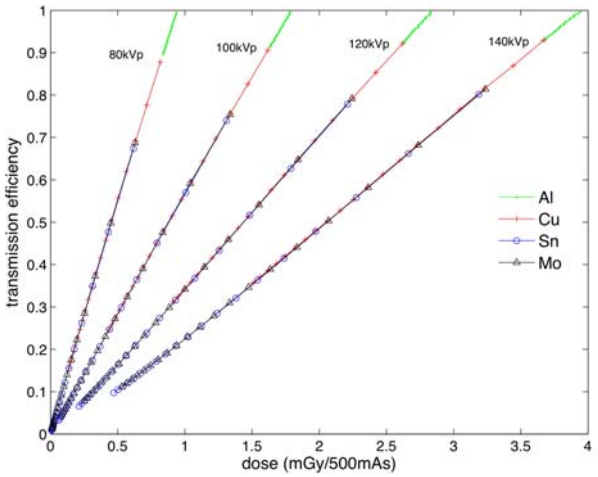


Figure 3. Transmission efficiency (% of surviving photons after filtration relative to no filtration) versus absorbed dose for a fixed tube current. Each of 4 filtration materials is varied from 0.05 to 1mm thickness in 0.05mm increments. Results are presented for four voltage potential.

C. CT and PET Images

Figure 4 presents anecdotal CT images generated with different spectra resulting in matched absorbed dose. The higher energy spectra (140kVp, 1mm Cu) results in a CT image with visually less streak artifacts and noise.

Figure 5 summarizes the CT and PET bias and RMSE for different kVp and Cu filtration versus absorbed dose. Each data point for a given spectra is generated with a different tube current, which linearly changes the absorbed dose. For a fixed absorbed dose, all metrics benefit from having a more narrow and higher energy spectra.

For all spectra, the CT noise increases severalfold as the dose is decreased to below 0.1mGy. This noise does not propagate into the PET images as much as would be expected because the downsampling to PET resolution suppresses noise.

For the lowest energy spectra (80kVp, no filtration), PET bias is below 5% when dose is greater than 0.30mGy. On the other end of the parameters, for the highest/narrowest energy spectra (140kVp, 1mm Cu), PET bias is below 5% when dose is greater than 0.11mGy.

IV. Discussion

We would like to find the x-ray generation scheme that provides the most efficient use of dose for attenuation correction of PET emission data. A more dose efficient scheme would offer improved image quality for a matched absorbed dose or vice versa. These results indicate that more narrow, higher energy spectra are more dose efficient for these metrics where subtle contrast discrimination in the CT image is not necessary and overall image bias and noise are important.

With the choice of x-ray tube determined by clinical CT image criteria, the selection of flat filtration material offers an alternative method to shape optimal spectral properties, such as providing better transmission efficiency with matched absorbed dose. Figure 3 suggests that for the four materials studied, there is no clear advantage to a particular material. That is, transmission efficiency versus dose shows an identical slope for all materials and the only variant is the material thickness. Assuming that finite, tube-current x-ray sources require a transmission efficiency of greater than 0.5, copper offers a good compromise of transmission and dose efficiency for a range of thicknesses [0-0.25mm]. On the other hand, only 0.1mm of the heavier metals of molybdenum and tin are required to reduce the transmission of efficiency of an 80kVp beam to less than 50%.

This simulation study of different x-ray spectra indicates that ultra-low dose CTAC imaging is possible with dose on the order of 0.11mGy, with minimal introduction of PET bias. Prior work [1, 2] has confirmed that additional steps for the reduction CT noise and artifacts offers the potential for even further reduction of dose. Furthermore, this study demonstrates that higher energy, more narrow x-ray spectra are preferred for CT-attenuation correction of PET images.

Acknowledgments

We thank the research staff at GE Global Research Center for help with Catsim and helpful discussions: Drs Jed Pack and Bruno De Mann. We also thank Drs. Patrick La Riviere, Jiang Hsieh, James Colsher, and Bruce Whiting for helpful discussions.

References

[1] T. Xia, A. M. Alessio, and P. E. Kinahan, "Limits of Ultra-Low Dose CT Attenuation Correction for PET/CT," *Nuclear Science Symposium Conference Record, 2009. NSS/MIC, IEEE*, 2009, pp.3074-3079

[2] J. G. Colsher et al., "Ultra low dose CT for attenuation correction in PET/CT," *Nuclear Science Symposium Conference Record, 2008. NSS '08. IEEE*, 2008, pp. 5506-5511.

[3] P. J. La Riviere, "Penalized-likelihood sinogram smoothing for low-dose CT," *Med Phys*, 32, 2005, pp. 1676-1683.

[4] G. H. Chen, J. Tang, and S. Leng, "Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic CT images from highly undersampled projection data sets," *Med Phys*, 35, 2008, pp. 660-663.

[5] P. E. Kinahan, A. M. Alessio, and J. A. Fessler, "Dual energy CT attenuation correction methods for quantitative assessment of response to cancer therapy with PET/CT imaging," *Technol Cancer Res Treat*, 5, 2006, pp. 319-327.

[6] P. E. Kinahan, B. H. Hasegawa, and T. Beyer, "X-ray-based attenuation correction for positron emission tomography/computed tomography scanners," *Semin Nucl Med*, 33, 2003, pp. 166-179.

[7] B. De Man et al., "CatSim: a new computer assisted tomography simulation environment," *6510(1)*, San Diego, CA, USA, 2007, SPIE, pp. 65102G-651028.

[8] C. Comtat et al., "Simulating Whole-body PET scanning with rapid analytical methods," *3*, Seattle, WA, 1999, pp. 1260-1264.

[9] B. De Man et al., "An iterative maximum-likelihood polychromatic algorithm for CT," *IEEE Trans Med Imaging*, 20, 2001, pp. 999-1008.

[10] B. R. Whiting et al., "Properties of preprocessed sinogram data in x-ray computed tomography," *Med Phys*, 33, 2006, pp. 3290-3303.
[11] C. Burger et al., "PET attenuation coefficients from CT images: experimental evaluation of the transformation of CT into PET 511-keV attenuation coefficients," *Eur J Nucl Med Mol Imaging*, 29, 2002, pp. 922-927.

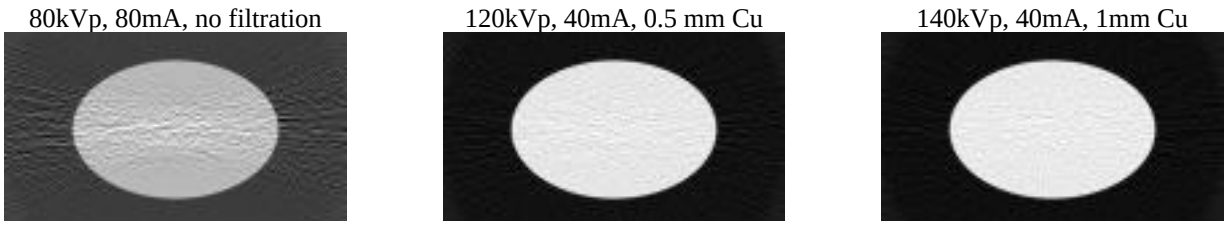


Fig 4. Representative reconstructed CT images of elliptical phantom acquired with different techniques and x-ray spectra. These techniques deliver the same mean absorbed dose to object.

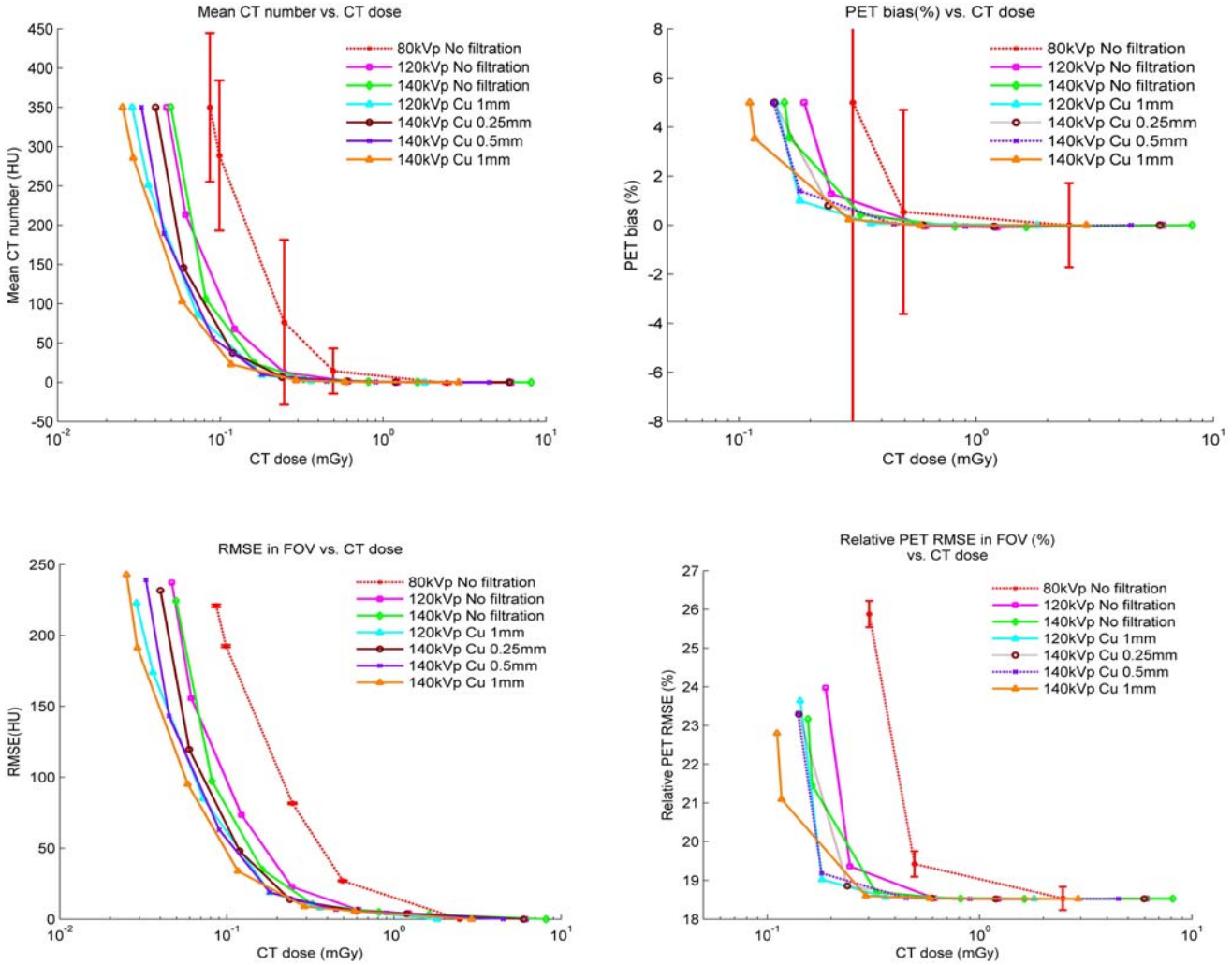


Fig 5. Bias (top), RMSE (bottom) of CT (left) and PET (right) versus radiation absorbed dose from CT. At matched absorbed dose, the lower energy, unfiltered spectra (for example 80kVp no filtration) tend to have more bias, noise and RMSE than the higher energy, filtered spectra

Noise-Reduction for Low-Dose Cone-Beam CT Sinograms

Yi Fan, Hongbin Zhu, Hongbing Lu, Jing Wang, and Zhengrong Liang

***Abstract* — Cone-beam computed tomography (CBCT) can employ a large area X-ray beam in conjunction with a flat panel X-ray detector to scan a volumetric acquisition. It can short examination times dramatically and improve visibility of vascular structures, thus have had a dramatic impact on body imaging in the past decades. Though CBCT has so many advantages over conventional stop-and-shoot CT, the X-ray radiation dose to the body is still a major concern. This work aims to lower the X-ray tube current (mAs) for radiation reduction and proposes a penalized weighted least-square (PWLS) scheme to obtain the noise-reduced low-mAs sinogram with decent image quality. Considering the continuous sampling of CBCT projection as detection system rotates, a Karhunen–Loéve (KL) transform is applied along the rotation direction for neighboring three projections. In the KL domain, each two-dimensional (2D) component is considered by an anisotropic Markov random filed (MRF) Gibbs function over the transverse and axial directions and a penalty is designed for the correlation over these two directions. Gauss-Seidel update strategy is adapted to minimize the objective function. A simulation study was performed to demonstrate the effectiveness of the proposed (KL-PWLS-2D) approach with comparison to a full 3D Gibbs smoothing strategy without considering the continuous sampling correlation (PWLS-3D). The improvement gained by the proposed algorithm was documented by the profiles, contrast-to-ratio (CTR) and noise-resolution tradeoff studies.**

***Index Terms*— Noise reduction, cone-beam computed tomography, low-dose, Karhunen-Loeve transform, penalized weighted least-squares, noise-resolution tradeoff**

I. Introduction

Cone-beam computed tomography (CBCT) has demonstrated a dramatic impact on body imaging in the past decades. Unlike conventional stop-and-shoot CT, CBCT provides continuous volumetric data acquisition while the patient bed moves through the gantry during scanning. The advantages of using CBCT include (1) dramatically shortened examination times, (2) improved visibility of vascular structures, and (3) only selected region of interest (ROI), like breast, is exposed to radiation which results in improved image quality and spars the rest of the body from unnecessary radiation exposure, etc. [1]. With more and more concerns on the potential risk of X-ray radiation exposure to the body, reducing the X-ray radiation becomes one of the major efforts in the CT field. Generally, minimizing X-ray exposure of CBCT to the patients could be achieved by lowering the X-ray tube current (mA) or shorter exposure time. However, the image quality of low mA acquisition protocol will be severely degraded due to excessive X-ray quantum noise [2, 3]. Thus, the reconstruction of low-mAs (or low-dose) CBCT images is essentially a noise problem.

In this study, we present a statistical noise reduction algorithm for low-dose CBCT by applying a Karhunen–Loéve (KL) domain penalized-weighted least-square (PWLS) filtering scheme to reduce the noise of cone-beam sinograms. The KL transform reduces the 3D sinogam to 2D principal components by de-correlating the correlation of continuous sampling around the rotation axis. The PWLS objective function is constructed based on the noise properties of the CBCT sinogram. The weight in the PWLS objective function is estimated in high accuracy using an analytical formula that describes the relationship between data sample mean and variance [4-7]. The penalty in the PWLS objective function is adaptive to the signal-to-noise ratio (SNR) for each 2D component. The objective function in the KL domain is minimized by the Gauss-Seidel update strategy [8-13]. The CBCT reconstructed images could be obtained by using the Feldkamp-Davis-Kress (FDK) algorithm [14] after the sinogram is processed by the proposed algorithm. The effectiveness of the KL-PWLS-2D noise reduction algorithm is shown by a phantom simulation study and evaluated by profile through selected area and noise-resolution tradeoff studies with comparison to a fully 3D statistical smoothing (PWLS-3D) method without using the KL transform.

II. Method

In this study, the KL transform is first applied to account for the correlative information of continuous data sampling along nearby views of the CBCT sinogram. For the *i*-th view of the sinogram data, its nearby views (the (*i-1*)-th and the (*i+1*)-th views) are selected to perform the KL transform. The implementation of the KL process could be described as:

$$\tilde{y} = Ay \tag{1}$$

This work was partially supported by NIH Grant #CA082402 and #CA120917 of the National Cancer Institute.

Yi Fan and Hongbin Zhu are with the Department of Radiology, State University of New York, Stony Brook, NY 11794, USA (e-mail: yifan@mil.sunysb.edu).

Hongbing Lu is with the Department of Computer Application, Fourth Military Medical University, Xi'an, Shaanxi 710032, China.

Jing Wang is with the Division of Medical Physics and Engineering, Department of Radiation Oncology, University of Texas Southwestern Medical Center, TX 75390, USA

Zhengrong Liang is with the Departments of Radiology and Computer Science, State University of New York, Stony Brook, NY 11794 USA

where $\tilde{y}$ and y are $3\times N$ vectors denoting the KL transformed components and the corresponding original sinogram in spatial domain, respectively, N is the product multiplied by the number of detector bins (transverse direction) and slices (axial direction). Each row of y is the vector of one complete dataset obtained at one rotation angle. $\boldsymbol{A}$ is the transform matrix of dimension 3×3 and could be computed directly from the sinogram.

In the KL domain, the PWLS criterion can be used to estimate the corresponding ideal sinogram by minimizing the following objective function [8]:

$$\Phi_l(\tilde{\lambda}_l) = (\tilde{y}_l - \tilde{\lambda}_l)'\tilde{\Sigma}_l^{-1}(\tilde{y}_l - \tilde{\lambda}_l) + (\frac{\alpha}{d_l})R(\tilde{\lambda}_l) \quad (2)$$

Formula (2) consists of two terms. The first term denotes the weighted least-square (WLS) measurement. Notation $\tilde{y}_l$ and $\tilde{\lambda}_l$ are the *l*-th component of the KL transformed noisy and ideal sinogram, respectively. $\tilde{\Sigma}$ is the diagonal variance matrix. Based on our previous analyses [7, 8], the calibrated projection data of low-dose CT follow approximately a Gaussian distribution with an associated relationship between the sample mean and variance which can be described as:

$$\sigma_i^2 = \exp(\lambda_i)N_{i0} \quad (3)$$

where N_{i0} denotes the fixed photon number.

In the second term, d_l is the eigenvalue and α is the parameter which controls the penalty R which takes the quadratic form as used in [8]:

$$R(\tilde{\lambda}_l) = \tilde{\lambda}_l^T \tilde{R}\tilde{\lambda} = \sum_i \sum_{m\in N_i} w_{im}(\tilde{\lambda}_{i,l} - \tilde{\lambda}_{m,l})^2 \quad (4)$$

where index *i* runs over all projection views in the sinogram spaces, N_j represents the first order nearest neighbors of the *i*-th pixel in the KL transformed 2D component. The weight $w_{i,m}$ is determined by the magnitude of difference between neighbors and the concerned. In this study, it is simplified by setting 1 for the four neighbors along vertical and horizontal direction, 0 for others.

The implementation of estimating $\tilde{\lambda}_l$ from $\tilde{y}_l$ could be achieved by an iterative Gauss-Seidel (GS) algorithm [7, 11], as:

$$\tilde{\lambda}_i^{(n+1)} = \frac{\tilde{y}_i + \beta\sigma_i^2\left(\sum_{m\in N_i^1} w_{i,m}\tilde{\lambda}_m^{(n+1)} + \sum_{m\in N_i^2} w_{i,m}\tilde{\lambda}_m^{(n)}\right)}{1+\beta\sigma_i^2\sum_{m\in N_i} w_{i,m}} \quad (5)$$

where *n* represents the iterative number. N_i^1 denotes the upper and left neighbors of $\tilde{\lambda}_i$, and N_i^2 denotes the right and lower neighbors of $\tilde{\lambda}_i$.

The implementation procedure of the proposed KL-PWLS-2D filtering scheme for low-dose CBCT sinograms is summarized as following:

1. For any chosen *i*-th view from the calibrated projection data, the *(i-1)*-th and *(i+1)*-th views of the projection are selected for KL transform.
2. Calculate the KL transform matrix $\boldsymbol{A}$ and apply KL transform on the three neighboring views.
3. In the KL domain, perform PWLS minimization on each KL component.
4. Apply inverse KL transform on the processed KL components to estimate the ideal sinogram for the chosen *i*-th view.
5. Run *i* from the first or beginning view until all the sinogram views are restored.

The CBCT reconstructed images is then obtained by using the Feldkamp-Davis-Kress (FDK) algorithm [14].

III. Results

An air scan was performed to estimate the incident intensity over the field of view (FOV) at a specific mAs level. Figure 1 shows the incident X-ray intensity with the tube current 10 mA and duration of pulse 10 ms. The incident X-ray intensity can then be used for estimation of the sinogram data variance.

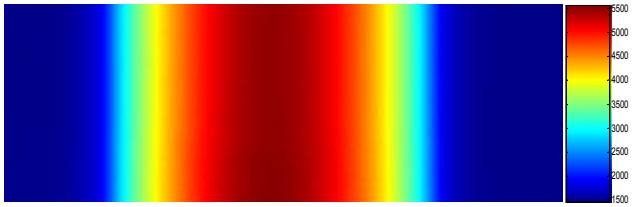


Fig. 1. Incident X-ray intensities across the field of view with 10 mA tube current and 10 ms pulse time.

We tested the proposed algorithm on the CatPhan phantom, see Figure 2. For comparison purposes, the reconstruction of 80 mA tube current is included. For display purpose, one slice was selected containing several strips with different sizes and contrasts. It can be observed that the reconstructed images of 10 mA data from the proposed KL-PWLS-2D show noticeable noise suppression and resolution preservation over the fully 3D PWLS-3D smoothing method without KL transform. The result of KL-PWLS-2D for 10 mA data is comparable to the output from a commercial CT scanner with the 80 mA protocol in terms of the detectability of the strips.

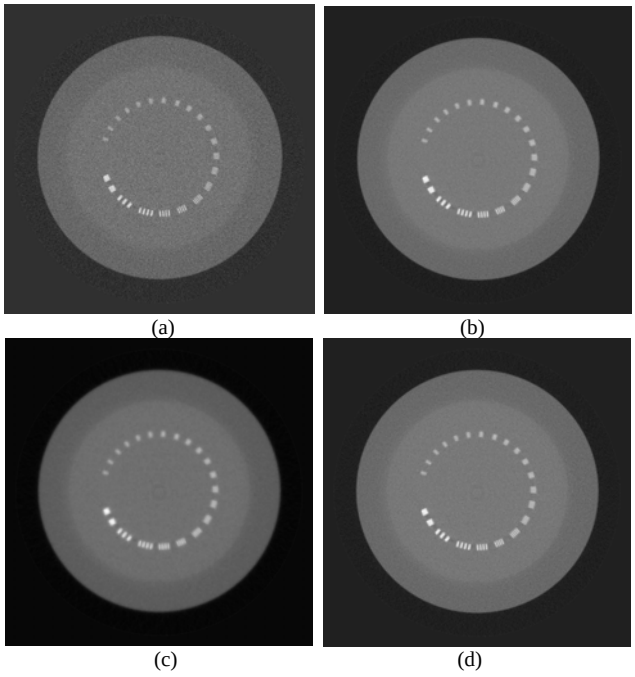


Fig. 2. Reconstructed transverse slice of the 3D CatPhan phantom: (a) from projection images acquired with 10 mA tube current, (b) from projection images acquired with 80 mA tube current, (c) from PWLS smoothed sinogram without KL transform using a 3D MRF penalty, (d). from KL-PWLS-2D filtered sinogram.

To quantitatively evaluate the effectiveness of the proposed KL-PWLS-2D sinogram smoothing algorithm, the noise-resolution tradeoff of the presented approach and the PWLS-3D were computed. The evaluation method used is similar to that used in [11]. A selected transverse slice contain small point-like circles was selected for this study, as shown in Figure 3(a).

The reconstructed image resolution was analyzed by the edge spread function (ESF) along the central vertical profile of the disk pointed by an solid arrow in Figure 3(a). The comparison of profiles through the kernel is presented in Figure 3(b). A Gaussian kernel with standard deviation equal to σ is used to fit the profiles to an error function. Then the image resolution could be reflected by full-width at half-maximum (FWHM) which equal to 2.35σ, see Figure 3(c).

We also calculated the contrast-to-ratio (CNR) at different ROI in the images shown in Figure 3(a). The CNR is defined as:

$$CNR = \frac{|\mu_s - \mu_b|}{\sqrt{\sigma_s^2 + \sigma_b^2}} \qquad (5)$$

where μ_s is the mean value of the signal and μ_b is the mean value of the background. Four selected disks (indicated by the dashed line and hollow arrows) with different intensities were chosen to calculate CNRs. Table 1 lists the CNRs of these four ROIs. It can be observed that the CNR obtained by the proposed KL-PWLS-2D achieves better performance than PWLS-3D and is comparable to that of the image acquired with 80 mA protocol.

Table 1. CNRs of selected ROIs in Figure 3(a)

	ROI1	ROI2	ROI3	ROI4
80 mA	4.81	1.25	7.38	1.77
10 mA	1.77	0.49	2.73	0.62
PWLS-3D	4.25	0.90	5.85	1.36
KL-PWLS-2D	4.33	1.11	6.23	1.54

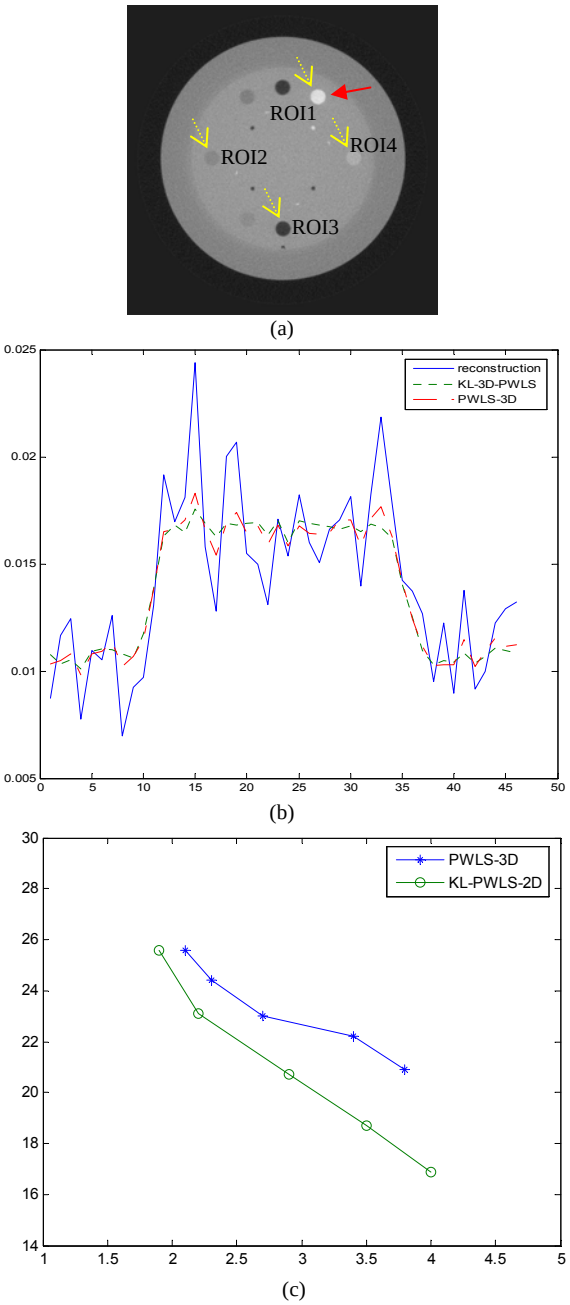


Fig. 3. (a) Phantom used for the evaluation studies. (b). Vertical profiles through the center of pointed disk in (a). (c). Noise-resolution tradeoff curves for the presented approaches and PWLS-3D method.

IV. Conclusion

In this study, we presented a KL domain PWLS scheme to treat noise in low-dose CBCT sinogram. Other than modeling the noise properties of the first and second statistical moments of the low-dose CT sinogram by sample mean and variance,

we obtained the property parameters directly by using commercial CT scanner through in-air scan. The KL transform was first applied along the neighboring views in the 3D sinogram to consider the continuous sampling correlative information among the rotated sampling. Then the PWLS objective function was minimized adaptively for each 2D principal component to estimate the ideal projection at each specified corresponding view. The image reconstruction from smoothed sinogram could be performed by a cone-beam reconstruction algorithm.

Gain has been observed by an experimental phantom study. By the profile through selected ROIs from the reconstructed results of the presented KL-PWLS-2D and the reference PWLS-3D, it is observed that the KL-PWLS-2D has a better performance than the reference. The noise-resolution tradeoff study and CNR of selected ROI further showed the effectiveness of the presented method. The differences should attribute to the KL transform for the consideration of the signal correlation among neighboring views of the CBCT while in the PWLS-3D method these correlations are considered by the penalty terms in the objective function. It is also worthwhile to point out that the computation burden can be greatly decreased by the proposed algorithm with comparison to the 3D penalty modeling approach. It costs only less than 10 minutes to perform the sinogram restoration with size $240\times924\times634$ on a PC with 2.33 GHz and 3 GB memory.

REFERENCES

1. Science News, "Cone-Beam CT Faster, Potentially More Accurate Than Conventional Mammography", *ScienceDaily*, Jan.1, 2008.
2. K. Jung, K. Lee, S. Kim, T. Kim, Y. Pyeun, and J. Lee, "Low-dose, volumetric helical CT: image quality, radiation dose, and usefulness for evaluation of bronchiectasis", *Invest. Radiology*, vol. 35, pp. 557-563, 2000.
3. J. Hsieh, "Nonstationary noise characteristics of the helical scan and its impact on image quality and artifacts", *Medical Physics*, vol. 24, pp. 1375-1384, 1997.
4. H. Lu, I. Hsiao, X. Li, and Z. Liang, "Noise properties of low-dose CT projections and noise treatment by scale transformations", *Conf. Record IEEE NSS-MIC*, in CD-ROM, 2001.
5. H. Lu, X. Li, and Z. Liang, "Analytical noise treatment for low-dose CT projection data by penalized weighted least-square smoothing in the K-L domain", *SPIE Medical Imaging*, vol. 4682, pp. 146-152, 2002.
6. T. Li, J. Wang, J. Wen, X. Li, H. Lu, J. Hsieh, and Z. Liang "SNR-weighted sinogram smoothing with improved noise-resolution properties for low-dose X-ray computed tomography", *SPIE Medical Imaging*, vol. 5370, pp. 2058-2066, 2004.
7. T. Li, X. Li, J. Wang, J. Wen, H. Lu, J. Hsieh, and Z. Liang, "Nonlinear sinogram smoothing for low-dose X-ray CT", *IEEE Trans. Nuclear Science*, vol. 51, pp. 2505-2513, 2004.
8. J. Wang, T. Li, H. Lu, and Z. Liang, "Noise reduction for low-dose single-slice helical CT sinogram", *IEEE Tans. Nuclear Science*, vol. 53, pp. 1230-1237, 2006.
9. J. Wang, H. Lu, D. Eremina, G, Zhang, S. Wang, J, Chen, J. Manzione and Z. Liang, "An experimental study on the noise properties of X-ray CT sinogram data in the Radon space", *SPIE Medical Imaging*, vol. 6913, 2008
10. J. Wang, T. Li, H. Lu, and Z. Liang, "Noise reduction for low-dose helical CT by fully 3D penalized weighted least-square sinogram smoothing", *SPIE Medical Imaging*, vol. 6142, 2006.
11. J. Wang, T. Li, H. Lu, and Z. Liang, "Penalized weighted least-square approach to sinoram noise reduction and image reconstruction for low-dose X-ray computed tomography", *IEEE Trans. Medical Imaging.*, vol. 25, no. 10, pp. 1272-1283, Oct. 2006.
12. K. Sauer and C. Bouman, "A local update strategy for iterative reconstruction form projections," *IEEE Trans. Signal Process.*, vol. 41, no. 2, pp. 534–548, Feb. 1993.
13. P. J. La Rivière and D. M. Billmire, "Reduction of noise-induced streak artifacts in X-ray computed tomography through Spline-based penalized-likelihood sinogram smoothing," *IEEE. Trans. Med. Imag.*, vol. 24, no. 1, pp. 105–111, Jan. 2005.
14. L. Feldkamp,L. Davis and J. Kress, "Practical cone-beam algorithm *J. Opt. Soc. Am.* A **1** 612–9, 1984.

Confidence Intervals for ROC-based Image Quality Assessment in CT

Adam Wunderlich and Frédéric Noo

***Abstract*—We introduce new confidence interval estimators for ROC curves and for ROC summary measures that may be used to characterize observer performance in task-based image quality studies involving binary classification of two classes of images. Our new estimators require continuous-valued observer ratings and rely on two assumptions: (i) the observer ratings are normally distributed for each class of images, and (ii) the variance of the observer ratings is the same for each class of images. These assumptions apply, for example, to ratings produced by linear model observers applied to low-contrast lesion detection tasks. Importantly, the new estimators yield confidence intervals with exact coverage probabilities and they may be easily calculated using a computer. The utility of our new interval estimators is illustrated through an image quality evaluation example using real x-ray computed tomography images.**

I. Introduction

Task-based assessments of image quality often involve binary classification. For example, studies of medical image quality frequently evaluate a lesion detection task in which an observer attempts to discriminate between two classes of images: those images that contain a lesion and those that do not. Observer performance for a binary classification task can be expressed using a receiver operating characteristic (ROC) curve, which plots the true positive fraction (TFP) as a function of false positive fraction (FPF) [1]. For purposes of simplification, the whole ROC curve is often reduced to a single number, called an ROC summary measure [1], [2]. In practice, ROC curves and ROC summary measures must be estimated from observer rating data obtained experimentally. Consequently, they suffer from statistical variability that must be characterized in order to make inferences about observer performance.

One way that estimator variability may be summarized is through the use of confidence intervals. As opposed to point estimates, confidence intervals provide a probabilistic guarantee of covering the parameter of interest [3]. Moreover, as observed in [4], a virtue of confidence intervals is that they convey more information than hypothesis testing (together with p-values) in two ways. First, confidence intervals communicate the amount of statistical precision involved in an experiment. Second, they communicate the relative size of an experimental effect, i.e., they show how significant experimentally-observed differences are in terms of their magnitude. In this work, we present fully parametric estimators that yield confidence intervals for ROC summary measures and confidence bands for ROC curves. Our new estimators are designed for continuous-valued observer ratings under the dual assumptions that (i) the observer ratings are normally distributed for each class of images, and (ii) the variance of the observer ratings is the same for each class of images.

Although the aforementioned assumptions appear to be restrictive, they are generally satisfied for CT image evaluation tasks involving detection of small, low-contrast lesions with linear computerized model observers. Indeed, because most linear model observers compute each observer rating as a linear combination of a large number of image pixel values, the central limit theorem [3] implies that the ratings will tend to be normally-distributed for each class of images. Furthermore, because the absence or presence of a small, low-contrast lesion will have only a small impact on the image covariance matrix, the variance of ratings produced by linear model observers will change little from one class of images to the other. Thus, the variance of the rating data may be assumed to be the same for each class of images.

Previous work that examined confidence intervals with application to ROC analysis has primarily focused on estimation of the area under the ROC curve (AUC), a widely-used summary measure; see, e.g., [5], [6] for overviews of the literature on this topic. Also, confidence bands for the entire ROC curve have been investigated by [7], [8]. A drawback of the previously investigated interval and band estimators is a reliance on either asymptotic approximations or resampling techniques; these approaches result in imprecise knowledge of the coverage probabilities for the resulting confidence intervals. By contrast, the coverage probabilities for the confidence intervals that we propose in this work are known exactly.

After reviewing summary measures for ROC curves, we present our new confidence interval estimators. The remainder of the paper illustrates the utility of these estimators in the context of image quality evaluation with real x-ray CT images.

II. ROC Figures of Merit

We assume that the reader is familiar with the normal and noncentral t probability distributions. If a random variable X is normally distributed with mean μ and variance σ^2, then we write $X \sim \mathcal{N}(\mu, \sigma^2)$. Similarly, if X follows a non-central t distribution with ν degrees of freedom and non-centrality parameter δ, it is written as $X \sim t'_\nu(\delta)$. For the standard normal distribution, $\mathcal{N}(0,1)$, we employ the usual notation $\Phi(x)$ and $\Phi^{-1}(p)$ for the cumulative distribution function (cdf) and the inverse cdf, respectively.

Suppose that an observer (a.k.a. classifier) is tasked with the discrimination of two classes of noisy images, denoted as

This work was partially supported by NIH grants R01 EB007236, R21 EB009168 and by a generous grant from the Ben B. and Iris M. Margolis Foundation. Its contents are solely the responsibility of the authors.

The authors are with the Utah Center for Advanced Imaging Research, Department of Radiology, University of Utah, Salt Lake City, UT, USA. (email: awunder@ucair.med.utah.edu, noo@ucair.med.utah.edu)

class 1 and class 2. In medical applications, these two classes may correspond to images of non-diseased and diseased tissues, respectively. A simple model is usually used for the observer's decision process. Namely, for each image, suppose that the observer generates a continuous-valued rating statistic, y, with larger values of y indicating a preference for class 2 and smaller values of y indicating a preference for class 1. To make its decision, the observer compares the rating statistic, y, to a threshold, c. If $y > c$, then the observer concludes that the image belongs to class 2, otherwise, the observer decides that the image belongs to class 1 [1], [2].

We assume that $y \sim \mathcal{N}(\mu_1, \sigma^2)$ for images from class 1 and $y \sim \mathcal{N}(\mu_2, \sigma^2)$ for images from class 2, with $\mu_2 > \mu_1$. In this setting, we investigate the figures of merit listed below; see [1] and [2] for more details on these metrics.

- observer signal-to-noise ratio (SNR):

$$\text{SNR} = \frac{\mu_2 - \mu_1}{\sigma} \tag{1}$$

- area under the ROC curve (AUC):

$$\text{AUC} = \Phi(\text{SNR}/\sqrt{2}\,) \tag{2}$$

- true positive fraction (TPF):

$$\text{TPF}(\text{FPF}) = \Phi(\text{SNR} + \Phi^{-1}(\text{FPF})) \tag{3}$$

  where the false positive fraction (FPF) is a fixed value.

- partial area under the ROC curve (pAUC):

$$\text{pAUC}(\text{FPF}_0, \text{FPF}_1) = \int_{\text{FPF}_0}^{\text{FPF}_1} \text{TPF}(\text{FPF})\; d(\text{FPF}) \tag{4}$$

  where the limits FPF_0 and FPF_1 are fixed.

III. Construction of Interval Estimators

Suppose that an observer rates n_1 images from class 1 and n_2 images from class 2. Denote these ratings for classes 1 and 2 as $y_1^{(1)}, y_2^{(1)}, \ldots, y_{n_1}^{(1)}$ and $y_1^{(2)}, y_2^{(2)}, \ldots, y_{n_2}^{(2)}$, respectively. We wish to estimate confidence intervals for summary measures of observer performance from this finite sample of rating data. In this section, we introduce our estimators assuming that $y_i^{(1)} \sim \mathcal{N}(\mu_1, \sigma^2)$ and $y_j^{(2)} \sim \mathcal{N}(\mu_2, \sigma^2)$, for $i = 1, 2, \ldots, n_1$, $j = 1, 2, \ldots, n_2$, where μ_1, μ_2, and σ^2 are unknown. Each of our interval estimators is based on a point estimator for SNR, which is introduced first. For the sake of brevity, we omit all proofs.

A. Point Estimation of SNR

Let the sample mean and the sample variance for class j be $\overline{y}_j = (1/n_j)\sum_{i=1}^{n_j} y_i^{(j)}$ and $s_j^2 = [1/(n_j - 1)]\sum_{i=1}^{n_j}(y_i^{(j)} - \overline{y}_j)^2$, respectively. Also, define a pooled estimator for σ^2 as $s^2 = [1/(n_1+n_2-2)]\left((n_1-1)s_1^2 + (n_2-1)s_2^2\right)$. With these definitions in place, we define an estimator of SNR as

$$\widehat{\text{SNR}} = \gamma\,(\overline{y}_2 - \overline{y}_1)/s, \tag{5}$$

with

$$\gamma = \frac{\sqrt{2\pi/(n_1+n_2-2)}}{B((n_1+n_2-3)/2, 1/2)}, \tag{6}$$

where $B(x, y)$ is the Euler beta function. The multiplicative factor γ is chosen to make the estimator unbiased. The following theorem characterizes the sampling distribution of $\widehat{\text{SNR}}$.

Theorem 1. *Suppose that $\widehat{\text{SNR}}$ is computed from independent samples $y_i^{(1)} \sim \mathcal{N}(\mu_1, \sigma^2)$ and $y_j^{(2)} \sim \mathcal{N}(\mu_2, \sigma^2)$, where $i = 1, 2, \ldots, n_1$ and $j = 1, 2, \ldots, n_2$. Then*

$$\eta\,\widehat{\text{SNR}} \sim t'_\nu(\delta),$$

where

$$\eta = \frac{1}{\gamma}\sqrt{\frac{n_1 n_2}{n_1 + n_2}}\,, \qquad \nu = n_1 + n_2 - 2,$$

and

$$\delta = \text{SNR}\sqrt{\frac{(n_1 n_2)}{(n_1 + n_2)}}\,.$$

The expression for the mean of a noncentral t deviate [9] together with the above theorem may be used to show that $\widehat{\text{SNR}}$ is unbiased.

It turns out that $\widehat{\text{SNR}}$ is the uniformly minimum variance unbiased (UMVU) estimator of SNR. This result is stated precisely in the next theorem.

Theorem 2. *Suppose that $\widehat{\text{SNR}}$ is computed from independent samples $y_i^{(1)} \sim \mathcal{N}(\mu_1, \sigma^2)$ and $y_j^{(2)} \sim \mathcal{N}(\mu_2, \sigma^2)$, where $i = 1, 2, \ldots, n_1$ and $j = 1, 2, \ldots, n_2$. Then $\widehat{\text{SNR}}$ is the unique* UMVU *estimator of* SNR.

B. Confidence Interval Estimation

Given a random variable, X, with a distribution depending on a parameter, θ, one may define a random interval estimate $[\theta_L(X),\ \theta_U(X)]$ for θ. This interval is said to be a $1 - \alpha$ confidence interval for θ if $P(\theta \in [\theta_L(X),\ \theta_U(X)]) = 1 - \alpha$ for all values of θ [10].

We find confidence intervals for the ROC figures of merit summarized in section II using the next two lemmas. The first lemma enables calculation of confidence intervals for the noncentrality parameter of a noncentral t distribution. It may be proved using [3, Theorem 9.2.12, p. 432] together with a property of the noncentral t cdf.

Lemma 1. *Suppose that $T \sim t'_\nu(\delta)$ and that the cdf for T is $F_T(t\,;\,\nu, \delta)$. Also, let α_1 and α_2 be fixed, nonnegative values that satisfy $\alpha_1 + \alpha_2 = \alpha$ with $0 < \alpha < 1$. Then for each $t \in (-\infty, \infty)$, there exist unique values $\delta_L(t)$ and $\delta_U(t)$ satisfying*

$$F_T(t\,;\,\nu, \delta_U(t)) = \alpha_1 \quad \textit{and} \quad F_T(t\,;\,\nu, \delta_L(t)) = 1 - \alpha_2.$$

that are the lower and upper bounds, respectively, of a $1 - \alpha$ confidence interval for δ.

Above, if $\alpha_1 = 0$, then $\delta_U(t) = \infty$ and if $\alpha_2 = 0$, then $\delta_L(t) = -\infty$. In either of these cases, the confidence interval defined in Lemma 1 is said to be one-sided. Otherwise, the interval is said to be two-sided [10]. Although the noncentral t cdf has a relatively complicated analytical form, the equations in Lemma 1 may be solved iteratively for $\delta_L(t)$ and $\delta_U(t)$.

The second lemma is a well-known property of confidence intervals that, as observed in [4], is rarely formalized.

Lemma 2. *Let $g(\theta)$ be a continuous, strictly increasing function of θ. If $[\theta_L,\ \theta_U]$ is a $1-\alpha$ confidence interval for θ, then $[g(\theta_L),\ g(\theta_U)]$ is a $1-\alpha$ confidence interval for $g(\theta)$.*

Since, by Theorem 1, $\widehat{\text{SNR}}$ is a strictly increasing transformation of a noncentral t random variable with a noncentrality parameter that is a constant multiple of SNR, Lemmas 1 and 2 suggest a method to calculate confidence intervals for SNR. Furthermore, the formulae in section II indicate that AUC, TPF, and pAUC are strictly increasing transformations of SNR. Hence, Lemma 2 may be used together with these formulae to compute confidence intervals for AUC, TPF, and pAUC from any confidence interval for SNR. To summarize, exact confidence intervals for the ROC figures of merit in section II, may be calculated by performing the following steps:

(1) Compute a value of $\widehat{\text{SNR}}$ from the observer ratings.
(2) Multiply the value of $\widehat{\text{SNR}}$ from step (1) by η as defined in Theorem 1 to get a noncentral t deviate.
(3) Apply Lemma 1 to compute a $1-\alpha$ confidence interval for the noncentrality parameter, δ.
(4) Use the formula relating δ to SNR in Theorem 1 together with Lemma 2 to transform the confidence interval for δ into a confidence interval for SNR.
(5) Compute confidence intervals for AUC, TPF, and pAUC from the confidence interval for SNR by using Lemma 2 and the formulae in section II.

A confidence band for the entire ROC curve may be found from a confidence interval for SNR in the sense of the next theorem. Below, we use the notation $\text{TPF}(\text{FPF}\,;\,\text{SNR})$ to emphasize that the TPF depends on SNR in addition to FPF. We denote the collection of points on the ROC curve as $\Omega_{\text{ROC}} = \big\{(\text{FPF}, \text{TPF})\ :\ \text{FPF} \in [0,1]\big\}$.

Theorem 3. *Suppose that $y \sim \mathcal{N}(\mu_1, \sigma^2)$ for images from class 1 and $y \sim \mathcal{N}(\mu_2, \sigma^2)$ for images from class 2. Let $[\text{SNR}_L,\ \text{SNR}_U]$ be a $1-\alpha$ confidence interval for* SNR*, and define the set*

$$\widehat{\Omega}_{\text{ROC}} = \big\{(\text{FPF}, T)\ :\ \text{FPF} \in [0,1] \text{ and } T \in \mathcal{I}\big\}$$

where

$$\mathcal{I} = [\text{TPF}(\text{FPF}\,;\,\text{SNR}_L),\ \text{TPF}(\text{FPF}\,;\,\text{SNR}_U)].$$

Then $\widehat{\Omega}_{\text{ROC}}$ is a $1-\alpha$ confidence band for the ROC curve in the sense that, for any value of SNR*, Ω_{ROC} is contained in $\widehat{\Omega}_{\text{ROC}}$ with probability $1-\alpha$, i.e., $P(\Omega_{\text{ROC}} \subset \widehat{\Omega}_{\text{ROC}}) = 1-\alpha$.*

Observe that the $1-\alpha$ confidence band defined in the above theorem is equivalent to the union over all FPF values of $1-\alpha$ confidence intervals for TPF.

IV. Application to Task-Based Image Quality Evaluation

Now, we present an example of how our estimators can be used for task-based image quality evaluation in CT.

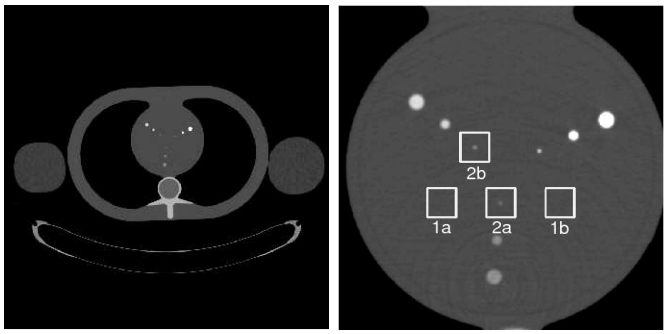


Fig. 1. Mean images of the QRM phantom displayed with a grayscale window of $[-200, 600]$ HU. Whole phantom (left) and reconstruction focused on the heart insert (right) with regions of interest marked with white boxes. ROI-1a and ROI-1b contain no lesion. ROI-2a and ROI-2b contain a low-contrast, and a high-contrast lesion, respectively.

A. Data Acquisition, Image Reconstruction, Task, and Observer

A Siemens *SOMATOM Sensation* 64 CT scanner was used to repeatedly scan a thorax phantom 186 times over a circular source trajectory to collect fan-beam data sets. The x-ray tube settings were 25 mAs and 120 kVp, and the data acquisition was performed with no tube current modulation to accentuate noise correlation in the image. The phantom consisted of a torso constructed by QRM (Möhrendorf, Germany) [11] together with two different water bottles attached to the sides to simulate arms. A mean image of the whole phantom estimated from 186 reconstructions is shown in Figure 1 (left).

The CT data was reconstructed with our implementation of the classical filtered backprojection (FBP) algorithm for direct reconstruction from either short-scan or full-scan fan-beam data [12]. The reconstruction was performed on a grid of 550×550 square pixels of size 0.02 cm $\times\, 0.02$ cm that was centered on the heart insert, as shown in Figure 1 (right).

For the image evaluation task, we considered a signal-known-exactly/background-known-exactly (SKE/BKE) lesion detection task [1]. To implement this task in various situations, we identified four regions of interest (ROI) in the heart insert: two regions without lesions, labeled as ROI-1a and ROI-1b, and two regions with a lesion at their center, labeled as ROI-2a and ROI-2b. The lesion diameter is 1 mm for both ROI-2a and ROI-2b, but the contrast varies from one ROI to the other, namely, the contrast is 210 HU for ROI-2a and 455 HU for ROI-2b. In each case, the background value is 40 HU. Each ROI when viewed as an image consists of 50×50 pixels.

We used the ROIs to investigate two tasks: (i) the task of detecting the 210-HU lesion in ROI-2a, and (ii) the task of detecting the 455-HU lesion in ROI-2b. In either case, the images with the lesion are referred to as class-2 images, and the images without a lesion as class-1 images. Thus, images of ROI-2a are the class-2 images for the first task, and images of ROI-2b are the class-2 images for the second task. On the other hand, images of ROI 1-a and 1-b define class-1 images for both tasks. For each task, any given CT data set yields one image of class-2, and also two images of class-1 that may or

may not be both used for the analysis.

Ideally, the class-1 images would cover the same spatial region as the class-2 images. Here, the phantom does not offer such a flexibility. To circumvent the issue, we rely on knowledge that the noise properties vary slowly in the reconstruction over the spatial scale of interest. That is, we assume that the noise properties are the same in all four ROIs. Furthermore, we assume that the images of the four ROIs obtained from one given data set are independent. This is justified by a previous study which indicated that correlations between image pixels are negligible over the distance (1 to 2 cm) that separates the ROIs [12].

For the observer, we used a trained channelized Hotelling observer (CHO), which is a popular type of linear model observer; for details, see [1]. Our CHO was implemented with three difference of Gaussian (DOG) channels using the parameters given in [13] for sparse DOG channels.

Below, we present results for two case studies. The first study considers the problem of evaluating the ROC parameters for detection of the low-contrast lesion with a short-scan reconstruction. For this study, the CT data sets were split into two groups of 93 elements each. The first group was used for training the observer, i.e., computing the template of the observer, while the second group was used for testing the observer, i.e., evaluating the ROC parameters of the observer. The training used images of ROI-1a for class 1 and of ROI-2a for class 2; the images of ROI-1b were discarded. The testing was performed under two scenarios: (i) using only ROI-1a for class 1, and (ii) using both ROI-1a and ROI-1b for class 1.

The second case study compares the full-scan and short-scan reconstruction strategies in terms of how they impact the detection of each lesion. For this study, the 186 CT data sets were split into two groups of 93 elements each, with the first group being used for full-scan reconstruction, and the second group for short-scan reconstruction. Such a separation ensured statistical independence between the short-scan and the full-scan ROC results. Each group of 93 CT data sets was further split into 46 elements used for training and 47 elements used for testing. As in the first case study, the training only used images of ROI-1a for class 1. However, the testing was performed under a single scenario, namely using both ROI-1a and ROI-1b for class 1.

B. First Case Study

Table 1 contains the 95% confidence intervals that resulted for SNR, AUC, and pAUC under each scenario. The intervals correspond to scenario 1: $n_1 = 93$, $n_2 = 93$, and scenario 2: $n_1 = 186$, $n_2 = 93$ with $\alpha_1 = \alpha_2 = 0.025$, $\text{FPF}_0 = 0$, and $\text{FPF}_1 = 0.2$. In addition, Figure 2 displays estimated 95% confidence bands for the entire ROC curve.

Examining both Table 1 and Figure 2, we see that a clear decrease in the size of each confidence interval and in the ROC confidence band was gained by increasing the number of class 1 images.

C. Second Case Study

Tables 2 and 3 give the 95% confidence intervals that resulted for SNR, AUC, and pAUC for the low-contrast (210 HU) and the high-contrast (455 HU) lesions, respectively. The intervals were estimated from $n_1 = 94$ class-1 ratings and $n_2 = 47$ class-2 ratings with $\alpha_1 = \alpha_2 = 0.025$, $\text{FPF}_0 = 0$, and $\text{FPF}_1 = 0.2$. In addition, the estimated 95% confidence bands for the entire ROC curve under each lesion-contrast scenario are displayed in Figure 3.

To compare the short-scan and full-scan results against each other for a given lesion, recall that the study was designed so that there is no statistical dependence between the 95% confidence interval (band) estimates obtained for the two reconstruction strategies. This feature greatly simplifies

	Scenario 1	Scenario 2
SNR	[1.0625, 1.7031]	[1.1213, 1.6707]
AUC	[0.7738, 0.8858]	[0.7861, 0.8813]
pAUC(0, 0.2)	[0.0770, 0.1229]	[0.0811, 0.1206]

TABLE I
95% CONFIDENCE INTERVALS IN THE FIRST CASE STUDY

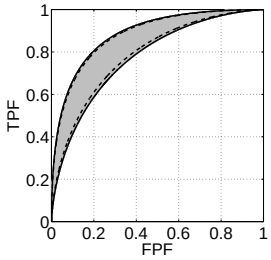


Fig. 2. Comparison of 95% confidence bands for an ROC curve estimated with different numbers of class 1 ratings in the first case study. The bands correspond to scenario 1: $n_1 = 93$, $n_2 = 93$ (solid lines) and scenario 2: $n_1 = 186$, $n_2 = 93$ (dashed lines).

	Full-scan	Short-scan
SNR	[1.6433, 2.4962]	[1.0008, 1.7734]
AUC	[0.8774, 0.9612]	[0.7604, 0.8951]
pAUC(0, 0.2)	[0.1187, 0.1682]	[0.0726, 0.1277]

TABLE II
95% CONFIDENCE INTERVALS ESTIMATED FOR THE LOW-CONTRAST LESION IN THE SECOND CASE STUDY

	Full-scan	Short-scan
SNR	[3.0605, 4.1607]	[1.9573, 2.8578]
AUC	[0.9848, 0.9984]	[0.9168, 0.9783]
pAUC(0, 0.2)	[0.1863, 0.1984]	[0.1397, 0.1811]

TABLE III
95% CONFIDENCE INTERVALS ESTIMATED FOR THE HIGH-CONTRAST LESION IN THE SECOND CASE STUDY

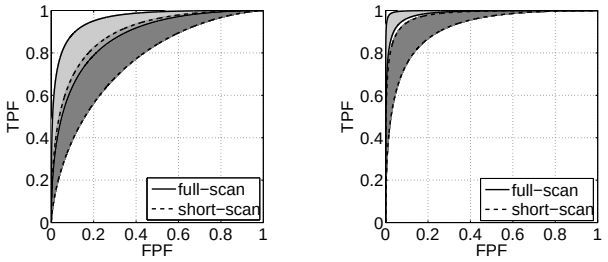


Fig. 3. 95% confidence bands for the ROC curves corresponding to full-scan and short-scan reconstructions of low-contrast (left) and high-contrast (right) lesions

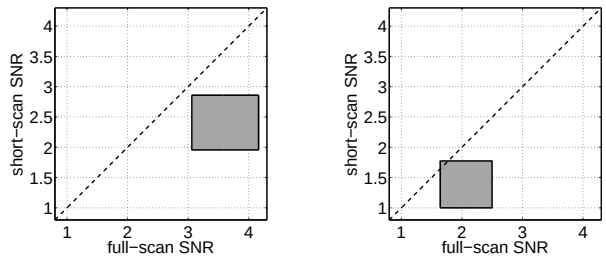


Fig. 4. 90.25% confidence regions for $(\mathrm{SNR}_{fs}, \mathrm{SNR}_{ss})$ in the high-contrast (left) and low-contrast (right) lesion cases

the analysis as there is no need to worry about correlations between these estimates. Suppose, for example, that we wish to compare the SNR values obtained for the two reconstruction strategies. Let SNR_{fs} and SNR_{ss} be these two values, and let $[L_{fs}, U_{fs}]$ and $[L_{ss}, U_{ss}]$ be their 95% confidence intervals. Then, by the rules governing probabilities of independent events, we can claim a $0.95^2 = 0.9025$ confidence that the region $[L_{fs}, U_{fs}] \times [L_{ss}, U_{ss}]$ covers the pair $(\mathrm{SNR}_{fs}, \mathrm{SNR}_{ss})$; see Figure 4. Both the size and the position of the confidence region determine how the results should be interpreted. First, a smaller region covering a given pair of SNR values indicates a higher statistical precision. Second, if the confidence region does not intersect the line at 45 degrees in the plane of possible values for $(\mathrm{SNR}_{fs}, \mathrm{SNR}_{ss})$, then there is evidence that the reconstruction strategies yield dissimilar detection performance. Consequently, different conclusions are to be drawn between the low-contrast and the high-contrast lesion cases. For the high-contrast lesion, because the SNR confidence intervals do not overlap, the SNR confidence region does not intersect the 45 degree line as shown in Figure 4 (left), and we can reject the hypothesis that $\mathrm{SNR}_{fs} = \mathrm{SNR}_{ss}$ with a confidence level of 0.9025. Likewise, the same conclusion can be made for the other figures of merit. Conversely, for the low-contrast lesion, as shown in Figure 4 (right), the SNR confidence region intersects the 45 degree line, and we can only state that there is not enough evidence at the 0.9025 level to reject the hypothesis that $\mathrm{SNR}_{fs} = \mathrm{SNR}_{ss}$. In either case above, it should be observed that the conclusions were drawn with a fairly poor statistical precision, due to the small number of images used to test the observer.

V. Discussion and Conclusions

In this work, we proposed confidence interval estimators that may be used in ROC evaluations of task-based image quality studies. A strength of the new interval estimators is that they have exactly-known coverage probabilities. Our approach relies on two assumptions: (i) the observer ratings are normally distributed for each class of images, and (ii) the variance of the observer ratings is the same for each class of images. These assumptions apply, for example, to ratings produced by linear model observers applied to lesion detection tasks involving small, low-contrast lesions.

We demonstrated the use of the new confidence interval estimators with an example involving a trained CHO and a SKE/BKE lesion detection task with real x-ray CT images. In the context of this example, we have seen that increasing n_1 relative to n_2, which is relatively easy to do with CT images, can be used to reduce statistical variability. It should be emphasized that our choices for the task and for the observer are not necessarily optimal for image quality assessment in CT. Investigation of more sophisticated tasks and observers suitable for CT images is an important topic for future research.

Much previous work dealing with ROC estimation has focused on estimation of the so-called binormal parameters, a and b, which may be used to parameterize ROC curves; see, e.g., [2] for an overview. Our new confidence interval estimators may be understood in this context. Specifically, under our assumptions for the observer ratings, $a = \mathrm{SNR}$ and $b = 1$. Thus, the results in this paper correspond to a setting in which the ROC curve is parameterized by only SNR.

References

[1] H. H. Barrett and K. J. Myers, *Foundations of Image Science*. Wiley, 2004.
[2] M. S. Pepe, *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford Univ. Press, 2003.
[3] G. Casella and R. L. Berger, *Statistical Inference*, 2nd ed. Duxbury, 2001.
[4] J. H. Steiger and R. T. Fouladi, "Noncentrality interval estimation and the evaluation of statistical models," in *What if There Were No Significance Tests?*, L. L. Harlow, S. A. Mulaik, and J. H. Steiger, Eds. Lawrence Erlbaum, 1997.
[5] D. Bamber, "The area above the ordinal dominance graph and the area below the receiver operating characterstic graph," *J. Math. Psych.*, vol. 12, pp. 387–415, 1975.
[6] R. G. Newcombe, "Confidence intervals for an effect size measure based on the Mann-Whitney statistic. Part 2: Asymptotic methods and evaluation," *Stat. Med.*, vol. 25, pp. 559–573, 2006.
[7] G. Ma and W. Hall, "Confidence bands for receiver operating characteristic curves," *Med. Decis. Making*, vol. 13, no. 3, pp. 191–197, July-Sept. 1993.
[8] S. A. Macskassy, F. Provost, and S. Rosset, "ROC confidence bands: An empirical evaluation," in *Proc. 22nd Internl. Conf. on Machine Learning*, Bonn, Germany, 2005, pp. 537–544.
[9] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, 2nd ed. Wiley, 1995, vol. 2.
[10] L. J. Bain and M. Engelhardt, *Introduction to Probability and Mathematical Statistics*, 2nd ed. Duxbury, 1992.
[11] QRM GmbH, http://www.qrm.de/.
[12] A. Wunderlich and F. Noo, "Image covariance and lesion detectability in direct fan-beam x-ray computed tomography," *Phys. Med. Biol.*, vol. 53, no. 10, pp. 2471–2493, May 2008.
[13] C. K. Abbey and H. H. Barrett, "Human- and model-observer performance in ramp-spectrum noise: Effects of regularization and object variability," *J. Opt. Soc. Amer. A*, vol. 18, no. 3, pp. 473–488, March 2001.

Plenary Talk

Tuesday, June 8, 2:00pm - 2:45pm

Design Differences between a Standard CT-Scanner and a C-arm Based Cone-beam CT-Scanner

Michael D. Silver, Ph.D.
Vice-President Research
Toshiba Medical Research Institute USA, Inc., Vernon Hills, Illinois,USA

In my talk I'll compare both the more obvious differences between standard CT-scanners and C-arm x-ray imaging systems with a CT-option and differences that might not be widely appreciated. Let's start with the definitions of the two different systems. What is a "standard CT-scanner?" I could use the term MSCT for multi-slice CT but we now have "standard" systems with nearly as many detector rows and as wide a cone-angle as the C-arm based cone-beam CT-scanners, which are often designated by CBCT. This is becoming a distinction without descriptive value. However, for ease of a short descriptor, I'll use MSCT for "standard" CT-scanners and CBCT for C-arm based imagers.

What is a "standard" CT-scanner? The IEC in the CT-safety standard IEC60601-44-ed3 limits its coverage to "CT scanners intended to be used for both head and body characterized by an enclosure of the X-ray source(s) and imaging detector(s) in a common protective cover in the shape of a toroid." Thus the distinction with CBCT is CBCT's lack of "a common protective cover in the shape of a toroid." This obvious mechanical difference leads to more than a factor of ten difference in rotation speed. This and resolution requirements leads to different choices for the detector and that has subsequent consequences for data corrections and image quality. A more subtle difference related to the mechanical design is the choice in backprojection formulae.

Another way to characterize the two systems is by clinical application. We usually say that the standard CT-scanner is a diagnostic device. CBCT is not a diagnostic device; rather, it is mostly used for guidance during intervention. We'll list the differences in clinical applications and discuss support for advanced applications.

In the commercial environment the two imaging systems have very different cost targets, thus affecting design choices. It is likely that the developments of the two systems are under two different engineering departments. They may have availability to different levels of funding, engineering resources, and design restrictions pertaining to legacy software and hardware in previous and current generations of these imaging systems.

MSCT is explicitly designed for CT while the interventional scanner is designed for fluoroscopy and radiology with CT as an add-on option. How does doing CT on-the-cheap stack up with conventional CT? Or do their different applications make this a non-issue?

HOW TO DEFINE THE NEXT GENERATION CARDIAC CT ARCHITECTURE?

Hengyong Yu[1], Bruno De Man[2] and Ge Wang[1]

1. SBES Division and ICTAS Center for Biomedical Imaging, VT-WFU School of Biomedical Engineering and Sciences, Virginia Tech, Blacksburg, VA, 24060, USA
2. CT and X-ray Systems and Applications Lab, GE Global Research, Niskayuna, NY 12309, USA
hengyong-yu@ieee.org, bruno.deman@research.ge.com, ge-wang@ieee.org

ABSTRACT

Cardiovascular diseases are pervasive with high mortality and morbidity at tremendous social and healthcare costs. There are urgent needs for significantly higher fidelity cardiac CT with substantially lower radiation dose, which is currently not possible because of technical limitations. Although cardiac CT technology has improved significantly from 16 to 320 detector rows and from single to dual source, there remain technical challenges in terms of temporal resolution, spatial resolution, radiation dose, and so on, which motivates us to advance the state-of-the-art in cardiac CT dramatically and define the next generation cardiac CT system. Here we propose an evaluation scheme to single out the best candidate from several novel cardiac CT architectures with novel sources and scanning trajectories.

Index Terms— Cardiac CT, Evaluation scheme, Architecture innovation, Algorithm development, Performance evaluation.

1. INTRODUCTION

Worldwide there are growing concerns on radiation induced genetic, cancerous and other diseases [1]. Cardiac CT is considered as a radiation-intensive procedure, yet becoming more and more common. Despite the impressive advancement in CT technology, there are critical and immediate needs for cardiac CT with significantly better image quality at lower radiation dose. This motivates us to advance the state-of-the-art in cardiac CT dramatically and define the next generation cardiac CT system.

Until recently, all manufacturers offered products with up to 4cm of longitudinal coverage, 330ms rotation time and about 8lp/cm (5% MTF cutoff) native spatial resolution. Because 50ms temporal resolution and 16cm coverage are very aggressive performance goals, all architectures in laboratory that have been proposed to achieve this level of performance come with severe tradeoffs in terms of cone-beam artifacts, scatter, and complexity. To define the state-of-the-art of cardiac CT, we set 16cm nominal coverage, 50ms nominal temporal resolution, 20lp/cm nominal spatial resolution, 10HU noise level, and 5mSv effective dose as the main technical target.

In recent years, a number of new CT technologies have been developed, including distributed X-ray sources, field emitters based on carbon nano tubes, photon-counting detectors, fast CT gantries with high-performance bearings, interior ROI reconstruction, *etc*. To single out the best candidate from all the available novel cardiac CT architectures, here we propose an evaluation scheme as

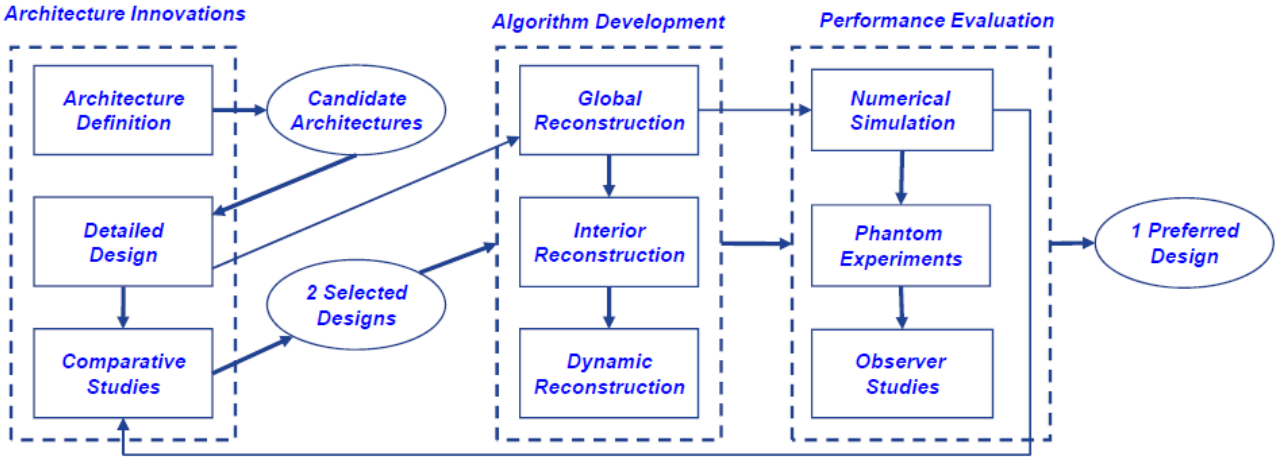


Figure 1. Process for selection of next generation cardiac CT architectures.

summarized in Fig. 1.

2. ARCHITECTURE INNOVATION

2.1. Architecture Definition

The overall goal of the architecture definition is to design, analyze, refine and identify superior cardiac CT architectures. In this first phase, we will define a number of candidate topologies: (1) baseline architectures, (2) saddle-curve scanning architectures, (3) triple-source architectures, (4) arc-source architectures, (5) nanotube-based architectures, (6) interior imaging schemes, and (7) instant/ultrafast CT schemes. We will also apply the TRIZ-methodology, a model-based technology for generating innovative solutions, to find any other promising architecture. These architectures will include flux optimization techniques such as dynamic source collimation, aggressive bowtie, virtual bowtie, dynamic bowtie, and advanced mA modulation [2]. For each of the architectures, we will develop a high-level conceptual design targeting cardiac CT with 16cm nominal coverage, 50ms nominal temporal resolution, and 20lp/cm nominal spatial resolution. This high-level design will include geometrically accurate drawings of each architecture and component selection.

Baseline Architectures. This includes all commercially available CT architectures and will serve as the baseline for our purpose. Most commercial CT scanners use the third-generation geometry. To meet the nominal coverage, this system will require a very large detector, and cone-beam artifacts will be a major challenge. We will investigate a triple-source CT architecture explained below. The best temporal resolution today is achieved with the EBCT scanner, which deflects an electron-beam to produce x-rays along semi-circular target rings. This architecture is fairly complex and limited in source intensity, but it provides extremely fast imaging capability. We will investigate a newer approach based on discrete electron emitters as explained under stationary architectures below.

Saddle-curve Scanning Architectures. Line sources with multiple longitudinally offset focal spots were proposed several years ago for CT [3]. They are a natural solution for achieving large volumetric coverage with limited cone-beam artifacts. If the number of focal spots is large they can even be used to implement a saddle trajectory [4-5]. In 2007, we invented a composite-circling scanning mode (Fig. 2) and associated cone-beam reconstruction methods to solve the quasi-short object problem [6]. This approach to cone-beam CT may have significant advantages in artifact reduction over existing cardiac CT scanners and in scatter rejection over the standard saddle-curve scheme, respectively.

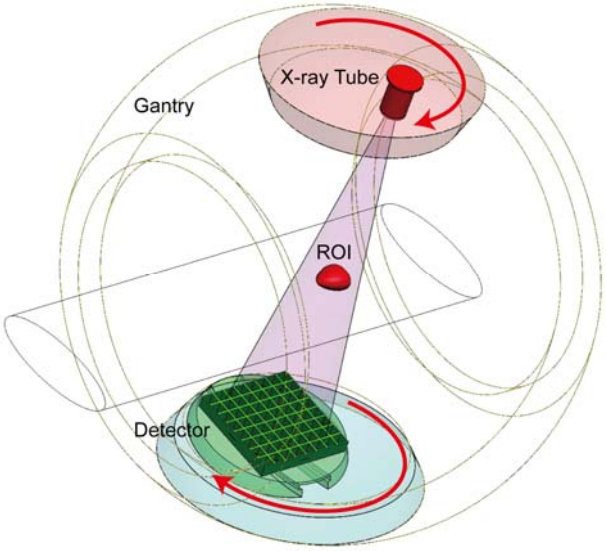


Figure 2. Compositing-circling mode. In such a CT system, the scanning trajectory is a composition of two circular motions: while an x-ray focal spot is rotated on a plane facing an object to be reconstructed, the x-ray source is also rotated around the object on the gantry plane.

Triple-source Architecture. The Siemens dual-source scanner has received a major attention in the field. A natural extension of the dual-source system is a triple-source cone-beam CT (CBCT) scanner. A first advantage of the triple source configuration is obviously the further improvement in temporal resolution. In addition, this architecture offers specific advantages in cone-beam reconstruction.

Arc-source Architectures. A different concept with a potential for cardiac CT is based on an arc source, or a source with multiple azimuthally offset focal spots. In one design, each focal spot irradiates the same cardiac ROI alternatively. Since the spots are azimuthally offset, they can acquire data over a large view range in a limited amount of time (~ milliseconds). To acquire sufficient data for a half-scan reconstruction several datasets need to be combined at different angular intervals, such as used with multi-sector reconstruction. This architecture can potentially yield greatly improved temporal resolution, but may suffer from increased scatter and limited flux.

Nanotube-based Architectures. Recently, Dr. Otto Zhou's group reported the development of a micro-CT scanner based on a carbon nanotube (CNT) based x-ray source being rotated around a stationary mouse bed [7]. Their system can provide high temporal and spatial resolutions in small animal imaging. We believe that the CNT-based source technology can be extended for clinical cardiac CT in the future.

Interior Imaging Schemes. The simple way to implement interior CT is to collimate the x-ray beam to a reduced FOV, using a collimator similar to the ones used today. This would require accurately centering the cardiac region-of-

interest. Several concepts have been proposed by us and others for aggressive and dynamic bowtie filters, which enable imaging protocols that target a central ROI.

Instant/Ultrafast CT Schemes. Similar to an EBCT scanner, an instant/ultrafast CT scanner is highly optimized for temporal resolution. We very recently introduced a specific concept for cardiac instant/ultrafast CT [8], shown schematically in Fig. 3. Interior tomography is particularly desirable for this scheme. Assuming a cardiac FOV of about 10cm, a magnification of two, and a detector width of 20cm, an architecture with several tens of source-detector pairs is conceivable. Challenges with scattered radiation can be addressed by optimizing the detector-patient air-gap and the number of sources that are fired simultaneously.

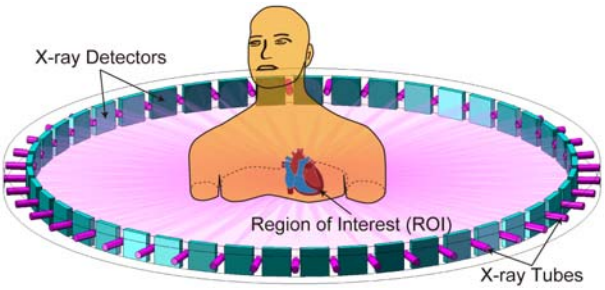


Figure 3. Instant tomography scheme as an extreme of the multi-source CT concept, combined with compressive sampling and interior tomography theories.

2.2. Detailed Design

After having defined the cardiac CT architectures at a high level in subsection 2.1, we will perform a detailed design, based on the target performance indexes.

Design Spreadsheet. For each of the cardiac architectures we will develop a detailed design spreadsheet including all detailed dimensions, scan times, sampling rates, photon fluxes, nominal image quality metrics, and their relationships. Then, we will optimize the design and address any critical design questions.

Component Assessment. The majority of the studied cardiac CT architectures will require innovations in the X-ray source, the collimation, the detector, the anti-scatter grid, the gantry and other components. For each design we will define the required subsystems in detail and assess the technical feasibility of these components, their performance bounds, and manufacturability issues.

System-level Considerations. After the characteristics of each key component are analyzed, we will further analyze the major system integration issues. For example, for the engineering implementation of our proposed composite-scanning mode, we recognize that the anti-scatter grid must be effectively addressed [6].

Image Quality Evaluation. The image quality is primarily characterized by two aspects: resolution and artifacts [9]. Image resolution has three aspects: high-contrast resolution (spatial resolution) for distinguishing adjacent objects of high-density, low-contrast resolution (contrast resolution) for differentiating an object from its background which is similar to the object in terms of gray-scale, and temporal resolution for capturing structures in motion which is most relevant to cardiac imaging. Image noise imposes a grainy appearance due to random fluctuations of the x-ray photon flux, and is a key factor in limiting low-contrast resolution. Image artifacts are structured or patterned interferences.

Dose Requirements. We will pay a special attention to the radiation dosage and its minimization. The performance criteria are fundamentally linked to patient dose. It will be key to assess the dose-efficiency of the various approaches, or more specifically whether the flux profile is well optimized and whether the detection efficiency is high. We will estimate the effective dose for a typical cardiac CT scan using each of the proposed architectures subject to the established clinical dose benchmark 5mSv.

Cost Estimation. Our long-term goal is to develop the next generation cardiac CT for large and medium-sized hospitals and clinics in the US. Based on the aforementioned engineering analysis, we will estimate the development or prototyping expenses, mass production costs, maintenance burdens, and so on. Then, we will predict the market prices of the cardiac CT designs of our interest (relative to today's commercial scanners), which will be used to establish the cost-effectiveness and competitive advantages of these architecture candidates.

2.3. Comparative Studies

All the designs will be entered in a standardized table including all performance criteria and the results from the overall design, engineering and cost analyses and preliminary numerical studies of key image quality and radiation dose indexes. Each performance criterion has a lower specification limit (LSL) and an upper specification limit (USL). A score will be given based on how well each specific design meets the specification. Each criterion also has a relative importance weight. Based on the weighted sum of all scores, the different concepts will be ranked. Where appropriate, multiple design iterations will be performed, until a quasi-optimal design is achieved for each type of architecture. This tradeoff study will result in a systematic quantitative comparison and enable us to obtain two promising architectures for the algorithm development described in section 3 and the performance evaluation described in section 4.

3. ALGORITHM DEVELOPMENT

Our algorithm development is categorized into three classes: (1) global reconstruction, (2) interior reconstruction, and (3) dynamic reconstruction. While global reconstruction algorithms will be used to produce key image quality and

radiation dose indexes for the aforementioned comparative studies (subsection 2.3), all these types of algorithms will be developed and optimized for the selected cardiac CT architectures. Furthermore, we will perform the algorithm development (for iterative algorithms in particular) in the compressive sampling framework [10-12].

4. PERFORMANCE EVALUATION

The architectures and algorithms will be evaluated at three levels. First, a more detailed evaluation will be conducted with numerical tests using the well-established comprehensive GE CT simulator CatSim [13]. The competing reconstruction algorithms will be compared for the selected architectures in terms of major quality and dose indexes. Second, where possible measurement experiments will be done using commercial CT scanners or research CT scanners to emulate the selected architectures and validate the associated algorithms with real projection measurements from various phantoms. This is to ensure all relevant physical effects are considered. Third, observer studies [14-15] will be performed to evaluate the images from the experimental studies as well as the selected system designs.

We will perform two types of observer studies: (1) an image evaluation study and (2) a system evaluation study. For the image evaluation study, we will focus on the simulation and phantom studies with the antropomorphic phantoms. The system evaluation study will be based on the "virtual brochure" method. We will produce virtual brochures for the two most promising candidates of our cardiac CT designs, including key specifications, system diagrams, scan protocols, simulated images and price estimates (relative to today's commercial scanners). Then, a marketing survey will be conducted to collect ratings and determine both the clinical demand and the commercial viability. Specifically, we will define six levels of efficacy as the probability of benefit to individuals from each system in the content of cardiac imaging, which are technical efficacy, diagnostic accuracy, diagnostic thinking efficacy, therapeutic efficacy, patient outcome, and societal efficacy [16-17]. Combining the observer studies with the quantitative studies we will have a very solid foundation to define the next-generation cardiac CT system.

5. CONCLUSIONS

In this paper, we have proposed a scheme for evaluation of promising candidate cardiac CT architectures to define the next generation cardiac CT architecture. It will help make a major step forward in cardiac CT performance and patient radiation dose. We invite critiques and suggestions from peers and welcome collaborative opportunities.

6. ACKNOWLEDGEMENT

This work is supported by NIH/NIBIB grant EB011785.

7. REFERENCES

[1] A. Berrington de Gonzalez, and S. Darby, Risk of cancer from diagnostic x-rays: Estimates for the UK and 14 other countries. Lancet, 2004. 363(9406): p. 345-351.
[2] B. De Man, et al., Computer assisted scan protocol and reconstruction (CASPAR), in RSNA. 2007: Chicago, IL.
[3] B. De Man, and S. Basu, Method and apparatus for employing multiple axial-sources, US patent 2005/0135550 a1. 2005.
[4] J.D. Pack, F. Noo, and H. Kudo, Investigation of saddle trajectories for cardiac ct imaging in cone-beam geometry. Phys Med Biol, 2004. 49(11): p. 2317-36.
[5] H.Y. Yu, et al., Exact BPF and FBP algorithms for nonstandard saddle curves. Med Phys, 2005. 32(11): p. 3305-3312.
[6] H,Y, Yu, and G. Wang, Cone-beam composite-circling scan and exact image reconstruction for a quasi-short object. Int. J. of Biomedical Imaging, 2007, Article ID: 87319, 10 pages.
[7] G. Cao, et al. A dynamic micro-ct scanner with a stationary mouse bed using a compact carbon nanotube field emission x-ray tube. in SPIE-Medical Imaging 2009.
[8] G. Wang, H.Y. Yu, and Y.B. Ye, A scheme for multi-source interior tomography. Med Phys, 2009. 36(8): p. 3575-3581.
[9] E.E. Christensen,and T.S. Curry, Physics of diagnostic radiology. 1990: Lea & Febiger,U.S.
[10] D.L. Donoho, Compressed sensing. IEEE Transactions on Information Theory, 2006. 52(4): p. 1289-1306.
[11] E.Y. Sidky, C.M. Kao, and X.C. Pan, Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT. J. of X-Ray Science and Technology, 2006. 14(2): p. 119-139.
[12] H.Y. Yu, and G. Wang, Compressed sensing based interior tomography. Phys Med Biol, 2009. 54(9): p. 2791-2805.
[13] B. De Man, et al. Catsim: A new computer assisted tomography simulation environment. Proceedings of SPIE, Vol. 6510, No. 65102G. 2007. San Diego, CA.
[14] A. Wunderlich, and F. Noo, Image covariance and lesion detectability in direct fan-beam x-ray computed tomography. Phys Med Biol, 2008. 53(10): p. 2471-2493.
[15] A. Wunderlich, and F. Noo, Estimation of channelized hotelling observer performance with known class means or known difference of class means. IEEE Transactions on Medical Imaging, 2009. 28(8): p. 1198-1207.
[16] D.G. Fryback, and J.R. Thornbury, The efficacy of diagnostic-imaging. Medical Decision Making, 1991. 11(2): p. 88-94.
[17] E.A. Krupinski, and Y.L. Jiang, Anniversary paper: Evaluation of medical imaging systems. Med Phys, 2008. 35(2): p. 645-659.

REGULARIZED 3D ITERATIVE RECONSTRUCTION ON A MOBILE C-ARM CT

Yongsheng Pan and Ross Whitaker[*]

Scientific Computing and Imaging Institute
University of Utah, Salt Lake City, Utah, 84112

Arvi Cheryauka and Dave Ferguson[†]

GE Healthcare-Surgery
Salt Lake City, Utah, 84116

ABSTRACT

3D iterative CT reconstruction is an active research area in medical imaging. Compared with analytic reconstruction methods such as FDK, iterative methods may provide better reconstruction results for incomplete and noisy projection data. The simultaneous algebraic reconstruction technique (SART), one of the most popular iterative reconstruction methods, is applied in the cone-beam geometry for high-resolution reconstruction, with the help of graphics hardware (GPU) and total variation (TV) regularization. GPU greatly improves the efficiency of SART, which is computationally intense for CPU, and thus makes it suitable for clinical applications. TV regularization reduces the effects of noise and helps the convergence of SART for noisy data. Experimental results for both synthetic and real data are provided to evaluate the accuracy and efficiency of the proposed framework.

Index Terms— Cone-beam CT, iterative reconstruction, SART, GPU, TV regularization

1. INTRODUCTION

Iterative CT reconstruction methods such as SART have been proposed since the late eighties [1]. These methods have advantages over analytical reconstruction methods such as FDK [2] for incomplete and noisy projection data. However, most industrial manufacturers have utilized FDK in their products so far because the high computational cost of SART hinders its practical application.

Iterative reconstruction methods have recently become active again due to the rapid developments of commodity hardware, such as GPU [3] [4] and Cell BE processor [5]. This hardware may greatly enhance the efficiency for SART and make SART appropriate for clinical applications.

On the other hand, regularizations are usually necessary for SART to reduce the effects of noise and enhance convergence, especially for projection data with strong noise. Total-variation (TV) minimization is a good method for the regularization of SART.

The performance of SART is studied in this paper. A GPU is utilized to improve its efficiency, while TV minimization [6] is utilized to regularize the SART algorithm. Reconstruction results for both synthetic and real data are presented.

The paper is organized as follows. Section 2 introduces the background information for SART and TV regularization. GPU implementation details are shown in Section 3. Experimental results are provided in Section 4, followed by the summary in Section 5.

2. BACKGROUND

Mathematical details of the SART algorithm and the TV regularization are provided in this section. These algorithms form the framework implemented in Section 3.

2.1. Introduction to SART

SART [1] takes every pixel in the object to be reconstructed as an unknown variable, and and it takes each projection measurement as a weighted summation of these variables. SART accomplishes CT reconstruction by solving the unknown variables from the acquired measurements. Specifically, SART is designed to solve the following simultaneous equation system

$$p_i = \sum_{j=1}^{N} w_{ij} v_j \tag{1}$$

where p_i represents the ith projection, w_{ij} represents the weight which the voxel v_j contributes its value to the projection i. Reconstruction is achieved by finding v_j from the equation system (1). The weight w_{ij} is assumed to be known.

The SART algorithm solves the equation system by iteratively applying a correction array to each voxel v_j as follows

$$v_j^{k+1} = v_j^k + \lambda \frac{\sum_i \left\{ w_{ij} \frac{p_i - \sum_{m=1}^{N} w_{im} v_m^k}{\sum_{m=1}^{N} w_{im}} \right\}}{\sum_i w_{ij}} \tag{2}$$

where λ is a constant coefficient. This process in Eq. 2 can be decomposed into two steps: the forward projection step and the backward projection step. The forward projection step computes a correction image for each ray, i.e.,

$$\Delta r_i = \frac{p_i - \sum_{m=1}^{N} w_{im} v_m^k}{\sum_{m=1}^{N} w_{im}} \tag{3}$$

[*] Send correspondence to Ross whitaker (whitaker@cs.utah.edu).
[†] Arvi.Cheryauka@med.ge.com, David.P.Ferguson@med.ge.com

The backward projection step updates each voxel by backprojecting to it the contribution of each correction image, i.e.,

$$v_j^{k+1} = v_j^k + \lambda \frac{\sum_i w_{ij} \Delta r_i}{\sum_i w_{ij}} \tag{4}$$

The SART algorithm has advantages over analytic reconstruction methods based on FDK [2], especially when few projections are available and when the projections are noisy. It is also more stable to the variations in the imaging geometry, such as the moving trajectory of imaging sources and detectors. However, SART is computationally intense, and it may require a lot of memory to store the weight information. Furthermore, regularization is usually necessary for noisy data.

2.2. Introduction to TV Regularization

Total variation regularization [6] is a nonlinear image regularization method that reduces the total variation of an image while keeping the regularized image similar to the original image. Given an image f defined on domain Ω, this method seeks a regularized image u which minimizes the following energy function

$$F(u) = \int_\Omega \|\nabla u\| + \frac{\alpha}{2} \int_\Omega \|u - f\|^2 \tag{5}$$

where α is a constant coefficient. This energy functional has the nice property of preserving straight, sharp edges, and thus allows solutions to have a piecewise flat property.

The minimization of the energy functional (5) is numerically calculated using the following updating scheme

$$\frac{u_{n+1} - u_n}{\triangle t} = \nabla \cdot \frac{\nabla u_n}{\|\nabla u_n\|} - \alpha(u - f) \tag{6}$$

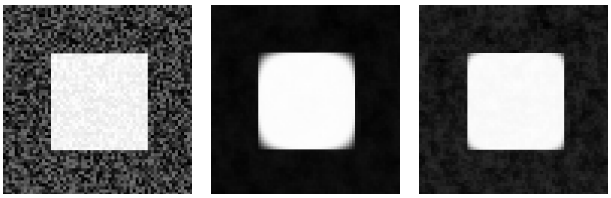
The minimization is iteratively processed until convergence.

Fig. 1 illustrates the regularization effects of TV. Fig. 1(b) shows the TV regularization results from the original image Fig. 1(a) with the coefficient $\alpha = 0.1$, and Fig. 1(c) shows the TV regularization results with the coefficient $\alpha = 0.5$. It can be seen that the TV regularization results contain less noise than the original image and that the results in Fig. 1(c) are closer to the original image with larger α.

3. IMPLEMENTATION DETAILS

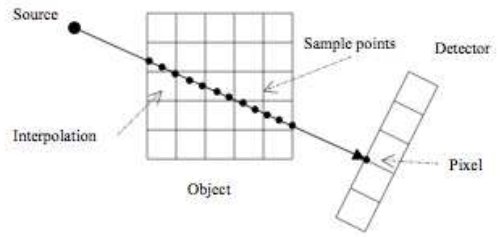
SART and TV are combined for iterative reconstruction in this paper. Specifically, Eq. 3, Eq. 4 and Eq. 6 are applied sequentially for a specified number of iteration until convergence. GPU implementation details for Eq. 3 and Eq. 4 are illustrated here, utilizing the methods in [7].

The mechanism of ray-based forward projection for Eq. 3 is illustrated in Fig. 2. For each projection, the ray from the source to each pixel in the detector plane is spatially determined, and the intersection points of the ray entering and exiting the object are calculated. The projection measurement is then calculated by accumulating the samples from the object along the ray between the intersection points using an equidistant step size. Interpolation methods such as trilinear interpolation may be applied to specify the object value at each sampling point. In the GPU implementation, the object data is stored as a 3D texture to utilize the hardware-accelerated interpolation functionality in the graphics card. Furthermore, GPU computes the weight information on the fly without storing it to save GPU memory.



(a) (b) (c)

Fig. 1. Illustration of TV regularization. (a) Original noisy image. (b) Image regularized using TV with $\alpha = 0.1$. (c) Image regularized using TV with $\alpha = 0.5$.



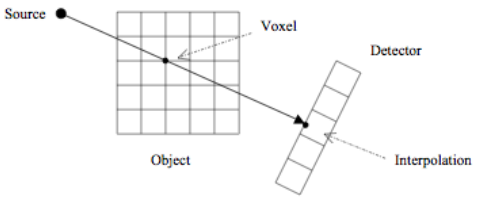
(a)

Fig. 2. Illustration of ray-based forward projection using GPU.

The mechanism of voxel-based backward projection for Eq. 4 is illustrated in Fig. 3. For each voxel in the object, the ray determined by the source and this voxel is utilized to calculate the intersection point with the detector plane. The correction value for this voxel is then calculated by interpolating the values in the detector plane. In the GPU implementation the correction data from the detector plane is stored as a 3D texture for fast data access and efficient hardware-accelerated interpolation.

4. EXPERIMENTAL RESULTS

Experimental results for cone-beam CT are presented here to illustrate the GPU implementation of SART. Nvidia CUDA is utilized for the GPU implementation on the Nvidia's



(a)

Fig. 3. Illustration of voxel-based backward projection using GPU.

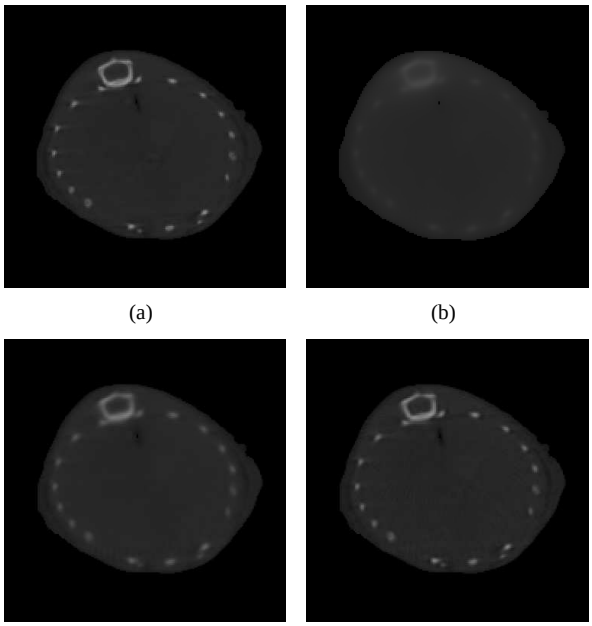
Geforce GTX 280 GPU, which has 240 processor cores with 1296MHz processor clock, 1GB GPU memory with 1107 MHz memory clock, and 141.7GB memory bandwidth.

Fig. 4 shows the SART reconstruction results for large real data using GPU. The object is a chest of a mouse of size 256×256×193. 100 projections are generated with the detector size 339×339. Fig. 4(a) shows the original image for slice 110 in the mouse data. Fig. 4(b)-Fig. 4(d) represent the SART reconstruction results after 5, 40, and 200 iterations respectively. The whole 200 SART iterations take 22.9 seconds using the Nvidia Geforce G280 graphics card, with 0.1 second for each iteration on the average. Fig. 5 shows the volume rendering of the corresponding reconstruction results in Fig. 4 using the same transfer function. These results show that the GPU implementation of SART may achieve accurate reconstruction in a very efficient way.

Fig. 6 demonstrates the effects of TV regularization on SART results for noisy projection data. Fig. 6(a) shows the 110th slice of the original data. Multiplicative noise is added to the generated projection data. If the value of a projection is p, its value is changed to be $p(1 - 0.3\alpha)$, where α is a random number between 0 and 1. Fig. 6(b) shows the 1st slice of the projection data before adding noise, while Fig. 6(c) represents the same slice after noise is added. Fig. 6(d) and Fig. 6(f) show the SART reconstruction results from noisy projection data after 20 and 200 iterations with no TV regularization. Fig. 6(e) and Fig. 6(g) show the SART results from noisy projection data after 20 and 200 iterations with TV regularization. It can be seen that TV regularization helps SART achieve much better reconstruction results, which contains much less noise. The reconstruction results in Fig. 6 with TV regularization take 25.8 seconds for 200 iterations.

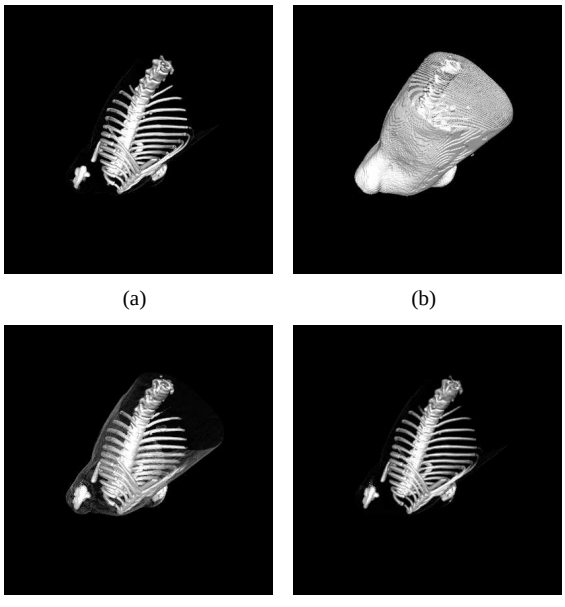
5. SUMMARY

Iterative CT reconstruction methods, which are robust to incomplete and noisy projection data, have great potential in real applications. SART, along with TV regularization, is utilized in the paper for reconstruction. CUDA GPU is utilized to speed up the computation. This paper presents the results on 3D real images. The results show that GPU-accelerated



(a) (b) (c) (d)

Fig. 4. GPU reconstruction results using SART for large mouse chest data. (Object volume: 256×256×193, Projection data: 100×339×339, no TV). (a) Original image for slice 110. (b) SART results after 5 iterations. (c) SART results after 40 iterations. (d) SART results after 200 iterations.



(a) (b) (c) (d)

Fig. 5. Volume rendering of GPU SART reconstruction results for a mouse chest. No TV. (Object volume: 256×256×193, Projection data: 100×339×339). (a) Original volume. (b) SART results after 5 iterations. (c) SART results after 40 iterations. (d) SART results after 200 iterations.

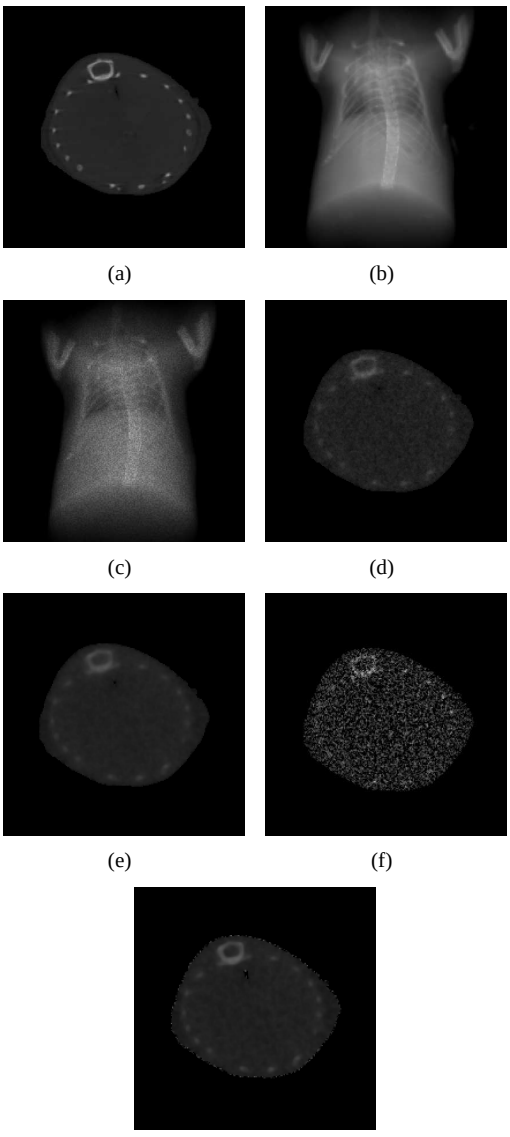


Fig. 6. 3D SART using TV regularization for mouse chest data. (Object volume: 256×256×193, Projection data: 100×339×339). (a) The 110th slice from the original volume. (b) The 1st slice from the projection data. (c) The 1st slice from the projection data with added noise. (d) SART results after 20 iterations with no TV regularization. (e) SART results after 20 iterations with TV regularization. (f) SART results after 200 iterations with no TV regularization. (g) SART results after 200 iterations with TV regularization.

SART algorithm with TV regularization may generate high-quality reconstructions with high potential for clinical applications. Future work will be focused on the SART convergence with respect to noise and new regularization methods.

6. REFERENCES

[1] A.C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*, IEEE Press, New York, 1999.

[2] L.A. Feldkamp, L.C. Davis, and J.W. Kress, "Practical cone-beam algorithm," *J. Opt. Soc. Am. A*, vol. 1, no. 6, pp. 612–619, 1984.

[3] K. Mueller and R. Yagel, "Rapid 3-D cone-beam reconstruction with the Simultaneous Algebraic Reconstruction Technique (SART) using 2-D texture mapping hardware," *IEEE Trans. on Medical Imaging*, vol. 19, no. 12, pp. 1227–1237, 2000.

[4] F. Xu and K. Mueller, "Real-time 3d computed tomographic reconstruction using commodity graphics hardware," *Physics in Medicine and Biology*, vol. 52, pp. 3405–3419, 2007.

[5] O. Bockenbach, M. Knaup, and M. Kachelrieß, "Implementation of a cone-beam backprojection algorithm on the cell broadband engine processor," in *Medical Imaging 2007: Physics of Medical Imaging*, February 2007, vol. 6510 of *Proceedings of SPIE*, pp. 1–10.

[6] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259–268, 1992.

[7] R. R. Galigekere, K. Wiesent, M. Kowarschik, and J. Hornegger, "Cone-beam reprojection using projection-matrices," *IEEE Trans. on Medical Imaging*, vol. 22, no. 10, pp. 1202–1214, October 2003.

Parallel ART

Cyril Riddell, GE Healthcare, 283 rue de la Minière, 78533 Buc cedex

I. Introduction

The Algebraic Reconstruction Technique (ART) is a fast algorithm used in tomographic image reconstruction [1]. Further developments have lead to the theory of the projection onto convex sets (POCS) that has grounded ART in a solid theoretical framework [2]. POCS projectors and projections are intended in the sense of an operator that generates "the unique projection Px of a point x on convex set C as that point in C closest to x".

Original ART principle is as follows:

a) A set of measurements is collected, from which an unknown vector must be estimated.
b) A set of projectors onto convex sets is defined, one projector per measurement, that projects an image vector onto a convex sub-space where all image vectors verify the measurement. Each measurement is therefore an equality constraint that the corresponding projector enforces on any given vector
c) The solution is found iteratively by applying, at each iteration, all projectors successively.

POCS theory adds to this basic framework the possibility of applying non-linear constraints. It thus extends ART to a tool for convex optimization of non-differentiable criteria (e.g. total variation). In the rapidly growing field of compressed sensing, dynamic tomographic applications have been demonstrated using updated versions of the ART algorithm associated with total variation minimization [3-4].

ART presents several advantages in tomography:

- No assumption on the acquisition pattern is required to provide optimal reconstruction.
- Theoretical derivation of the tomographic projector for one measurement is simple.
- Convergence is proven under the assumption that the solution exists.
- An update of the vector to be estimated is performed by each projector, which provides fast convergence.
- Additional constraints can be accommodated, especially non-linear, as long as they can be expressed as a projector onto a convex set.

One must also acknowledge several limitations:

- If the data is inconsistent, as in the presence of noise, and no solution exists, the algorithm will only cycle through a set of feasible solutions.
- Computationally speaking, an update after each single measurement may not be optimal: most efficient parallel implementations are obtained by simultaneous processing of a block of measurements. This has lead to an alternative algorithm: Simultaneous-ART (SART) [5], which, however, is not a POCS technique.
- Even though the projector formula is simple, it involves a norm computation that does not exist in SART and standard least-squares techniques.

In the following, we will propose a parallelization of the computation of ART for tomographic reconstruction, based upon the observation that application of constraints along measurements lines that do not intersect, lead to the same updated volume whether the projectors are applied simultaneously or successively. Although this observation is obvious in the continuous case, it has not been exploited in the discretization phase, since, after discretization, parallel lines of interpolated samples may share pixels. Thus, in the SART iteration, simultaneous use of measurements along lines that do not intersect is not equivalent to the successive application of each measurement projector defined in ART. It is therefore distinct from ART and associated POCS theory.

In the following, we shall revisit the discretization of ART through an extrinsic – intrinsic decomposition that allows for a "parallel ART" scheme that will maintain the equivalence between successive and simultaneous updates of a vector for a set of parallel measurement lines. Numerical schemes will be proposed that are arguably simpler and faster than standard ART implementations, while maintaining the theoretical properties of POCS, such as convergence and use of non-linear constraints.

II. Theory

A. Linear constraints

We consider a tomographic linear operator R that captures the geometry of the measurement system (parallel-, fan- or cone-beam geometry), the sampling (number and sizes of image pixels and measurement bins), the chosen interpolation (e.g. nearest neighbor, linear, Fourier), the trajectory (e.g. circular, helical, tomosynthesis) and, possibly, additional non-stationary physical aspects to be corrected for during the reconstruction.

A set of measurements defines a set of linear constraints.

Let us first recall the expression of the projection $P_s(f_0)$ of a vector f_0 as the closest vector that verifies a set of linear constraints $RP_s(f_0) = s$. This projector is defined as a constrained optimization problem:

$$P_s(f_0) = \arg\min_f \left\{ \|f - f_0\|^2 \text{ s.t. } Rf = s \right\} \quad (1)$$

The interest of the ART method lies in the fact that, if one considers only a subset of constraints, i.e. $R_\theta f = s_\theta$ where θ is a subset of measurement indices, inversion of $\left(R_\theta R_\theta^t\right)^{-1}$ might be easier to express and compute. A set of projectors can then be defined with the property that successive application of these projectors will lead to the global solution. Under the assumption that the solution exists, we have:

$$\lim_{n\to\infty}\left(\prod_\theta P_\theta\right)^n (f_0) = P_s(f_0) \quad (2)$$

Indeed, if we further explicit operator R as a system matrix $R=(r_{ij})_{1\le i\le I,1\le j\le J}$ relating J pixels of index j to I measurements of index i, and if the subset is reduced to one linear measurement, then: $R_\theta = r_j$ is a row of matrix R, $s_\theta = s_j$, $R_\theta f = r_j.f$ and $R_\theta R_\theta^t = r_j.r_j^t = \|r_j\|^2$.

The projector therefore reduces to:

$$P_j(f) = f + \left(\frac{s_j - r_j.f}{\|r_j\|^2}\right) r_j^t \qquad (3)$$

B. Change of variable

We denote U a unitary operator, i.e. such that $U^{-1}=U^t$ (symbol ^{t} is used indifferently for the transpose of a real operator or for the conjugate of a complex operator). We consider the same set of linear constraints $Rf = s$. A change of variable $\tilde{f} = U^t f$ leads to the equivalent constraints $\tilde{R}\tilde{f} = s$ with $\tilde{R} \equiv RU$. The associated projector $\tilde{P}$ is defined as:

$$\tilde{P}_s(\tilde{f}_0) = \arg\min_f \left\{ \|f - \tilde{f}_0\|^2 \text{ s.t. } \tilde{R}f = s \right\} \quad (4)$$

$\tilde{P}_s(\tilde{f}_0)$ is such that

$$\begin{aligned} \|\tilde{P}_s(\tilde{f}_0) - \tilde{f}_0\|^2 &= \|\tilde{P}_s(U^t f_0) - U^t f_0\|^2 \\ &= \|U\tilde{P}_s(U^t f_0) - f_0\|^2 \end{aligned} \qquad (5)$$

and

$$\tilde{R}\tilde{P}_s(\tilde{f}_0) = RU\tilde{P}_s(U^t f_0) = s \qquad (6)$$

therefore

$$U\tilde{P}_s U^t = P_s \Leftrightarrow \tilde{P}_s = U^t P_s U \qquad (7)$$

We conclude that:

$$\tilde{R} = RU \Rightarrow \tilde{P}_s = U^t P_s U \qquad (8)$$

We will use extensively this result thereafter.

III. Extrinsic-intrinsic decomposition

Extrinsic-intrinsic decomposition consists in splitting a tomographic acquisition into the succession of two generic operations: one, *extrinsic*, that samples the patient on a grid, the other, *intrinsic*, that models the tomographic acquisition of the patient on that grid. Standard reconstruction defines the extrinsic matrix once for all as the original grid of samples to be estimated, while the intrinsic operator models all measurements with respect to this grid. One ART projector is defined per row of the matrix of this intrinsic operator. Even though this approach reflects the physical reality of a still patient inside a rotating scanner, it is conceptually equivalent to consider a rotating object inside a still scanner. With this point of view in mind, there still is a unique, hopefully simplified, intrinsic operator that models the acquisition sampling pattern, together with as many extrinsic operators as rotation angles that rotate the object, i.e. re-samples the object over the grid defined by the intrinsic operator.

Let us consider again a subset of constraints $R_\theta f = s_\theta$ enforced by projector P_θ where θ is a subset of measurement indices. For each subset θ, we define an extrinsic operator, denoted U_θ. We denote R_0 the intrinsic operator such that $R_\theta = R_0 U_\theta$ and $P_{0,\theta}$ the projector enforcing constraint $R_0 f = s_\theta$.

We now assume that each extrinsic operator is *unitary*, i.e. such that $U_\theta^{-1} = U_\theta^t$. We further define $U_{-\theta} \equiv U_\theta^{-1}$. Each extrinsic operator is therefore used as a change of variable and we have:

$$R_\theta = R_0 U_\theta \Rightarrow P_\theta = U_{-\theta} P_{0,\theta} U_\theta \qquad (9)$$

This change of variable is easily captured using the matrix form of the projector:

$$\begin{aligned} P_\theta(f_0) &= f_0 + (R_0)^t \left(R_0 U_\theta (R_0 U_\theta)^t\right)^{-1} \left(s - R_0 U_\theta f_0\right) \\ &= U_{-\theta}\left(U_\theta f_0 + (R_0)^t \left(R_0 R_0^t\right)^{-1} \left(s - R_0 U_\theta f_0\right)\right) \\ &= U_{-\theta} P_0 (U_\theta f_0) \end{aligned} \qquad (10)$$

Therefore the set of projectors P_θ is replaced by the combined use of extrinsic unitary operators U_θ and a set of projectors $P_{0,\theta}$ derived from the intrinsic operator R_0. Thus, the ART iteration can be expressed as:

$$\prod_\theta P_\theta = U_{-\theta(N)} \left(\prod_{n=1,\dots,N-1} U_{\theta(n+1)} U_{-\theta(n)} P_{0,\theta} \right) U_{\theta(1)} \qquad (11)$$

Since each extrinsic operator U_θ is unitary and invertible, the ART iteration gets simplified since the inverse extrinsic operator of a subset can be combined to the extrinsic operator of the next subset, thus reducing computation. The use of rigid transforms for U_θ will accommodate any trajectory by sampling the object with a grid aligned with the detector sampling pattern.

For cone-beam geometry, it is standard in the field of "computer vision" to use the product of a rigid transform and a pure perspective projective transform [6]. The perspective transform is constant at all positions since it is defined by the camera characteristics, and the extrinsic operator is restricted to the rigid transform (modeling the camera displacement).

In the field of "computer graphics", it is standard to further factorize the perspective transform as the product of a pure elastic transform (change of scales in the continuous case and of sampling rates after discretization) and a so-called pure orthographic (i.e. parallel) projection transform [7]. This better reflects the conditions of C-arm-based cone-beam tomography, where deformations during the rotation introduce variations in the perspective geometry intrinsic matrix that is commonly used, which cannot be considered truly intrinsic anymore.

Cone-beam tomography can therefore be handled with extrinsic operators that combine all rigid and elastic transforms, while keeping the same intrinsic operator as for parallel geometry: a pure orthographic transform.

IV. Extrinsic operator

A. Elastic transform

Once sampled, the elastic transform is not invertible in general: sub-sampling results in information loss that cannot be recovered by the inverse transform. If the elastic transform is constant for all extrinsic operators, it can be factorized among all projectors so that it does not interfere with convergence. In other words, the reconstruction is performed with rigid transforms only, then the final reconstruction is obtained by applying the inverse extrinsic transform. An approximate inverse due to interpolation errors will only alter this final step of the reconstruction. This is applicable to tomosynthesis when the source trajectory remains in a plane [8].

B. Rigid transform

Rotations and translations can be performed without information loss depending on the interpolation scheme. Importantly, a rigid transform based on linear interpolation will *not* result in a unitary matrix: the transposed matrix is not the exact inverse, raising the same problems as seen with elastic transforms. In a tomographic projector, the rotation can be decomposed into an elastic transform and a succession of translations [9]. This elastic transform can be removed from the iterative process by pre-processing the data. Therefore, invertibility of the translation is sufficient. Invertible translations can be achieved at least with nearest neighbor and Fourier interpolations. Nearest neighbor interpolation accuracy can be increased through over-sampling.

Importance of the invertibility of the extrinsic transform is illustrated on fig. 1 where data is simulated using parallel geometry that is reconstructed by parallel ART with an extrinsic matrix made of translation only.

V. Orthographic intrinsic operator

The proposed extrinsic – intrinsic decomposition is most advantageous when using an intrinsic operator that is a pure orthographic transforms. This means that the extrinsic transforms (using elastic pre-processing or not) have aligned the reconstruction grid with a sampling pattern as in parallel geometry. Intrinsic operator R_0 is therefore reduced to a sum over either the rows or the columns of the sampling grid. It is arguably the most intrinsic tomographic operator!

We now denote $P_{0,j}$ the ART projector for a single measurement, that is, for any line of the grid. Denoting L_j the set of pixel samples in line *j* and l_j its cardinal, $P_{0,j}$ has the following expression:

$$\left(P_{0,j}(f)\right)_{i\in L_j} = f_i + \frac{1}{l_j}\left(s_j - \sum_{i\in L_j} f_i\right) \qquad (12)$$

The projector has therefore a simple expression where all pixels of line L_j are updated with the average difference over L_j between a pixel sample of index $i \in L_j$ and the measurement *j*. Importantly, it does not involve interpolation. Since the projectors are applied along parallel lines, whether they are all rows or all columns, they update the grid samples independently of the other lines. There is therefore no difference between applying *successively* and *simultaneously* projectors $P_{0,j}$ for a subset θ of parallel lines of the grid.

We can thus define a single projector $P_{0,\theta}$ per subset θ of parallel lines such that:

$$P_{0,\theta} \equiv \prod_{j\in\theta} P_{0,j} \qquad (13)$$

Computationally speaking, on a subset per subset basis, all measurements can be used in parallel with the same output as if they were used simultaneously. It must be noted that the simplicity of projector $P_{0,j}$ derives from the simplicity of the intrinsic tomographic operator R_0.

VI. Constrained intrinsic operator

A. Tomographic positivity constraint

In POCS theory, component-wise thresholding or clipping of the image according to component-wise thresholds defines a projector onto convex set. This allows for applying a positivity constraint to the solution. This constraint can be further refined for the tomographic case by considering that the integral of a positive function along a line is greater or equal to the maximum of the function over this line:

$$\begin{aligned} &\left(\forall x\ f(x) \geq 0\right) \\ &\Rightarrow \int f(x)dx = \int \left|f(x)\right|dx \geq \sup_x \left|f(x)\right| \end{aligned} \qquad (14)$$

We call this enhanced constraint the tomographic positivity constraint, and we combine it with the intrinsic projectors to yield:

$$\begin{aligned} &\left(P_{0,j}^{T+}(f)\right)_{i\in L_j} \\ &\equiv \max\left(0, \min\left(f_i + \frac{1}{l_j}\left(s_j - \sum_{i\in L_j} f_i\right), s_j\right)\right) \end{aligned} \qquad (15)$$

This property enforces a stronger constraint than the standard positivity case over the solution. In particular:

$$s_j = 0 \Leftrightarrow \left(P_{0,j}^{T+}(f)\right)_{i\in L_j} = 0 \quad \forall i \in L_j \qquad (16)$$

This constraint is expected to be particularly effective when reconstructing contrast enhanced subtracted X-ray acquisitions.

B. Support constraint

A support constraint simply consists in setting to zero the components of the vector known a priori to be outside of the object support. It is partly redundant with the tomographic positivity constraint, which also contributes to defining a reconstruction support, since a zero measurement will set to zero all samples of the integration line. Samples that do not contribute to the measurements are thus excluded a priori from the reconstruction process. The only visible trace of this constraint is the value l_j that was defined as the number of samples in line L_j used by projector $P_{0,j}$. More precisely, this value should be the number of non-zero samples in the row. In the case $l_j = 1$, one and only one sample must be updated, and it is simply set to s_j.

VII. CHANGE OF REPRESENTATION DOMAIN

Up to now, the intrinsic operator has been defined in the spatial domain: the grid samples the object in the Euclidean space. A simple result of tomography states that for a suitable *f* function and suitable *w* functions forming a basis *W*:

$$\begin{aligned} & s(y) = \int f(x,y)dx \\ & \Rightarrow \int s(y)w(y) = \int \left[\int f(x,y)w(y)dy \right] dx \end{aligned} \quad (17)$$

For the intrinsic operator, *f* is a re-sampled version of the object of interest such that it is aligned with the set of measurement lines s_θ defined here by $s_{\theta,j} = \int f(x, y_j)dx$. We denote $\hat{s}_w = \int\int f(x,y)w(y)dxdy$ and $\hat{f}_{i,w} = \int f(x_i, y)w(y)dy$ the samples of *f* and s_θ in basis *W*. Application of (16) states that constraint $R_0 f = s_\theta$ implies constraint $R_0\hat{f} = \hat{s}_\theta$. For each measurement $\hat{s}_w$ in basis *W*, the associated projector simply is:

$$P_{0,w}(\hat{f}) = \arg\min_f \left\{ \left\| f - \hat{f} \right\|^2 \text{ s.t. } R_0\hat{f} = \hat{s}_w \right\} \quad (18)$$

leading to similar intrinsic projectors for each function w of the W basis:

$$\left(\hat{P}_{0,w}\left(\hat{f}\right)\right)_{i\in L_w} = \hat{f}_{i,w} + \frac{1}{l_w}\left(\hat{s}_w - \sum_{i\in L_w}\hat{f}_{i,w}\right) \quad (19)$$

A. *Fourier orthographic intrinsic operator*

Eq. 19 shows that the intrinsic orthographic operator has the same implementation if the vector is Fourier-transformed in the direction orthogonal to the orthographic projection. Each projector applies a constraint over each sampled frequency *w* defined by the Fourier transform. All frequencies are equivalently updated successively or simultaneously.

If we consider the ramp filter used in analytical reconstruction, we see that it is the application of a diagonal operator in the Fourier space. It combines derivation and the Hilbert transform. Let us denote this filter $D = \text{diag}\left(\hat{d}_w\right)$, it is obvious that $R_0\hat{f} = \hat{s}_\theta$ is equivalent to $DR_0\hat{f} = R_0 D\hat{f} = D\hat{s}_\theta$, leading to the ramp-filtered intrinsic projector:

$$\left(P^D_{0,w}\left(\hat{f}\right)\right)_{i\in L_w} = \hat{f}_{i,w} + \frac{\hat{d}_w}{l_w}\left(\hat{s}_w - \sum_{i\in L_w}\hat{f}_{i,w}\right) \quad (20)$$

Parallel-ART in the Fourier domain can therefore combine algebraic reconstruction with ramp filtering (or any other relevant filtering). In the cases where the extrinsic transform is made of translations, the whole parallel-ART algorithm can be implemented in the Fourier domain exclusively with eq. (20) associated with phase-shifts for the translations of the extrinsic operator.

VIII. CONCLUSION

An extrinsic-intrinsic decomposition of the tomographic problem has been proposed that decouples sampling from tomographic inversion. It allows for simpler expression of ART projectors that can be applied in Fourier space and combined with Fourier filtering. It isolates interpolation issues from tomographic modeling and thus might serve as a tool for evaluating whether the system model or the sampling is key to high-quality reconstruction of real tomographic data.

IX. REFERENCES

[1] R. Gordon, R. Bender, and G.T. Herman, "Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography," J. Theoretical Biology, vol. 29, pp. 471-481, 1970.

[2] D. C. Youla and H. Webb. "Image restoration by the method of convex projections: part 1 – Theory", IEEE Trans. Med. Imaging, vol. 1, pp 81-94, 1982

[3] G.-H. Chen, J. Tang and S. Leng "Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic CT images from highly undersampled projectiondata sets," Med. Phys. 35(2) pp. 660-663, 2008

[4] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," Phys. Med. Biol., vol. 53, pp. 4777-4807, 2008.

[5] A. H. Andersen, "Algebraic Reconstruction in CT from Limited Views," IEEE Trans. Med. Imaging., vol. 8, no.1, pp. 50-55, 1989.

[6] O. Faugeras, Three Dimensional Computer Vision: A Geometric Viewpoint. Cambridge, MA: MIT Press, 1993.

[7] OpenGL Architecture Review Board, D. Shreiner, M. Woo, J. Neider and T. Davis, OpenGL Programming Guide 5th Edition. The official guide to learning OpenGL, Version 2, Addison Wesley, 2006

[8] B. E. Nett, S. Leng and G.-H. Chen, "Planar tomosynthesis reconstruction in a parallel-beam framework via virtual object reconstruction," SPIE Proc. Med. Imag. pp. 6559-6579, 2007

[9] C. Riddell and Y. Trousset, "Rectification for cone-beam projection and backprojection," IEEE Trans. Med. Imaging. vol. 25, pp.950-962, 2006.

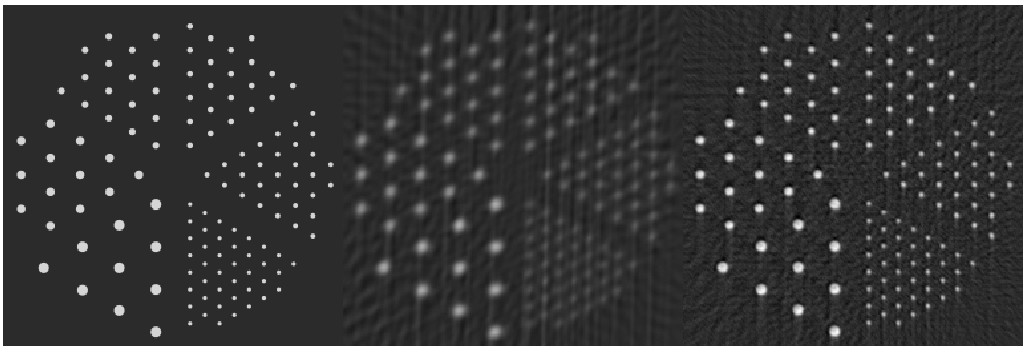


Fig 1: Reconstruction of simulated data with parallel-ART. Left image: reference image. Middle image: parallel ART reconstruction with non-invertible linear interpolation extrinsic matrix. Right image: parallel ART reconstruction with invertible nearest-neighbor interpolation extrinsic matrix.

3-D Geometry Calibration on a Mobile X-Ray C-arm for CT Imaging and Surgical Navigation

Prabhanjana Kalya, Rocco Cherone, and Arvi Cheryauka

Abstract– **The design of mobile X-ray C-arm equipment with image tomography and surgical guidance capabilities involves the retrieval of repeatable gantry positioning in three-dimensional space. Geometry misrepresentations can cause degradation of the reconstruction results with the appearance of blurred edges, image artifacts, and even false structures. It may also amplify surgical instrument tracking errors leading to improper implant placement. In our prior publications we have proposed a C-arm 3D positioner calibration method comprising separate intrinsic and extrinsic geometry calibration steps. Following this approach, in the present paper, we extend the intrinsic geometry calibration of C-gantry beyond angular positions in the orbital plane. Our method makes deployment of markerless interventional tool guidance with use of high-resolution fluoro images and electromagnetic tracking feasible at any angular position of the tube-detector assembly. Variations of the intrinsic parameters associated with C-arm motion are measured off-line as functions of orbital and lateral angles. The proposed calibration procedure provides better accuracy, and prevents unnecessary workflow steps for surgical navigation applications. With a slight modification, the Misalignment phantom, a tool for intrinsic geometry calibration, is also utilized to obtain an accurate 'image-to-sensor' mapping. We show simulation results, image quality and navigation accuracy estimates, and feasibility data acquired with the prototype system. The experimental results show the potential of high-resolution CT imaging (voxel size below 0.5 mm) and confident navigation in an interventional surgery setting with a mobile C-arm.**

I. Introduction

The design of mobile X-ray C-arm equipment with computer tomography and surgical guidance capabilities involves the retrieval of repeatable gantry positioning in three-dimensional space [1]. Geometry misrepresentation in the image reconstruction can cause degradation of the imaging results with the appearance of blur, artifacts, and even false structures. It may also amplify tracking error for surgical instruments and lead consequently to wrong implant placement. In our prior publications we have proposed a C-arm 3D positioner calibration method comprising separate intrinsic and extrinsic geometry calibration steps [2]. Following this approach, in the present paper, we extend the use of the intrinsic geometry calibration technique beyond the orbital plane into the 3D space of C-arm positioning. This geometric information is necessary for 2-D navigation, multi-modality data co-registration, and, potentially, for tomographic imaging deploying the C gantry motion trajectory into three-dimensional space [4].

Manuscript received February 1, 2010.

Prabhanjana Kalya, Rocco Cherone, and Arvi Cheryauka are with GE Healthcare-Surgery, Salt Lake City, UT 84116 USA.
Contact e-mail address: Arvi.Cheryauka@med.ge.com.

II. C-Arm Geometry Calibration

C-arm geometry characteristics can be derived by a) supremely accurate implementation of the gantry rotation, b) obtaining direct measurements using permanent sensors, or c) extracting markers from images projecting target of known geometry and material property. Following the image-based framework,

- The first step in our method is the intrinsic calibration that deals with delineation of geometry of the X-ray source-detector assembly.
- The subsequent motion characterization (extrinsic calibration) uses the previously obtained intrinsic parameters.

II.1. Cone-beam Camera Model

In the surgical room the C-arm gantry moves around the patient while the patient table is kept motionless. Since manufacture of a rigid gantry with highly accurate motion is extremely expensive, practical considerations necessitate the C-gantry to deviate from programmable position / velocity / acceleration profiles. We define a 3-D Camera pinhole model introducing 3 intrinsic (two plane coordinates of principal point and source-to-imager distance) and 6 extrinsic (3 translation and 3 rotation displacements) parameters needed to uniquely define a cone-beam (CB) projection in 3-D space [5]. The systematic deviations from the theoretical trajectory can be compensated in the backprojection step. In the next sections, we describe derivation of these 9 projection geometry values per X-ray projection from the computed assessment of the imaging results of the dedicated intrinsic and extrinsic calibration targets.

II.2. Intrinsic Geometry Calibration

Intrinsic parameters describe the relative positioning of the source and the detector, i.e. 3-D misalignment, which, in practice, can be caused by welding defects, mounting imperfectness, and deflection of the C-arc due to weight force. We have built a removable Misalignment target, which is rigidly attached to the detector surface during an off-line calibration sweep. The factual geometrical parameters of the

target used in image processing are measured by high-precision FARO-arm equipment [4]. The governing principle of image-based intrinsic calibration is to place a number of high-opacity markers in pre-defined positions with regard to detector pixels and then extract cone-beam geometry parameters from the observed shadow locations. The results of intrinsic geometry delineation, differences between observed and predicted positions of the marker's centroids, are shown in Figure 1.

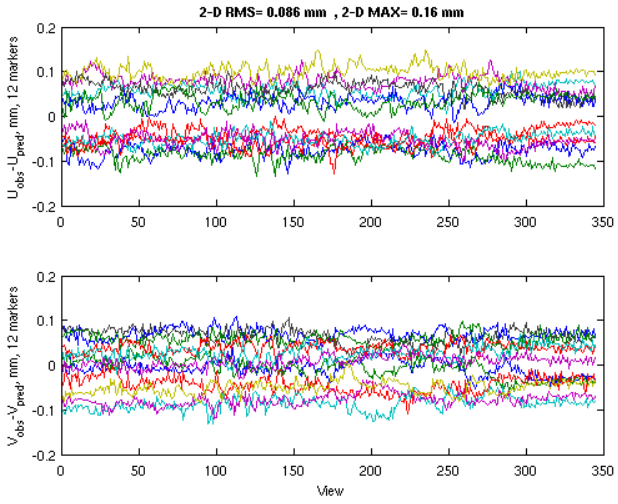


Fig. 1. Difference between observed and predicted position of the marker's centroids in the 2-D detector plane. During intrinsic geometry calibration 12 markers incorporated into Misalingment calibration target are tracked over 350 gantry positions.

It was confirmed that relative positioning of the x-ray tube and the detector has highly repeatable intrinsic geometry parameters. The maximum deviation of the intrinsic geometry parameters recorded over 2-week time period is less than 0.1 percent. Stable alignment characteristics are critical for maintaining accuracy in C-arm CT imaging and 2-D/3-D navigation over time.

II.3. Extrinsic Geometry Calibration

Extrinsic parameters deal with motion of the C-arm assembly in 3D space and relate the Camera and World coordinate systems. In a hypothetical cone-beam C-arm CT scan, gantry rotates 180+ degrees in one direction strictly in the orbital plane. In reality, the C gantry moves along some curve in 3-D space. Each time the gantry moves, the actual trajectory and its temporal characteristics can deviate from their pre-defined values. By acquiring experimental data the most probable trajectory can be identified and random deviations estimated.

In the first series of experiments, by running a series of orbital scans with the prototype C-arm CT system, and taking robotic arm measurements, we evaluate the systematic motion trajectory and the correspondent spectrum of the spatial deviations. The photo in Figure 2 captures measuring C gantry motion with use of the high-precision FARO-arm equipment. The robotic arm is connected to the C frame and tracks 3-D location of the ball-pointer located at the tip of the arm.

Fig. 2. High accuracy measurements of 3-D geometry with use of FARO-arm.

Here we summarize the recorded 10 cyclic orbital runs. Maximum angular velocity is equal to 12 degrees per second. In Figure 3, the 3-D markers project the locations of the bearing ball center expressed in the Reconstruction coordinate system, $\{X_{orb}, Y_{orb}, Z_{orb}\}$, associated with the C-arm mainframe. This coordinate system can be obtained by transaction of rotation and translation is based on least square (LS) fits of the coordinate point dataset to a plane and to a circle.

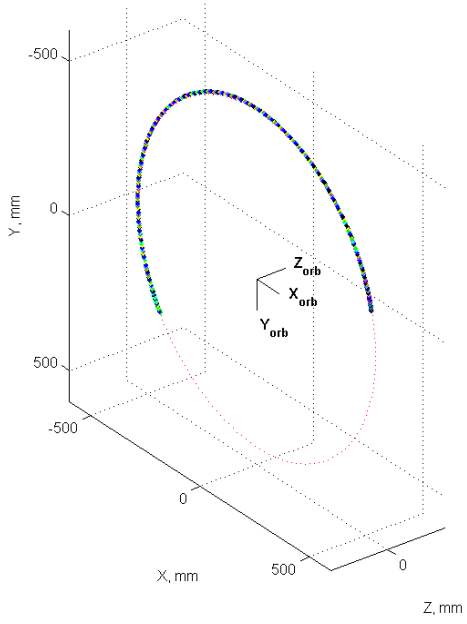


Figure 3. 3-D plot of the recorded data visualized as a cloud of markers. The data is expressed in the Reconstruction coordinate system, where X_{orb} and Y_{orb} axes define a LS plane of rotation and Z_{orb} axis represents a LS axis of rotation. The dotted line circle is a LS circle that fits the observed data.

The trajectories recorded from different runs overlap with a high degree of accuracy. The maximum and standard deviations for 3-D displacement from the systematic trajectory are equal to 1.09 mm and 0.17 mm, respectively.

For image-based extrinsic calibration, we have designed a Motion calibration target comprising a very small number of markers located at the periphery of FOV. The results of extrinsic geometry delineation, differences between observed and predicted positions of the marker's centroids, are shown in Figure 4.

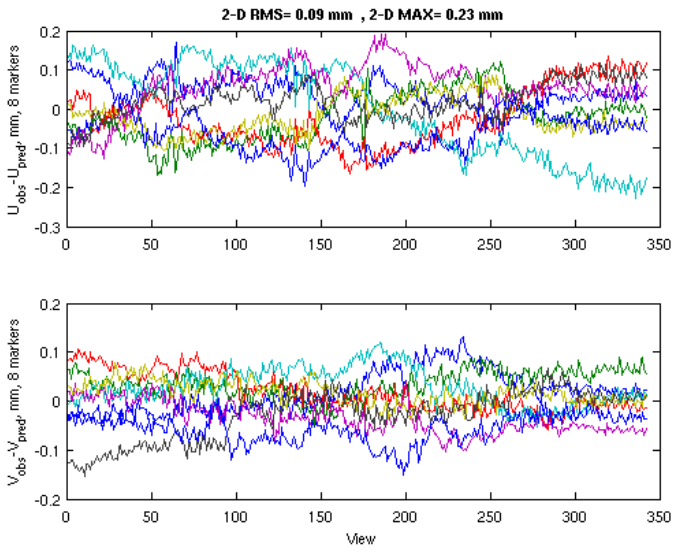


Fig. 4. Difference between observed and predicted position of the marker's centroids in the 2-D detector plane. During extrinsic geometry calibration 8 markers incorporated into the Motion calibration target are tracked over 350 gantry positions..

Overall maximum non-repeatability of the mechanical platform on the experimental prototype system has been observed to be less than 0.6 mm in 3-D space.

III. Orbital and Lateral C-Gantry Motion

The ISO-centric C-gantry of the prototype system has two servo-controlled motions: orbital motion and lateral motion. The orbital motion keeps the X-ray beam centered on the region of interest throughout the 180+° rotation about the horizontal axis of the C-gantry, while the lateral motion allows the C-gantry to rotate 360° about that axis.

The motion control system utilizes a distributed architecture with one motion controller and one servo-node for each axis. The feedback components of orbital and lateral motion axis include a motor Hall sensor for commutation feedback, a motor encoder for velocity feedback, a secondary incremental encoder for position feedback, and a potentiometer for indicating a reference absolute position. Position Velocity Time (PVT) mode is applied for the orbital scan, which is designed specifically to give a more direct control over the trajectory profile. The PVT mode is essentially a cubic position profile defined by the position and velocity at the start and end of a segment.

To acquire the position and alignment data in 3-D space, a special motion trajectory is considered, in which a full orbital scan is performed at the interval of 5-20 degree interval of lateral axis. To illustrate the trajectory, the orbital and lateral motions are mapped into a 2-D planar map, which is shown in Figure 5.

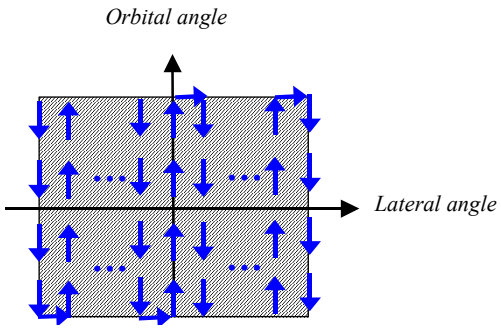


Figure 5. 2-D map of workspace without constraints.

IV. Surgical Navigation with EM Trackers

Surgical Electromagnetic (EM) navigation systems allow for virtual display of the surgical tool superimposed on 2-D or 3-D images of the patient anatomy of interest. Surgical navigation facilitates minimally invasive surgical procedures by allowing for controlled and precise manipulation of the surgical tools without requiring additional fluoroscopic visual control thus reducing the X-ray dose to the patient and staff. An EM-based surgical navigation system comprises the electromagnetic tracker, an EM transmitter rigidly attached to the patient's anatomy, an EM receiver attached to the surgical tool, and an EM receiver affixed to the imager. The key property of the system is the ability to render the instrument tip and trajectory in the image coordinate system (2-D image or 3-D reconstruction volume) via registration of the tracker coordinate system and the image coordinate system. Such registration is computed through a sequence of transforms [6].

The high C-gantry position allows for seamless incorporation of the surgical navigation system. Traditional C-arms, due to mechanical instability factors, cannot guarantee accurate positional repeatability and require dynamic calibration of each acquired image, which is performed using a calibration target attached to the input screen of the detector. The new mechanical design of the C-Arm has low non-repeatable variability and allows for robust off-line 3-D geometric calibration. The C-arm can now be used without the calibration target being permanently attached, increasing the surgical volume and improving the workflow and image quality.

The navigation error, perceived by the operator as the distance between the navigated surgical instrument tip in the image and the true instrument position in the image, is one of the key elements of the system. In this paper we consider the contribution of the imaging system calibration error to the overall navigation error in 2-D and 3-D navigation. Our goal is to determine the upper bound of this error under realistic conditions and thereby validate the intrinsic and extrinsic calibration algorithms presented earlier in this paper for the purpose of navigation. In this work, we utilize a numerical modeling approach. We construct the simulation framework that allows for realistic modeling of the error contribution under consideration. We take into account the following sources of error:

1. Uncertainty in the calibration marker shadows detected in the images. This uncertainty depends on the utilized image processing algorithms, marker size, and detector pixel size.
2. Uncertainty in the calibration marker mechanical locations.
3. Calibration phantom flexing. The magnitude of this effect depends on the phantom mechanical design and is modeled using a finite element approach.
4. Error due to finite elevation of the calibration phantom above the image plane and uncertainty in the knowledge of this elevation. This effect is modeled using mechanical detector design parameters.
5. Errors due to non-repeatable variations of the C-arm intrinsic and extrinsic parameters, modeled using C-arm prototype alignment and motion characteristics.

We compute the difference between the virtual point in the surgical volume in the x-ray image and simulated projection of this virtual point using calibrated camera parameters. In the 3-D case, this error is computed in the reconstruction coordinate system. In order to produce the final error estimates, we repeatedly inject random errors into the calibration algorithm inputs (marker locations, C-arm intrinsic and extrinsic parameters, etc.), and measure the average errors over a pre-specified set of virtual navigation points uniformly covering the surgical volume. We measure the errors in terms of Root Mean Square errors (RMS) over the surgical volume, and over the large set of experiment realizations.

Our simulations show that the 2-D navigation RMS errors introduced by the imaging chain are expected to be below 0.5 mm, which is satisfactory from the point of view of the total error budget. The 3-D navigation errors depend on the number of gantry positions used to acquire (register) the transform C in the process of acquiring projections. The navigation error generally decreases with the number of registered C-arm orbital positions, but depends on what specific position is used. In addition, the 3-D navigation errors depend upon the algorithm used to compute the final navigated point in the reconstruction coordinate system due significantly non-isotropic error covariance structure. A special algorithm (outside the scope of this paper) is used to produce the results reported here. Resulting simulated navigation errors for a different number of registered gantry positions are summarized in Table 1.

Number of registered gantry positions	1	2	160
RMS navigation error, [mm]	1.2	0.4	0.2

Table 1. 3-D error simulation results summary.

For two registered gantry positions, the navigation error is below 0.5 mm, in line with the error achieved in 2-D case.

V. CONCLUSIONS

Sequential calibration of the C-arm intrinsic and extrinsic geometries show highly accurate results and efficient performance. Overall maximum non-repeatability of the mechanical platform on the experimental prototype system has been observed to be less than 0.6 mm in 3-D space. By increasing the positioning accuracy of the C-gantry, high-resolution C-arm CT imaging (voxel size smaller than a half millimeter) as well as more confident 2-D and 3-D navigation in interventional and minimally invasive surgery becomes feasible.

ACKNOWLEDGMENT

The authors are thankful to Sebastien Breham, Douglas Reynolds, Zhonghua Wang, Johnny Barrett, and Tyson Baker for assistance in C-gantry alignment and motion measurements. We appreciate programming efforts done by Michael Stipmson, Jeffrey Moellmer, and Wayne Christensen. We are also grateful to Andrew Litvin, Ali Hamadeh, and Daniel Beaudet for working with us on positioning concepts and simulations for integrated EM tracking of surgical tools.

REFERENCES

[1] GE Healthcare–Surgery. About the OEC 9900 Elite, http://www.gehealthcare.com/euen/surgery/products/oec-9900-elite/index.html

[2] Cheryauka A., Breham, S. and W. Christensen, 2006, "Sequential Intrinsic and Extrinsic Geometry Calibration in Fluoro CT Imaging with a Mobile C-arm", *Proceedings of SPIE Medical Imaging*, vol. 6141**,** MI 6141-90 paper.

[3] FARO Technologies, Inc., http://www.faro.com/content.aspx?ct=us&content=pro&item=2

[4] Gorges, S., Kerrien, E., Berger,M-O, Trousset, Y., Pescatore, J., 2005, "An effective technique for calibrating the intrinsic parameters of a vascular C-arm from a planar object", *Proceedings of SPIE Medical Imaging*, MI 01 paper.

[5] Roger Y. Tsai, 1987, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses", *IEEE Journal of Robotics and Automation*, Vol. RA-3, No 4.

[6] Cheryauka A., Tubbs, D., Langille, V., Kalya, P., Smith B., Cherone, R., 2008, CT Imaging with a Mobile C-arm Prototype. *Proceedings of SPIE Medical Imaging*, vol. 6141**,** MI 6913-198 paper.

Accelerated Fluoroscopy and CT on a Mobile C-arm

Larry Anderton, Todd Brown, Arvi Cheryauka, Vinton Langille, and Alex Tokhtuev

***Abstract*– Mobile X-ray imagery is an omnipresent tool in conventional musculoskeletal and soft tissue applications. The next generation of mobile C-arm systems can provide clinicians of minimally-invasive surgery and pain management procedures with both real-time high-resolution fluoroscopy and C-arm CT imaging modalities. In this study, we consider real-time fluoroscopic and CT experimental system configurations and evaluate their imaging capabilities. To streamline tuning of the X-ray fluoroscopic and CT IP chains, improve effectiveness of observer studies, and support research efforts, we discuss Interactive Fluoroscopic Imaging Chain Simulator and CT Exerciser. A heterogeneous computing and visualization platform has been tested to simulate real-time fluoroscopy on a mobile C-arm. The Interactive GPU-Accelerated IP Chain Simulator, IP9900, was used for tuning anatomical profiles to address clinical requests and challenging cases. We consider two experimental C-arm CT setups: non-destructive evaluation configuration with a rotating stage and medical imaging configuration with a C gantry moving around the patient and the table. We connect the participating devices through a Mobile X-Ray Imaging Environment known as MOXIE. Anthropomorphic phantom volume renderings and orthogonal slices of reconstructed images are displayed. The experimental results show CT-like image quality that may be suitable for interventional procedures, and near real-time data management, and, therefore, have great potential for effective use on the clinical floor.**

I. Introduction

Mobile X-ray may be seen as the surgeon's eyes for imaging and navigating in interventional and minimally-invasive operational procedures [1]. The next generation of C-arm systems can provide clinicians with both real-time high-resolution fluoroscopy and in-room C-arm CT imaging capabilities [7-8]. In previous studies, we have discussed calibration of geometry, image processing algorithms, image registration, and image-guidance strategies [3-4]. In this study we investigate the imaging capabilities of the mobile C-arm prototype, enhanced with commodity hardware devices and in-house research software. The anthropomorphic phantom image reconstruction results are provided.

II. Fluoroscopy

Fluoroscopy applications may not be as computationally intensive as computationally loaded image registration, CT reconstruction, or volume rendering, but they impose a different set of constraints on image processing implementation. X-ray image detectors produce a video stream that must be processed without skipped frames or noticeable lag. Until recently, good fluoroscopic image quality was only attainable on highly specialized proprietary image processing hardware, which cannot be easily reprogrammed for acceleration of other computing tasks.

A fluoroscopy image processing chain is simulated for benchmarking purposes by executing following operations commonly found in modern C-Arms:

- Frame transfer into image processing device memory.
- Temporal averaging to reduce the amount of quantum noise.
- Spatial averaging to replace or supplement temporal averaging when motion-blurring artifacts become noticeable.
- Additional spatial filtration for algorithms such as edge detection or other image enhancements.
- Histogram accumulation for calculation of image metrics that can be used to make other operations adaptive.
- Grayscale conversion that can be used for gamma correction or histogram equalization.

Temporal averaging is implemented in a typical recursive fashion by adding the currently acquired frame F_a with the previously averaged frame F_p using appropriate weights α and $(1 - \alpha)$, so that the current temporal average F_c is the superposition of the new and previous frames:

$$F_c = F_a \cdot (1-\alpha) + F_p \cdot \alpha. \tag{1}$$

Spatial averaging and filtration are implemented by the application of convolution kernels, of various symmetric geometries, across an image frame. The general solution to a 2D convolution, g_{con}, is given as:

$$g_{con}(x_1, x_2) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} f_1(x_1, x_2) \cdot f_2(x_1 - k, x_2 - l). \tag{2}$$

Most practical convolutions are local and can be implemented for 1024x1024 images with kernels measuring less than 16 pixels in size, thus greatly simplifying calculations. Some widely used 2-D kernels, such as Gaussian blur, are also separable and can be replaced with the application of two 1-D kernels, further reducing computational load.

Manuscript received February 1, 2010.

Larry Anderton, Todd Brown, Arvi Cheryauka, Vinton Langille, and Alex Tokhtuev are with GE Healthcare-Surgery, Salt Lake City, UT 84116 USA.

Contact e-mail: Arvi.Cheryauka@med.ge.com.

III. Computer Tomography

The Modified Feldkamp-Davis-Kress (FDK) CBCT reconstruction algorithm is used for a quasi-circular scanning trajectory. The backprojection step is the most resource-consuming portion of the algorithm. In this study we use the same approach as in the earlier work, but implement it in different computing environment [3].

IV. Simulation Framework

Our methodology is to create a friendly framework utilizing a common workstation, to build tools to accelerate image analysis, and to improve user interactions. A heterogeneous computing and visualization platform has been tested and deployed to simulate real-time fluoroscopy on a mobile C-arm. Hardware components include an off-the-shelf CPU and GPU (Fig.1).

Fig. 1. Nvidia's Quadro FX 5800 comprising 240 thread processors and 4 Gb onboard graphics memory [6].

Figure 2 is a screenshot of the IP9900 Simulator. The software environment comprises a 64-bit Linux, GCC compiler, CUDA compiler, and MATLAB integrated toolkit. All image processing is performed using single floating-point operations. The program emulates the image processing implemented on Mobile X-ray C-Arm [1].

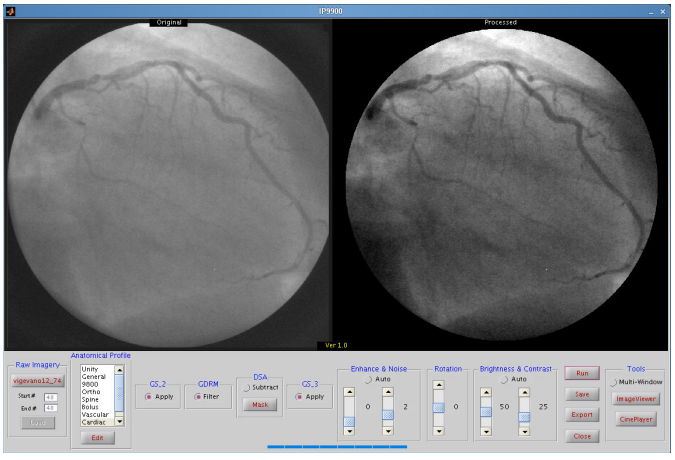


Fig. 2. Screenshot of the IP9900 Simulator

Algorithmic blocks include grayscale conversion, dynamic range management, digital subtraction, noise filtering, edge enhancement, image rotation, and brightness / contrast handling. Anatomical profiles, as they are described in [1], are prime tools to manage image processing and display. By using the Simulator we adjust the settings in the anatomical profiles to address clinical requests and challenging cases [2].

Figure 3 is a screenshot of the C-arm CT Exerciser. The graphics user interface of 2-D fan-beam tool is demonstrated.

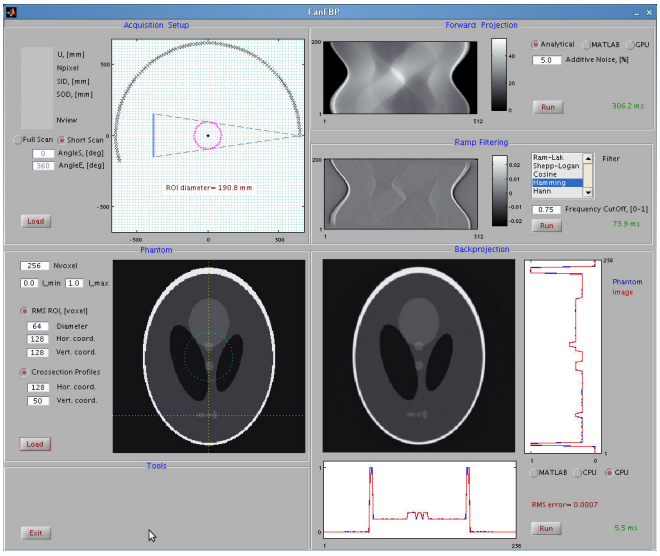


Fig. 3. CT Exerciser for fan-beam geometry.

The 2-D and 3-D CT simulation tools work in the filtered backprojection framework described in the "Computer Tomography" section. The forward and back projection algorithms are implemented in Matlab, C, and CUDA. The performance is bit-accurate. GPU-accelerated execution is more than two orders of magnitude faster than execution of Matlab scripts.

V. System Setup

To research imaging capabilities of mobile X-ray in C-arm CT configurations, we built two experimental systems. Our intention is to utilize standard C-arm industry components and evaluate image quality with correspondent feature performance estimates for systems that have various degrees of complexity but share the same hardware and software. This allows us to focus on specific subsystem blocks, evaluate their subsequent impacts independently, and streamline the experimental studies. Fostering this modular feasibility leads to a shorter examination period, more transparent management of the resources and, eventually, to more predictable outcomes. In the previous study we simulated the system's image chain and analyzed its characteristics [3-4]. The real world has many challenges and constraints that the simulated world does not have. Therefore, by substituting simulation blocks in the image chain with real devices and processes, the real world system can be gradually, thoroughly, and cost-effectively prototyped.

II.1. Electro-Mechanical Environment

Two setups are used comprising a mobile C-arm mainframe prototype and a mobile rotating table depicted in Figure 4.

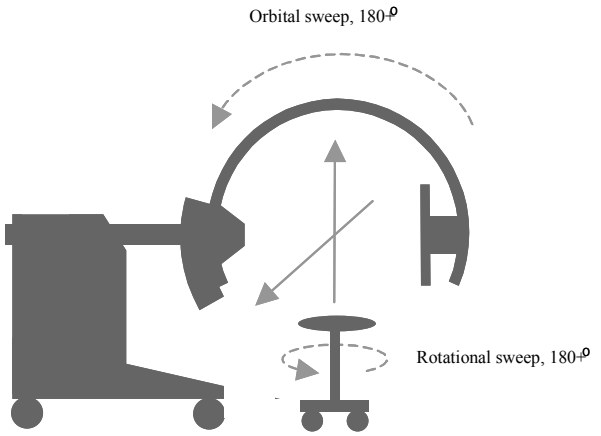


Fig. 4. C-arm orbital and turntable rotational scans.

The first system setup includes a stationary X-ray acquisition system and an object of interest placed on a rotating stage. A similar Non-Destructive Evaluation (NDE) configuration is used in the industrial CT, where projection geometry can be established and measured very accurately. Rotational sweep of an isometric object is an easily controllable process. Medical imaging CT protocol requires rotation of the X-ray acquisition assembly. Due to the current electro-mechanical design, actual motion of the C-arm gantry can deviate from the command motion in a systematic and a random fashion.

II.2. Software Environment

The participating devices are connected through the Mobile X-Ray Imaging Environment, MOXIE, which is a set of software applications for internal research in our facility (Fig. 5). The MOXIE suite consists of three primary applications: the MOXIE client that runs on a Windows computer; the ImageViewer that runs on one or more Windows computers; and the Test Framework that runs on the Real-Time OS computer located on the C-arm workstation prototype. Each MOXIE module communicates with the others over a 100Mbit Ethernet network. MOXIE supports up to four ImageViewer applications running on separate Windows desktop systems.

The MOXIE client is the program interface where user specifies parameters of X-ray generation and detection, as well as manages data transfer. The calibration procedures of the image chain consist of gain equalization, offset correction, and bad pixel mapping. The screenshot of the test image is demonstrated in Figure 6.

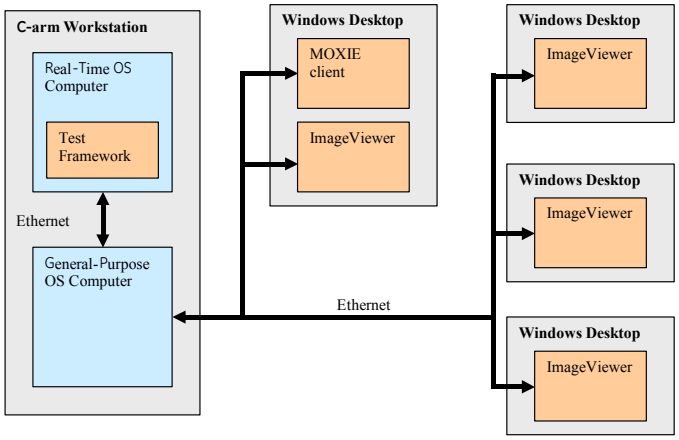


Fig. 5. Mobile X-ray Imaging Environment (MOXIE).

The ImageViewer application is used to display imagery being received from the C-arm workstation prototype (Fig. 6). ImageViewer provides the ability to adjust the brightness and contrast of the image, and has a number of magnification modes so the user can zoom in on an area of interest. Of particular interest is ImageViewer's ability to display the position and value of pixels in the X-ray image.

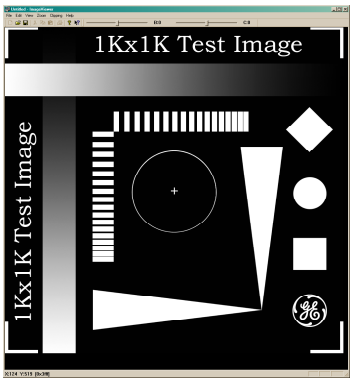


Fig. 6. Screenshot of the ImageViewer application.

The MOXIE and ImageViewer applications are driven by network connections, so once launched they will monitor the network, waiting for commands to process. It is the Test Framework, running on the Real-Time OS computer that will initiate communications with the hardware subsystems and issue commands. When scan data is transferred, hardware-accelerated image processing starts on a dedicated remote computer.

VI. Imaging with a Rotating Table

As it was outlined in the System Setup section, a scan with rotational sweep in the horizontal plane is functionally equivalent to one with orbital sweep in the vertical plane. An

anthropomorphic phantom is placed on the rotating stage between the detector and the x-ray tube. With the C-arm gantry stationary while the tabletop rotates based on a programmable velocity profile. This type of setup minimizes non-repeatable mechanical errors related to the positioner. Intrinsic and extrinsic geometries are calibrated prior to the scan. X-ray exposure rate, detector readout rate, and turntable motion are synchronized as well. In Figure 7, we demonstrate the imaging setup with the use of the stationary C-arm, rotating stage, and hand phantom.

Figure 7. The Hand phantom.

In Figure 8, we show the imaging results presented in the form of three orthogonal slices and volume rendering.

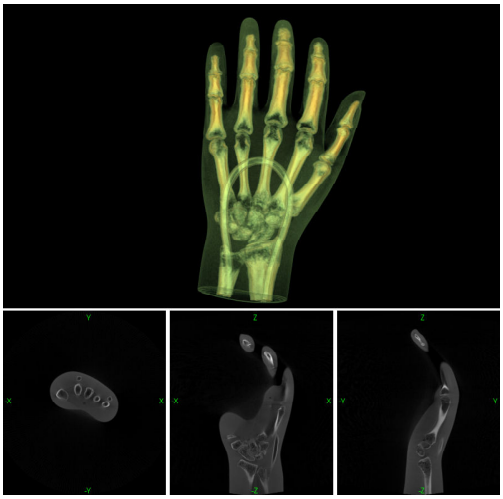


Fig. 8. 3-D reconstructed image of the Hand phantom.

CBCT acquisitions are done at half detector resolution with an effective pixel pitch of 0.4 mm and at the standard frame rates of 7.5 and 15 frames per second. The turntable rotates at a velocity of 12 degrees per second. A typical reconstruction from a scan of 225 binned images and of a volume dimension of 512^3 takes less than 15 seconds on Quadro FX 5800.

VII. Imaging with Rotating C-Gantry

Image reconstruction from data acquired with rotating X-ray Tube–Detector assembly relies on accurate geometrical positioning of imaging devices with respect to the patient anatomy. 3-D geometry calibration concepts, methods, and findings are summarized in the separate paper [5]. Continuation of feasibility research with moving C-arm gantry is in progress, and, in the foreseen future we plan to obtain the new imaging results.

Conclusions

Our investigations highlight the X-ray fluoroscopy and CT research through the use of the C-arm prototypes, in-house experimental developments, and off-the-shelf hardware/software components. The Interactive GPU IP Chain Simulator and CT Exerciser are found to be helpful in accelerating image evaluation and prototyping of new imaging solutions. Parallelization of computationally demanding tasks using high-performance on-chip devices could be a key to a practical and cost-effective mobile C-arm imaging. The NDE-type and medical imaging cone-beam X-ray prototypes are constructed to test the system capabilities. The results obtained with the anthropomorphic phantoms show CT-like image quality that may be suitable for interventional procedures, near real-time data management, and, therefore, have great potential for effective use on the clinical floor.

Acknowledgments

This research work is sponsored by GE Healthcare – Surgery. The authors appreciate support from our colleagues: L. Powell, J. Hamblin, M. James, R. Purcell, D. Reynolds, S. Brown, D. Riabkov, D. Tubbs, F. Brown, J. Barrett, S. Curtis, C. Atria, D. Ferguson, S. Maesato, R. Hauck, V. Jensen, D. Barker, F. Falco, J. Snyder, and D. Ball.

References

[1] GE Healthcare–Surgery. About the OEC 9900 Elite, http://www.gehealthcare.com/euen/surgery/products/oec-9900-elite/index.html

[2] Brown, T., Cheryauka, A., and L. Anderton, 2009, GPU-Accelerated IP Chain Simulator for Mobile X-Ray Imaging, *Proc. of HPMI workshop.*

[3] Riabkov D., Cheryauka A., Tokhtuev A., and T. Brown, 2008, Hardware Accelerated C-arm Fluoroscopy and CT: A Pilot Study, *Proc. of SPIE Medical Imaging*, vol. 6141**,** MI 6141-90.

[4] Cheryauka A., Tubbs, D., Langille, V., Kalya, P., Smith B., Cherone, R., 2008, CT Imaging with a Mobile C-arm Prototype. *Proceedings of SPIE Medical Imaging*, vol. 6141**,** MI 6913-198.

[5] Kalya, P., Cherone, R., and A. Cheryauka, 2010, 3-D Geometry Calibration on a Mobile X-Ray C-arm for CT Imaging and Surgical Navigation. *Submitted to CT Meeting*.

[6] NVIDIA Quadro Professional Solutions. http://www.nvidia.com/page/workstation.html

[7] Ritter D., Ritter, J. Orman J., C. Schmidgunst C., and R. Graumann. 2007, "3D Soft Tissue Imaging with a Mobile C-arm". *Computerized Medical Imaging and Graphics*, 31, 91-102.

[8] Siewerdsen J.H., M. J. Daly, G. Bachar, D. J. Moseley, G. Bootsma, K. K. Brock, S. Ansell, G. A. Wilson, S. Chhabra, D. A. Jaffray, and J. C. Irish. 2007, "Multi-mode C-arm Fluoroscopy, Tomosynthesis, and Cone-Beam CT for Image-Guided Interventions: From Proof of Principle to Patient Protocols", *Proc. of SPIE Medical Imaging*, 6510, 65101A-1.

Enhancement of CT Image Quality via Bilateral Filtering of Projections

Anna Gabiger ♠, Robert Weigel ♠, Steven Oeckl ♣, Peter Schmitt ♣
♠ Institute for Electronics Engineering, University of Erlangen-Nuremberg, Erlangen, Germany
♣ Fraunhofer Institute for Integrated Circuits, Erlangen, Germany
Email: Gabiger@lfte.de, Weigel@lfte.de, Steven.Oeckl@iis.fraunhofer.de, Peter.Schmitt@iis.fraunhofer.de

***Abstract*—In this paper the way to enhance the image quality of 3D-CT reconstruction by means of noise filtering in 2D projections is shown. The noise reduction is done by applying a bilateral filter, which is well known for its edge-preserving quality. During the measurement of the projections the noise signal of the system is superimposed to the image signal. This noise propagates into the reconstruction and can be measured in the cross sections of the reconstructed volume. In order to reduce the noise in the volume, the preprocessing of the projections is extended by noise filtering prior to the logarithm transform and the ramp filtering. The capability of the bilateral filter to reduce image noise without blurring or corruption of the image contents combined with the real-time implementation on FPGA makes this filter suitable for the application to the x-ray projections. As it is shown in this paper, not only the noise is suppressed in the reconstruction, but due to the edge-preserving filtering of noise in the projections the structural details in the reconstruction are resolved better.**

I. Introduction

The aim of noise filtering in 2D projections is the enhancement of the image quality of 3D-CT reconstruction. Thereby the emphasis is set to the real-time processing of the projections, so that there is no negative influence on the reconstruction time. The capability of real-time processing is reached through implementation of the bilateral filter on FPGA. The FPGA design has passed the according tests in our previous work [4].

The noise signal of the system is superimposed to the image signal during the acquisition of the projections. This noise propagates into the reconstruction and corrupts the image contents of the cross sections of the reconstructed volume. In order to lessen the image quality degradation in the volume, the preprocessing of the projections is extended by noise filtering prior to the logarithm transform and before the ramp filtering. In this work the reconstruction is done with the Feldkamp-Davis-Kress FDK algorithm, for the ramp filtering the Shepp-Logan filter is used.

The noise signal is considered as a random variable with Poisson statistics, which converges towards a random variable with Gaussian statistics for a large number of detected quanta. The bilateral filter, first presented in [1], can be adjusted to the noise statistics. Due to the possibility of adapting this filter to the noise statistics it is a well known noise reduction method in computed tomography. It has been used for both sinogram smoothing [2] and image space noise reduction [3]. In this paper the processing results of raw projection data are shown.

The capability of the bilateral filter to reduce image noise without blurring the x-ray projections combined with the fast implementation on FPGA makes this filter suitable for the application in a running computer tomograph. Without disturbing the data flow, the noise in the projections can be reduced exceedingly, which implicates the enhancement of the image quality of the 3D reconstruction. As it is shown in this paper, due to the edge-preserving filtering of the noise in the projections the structural details in the reconstruction are resolved better.

In the following section the theory of the bilateral filter is briefly described. For the application of the noise filter the Gaussian noise is simulated and added to the projections. In section III the criteria for the image quality evaluation are described. The reconstructions from noiseless, noise affected and filtered projections are compared by means of the calculation of the noise standard deviation followed by the comparison of the standard deviation to mean ratio. These results are shown in section IV. Additionally the comparison of bilateral filter and median filter is presented in section IV. Section V contains the conclusion.

II. Bilateral Filter

The denotation *bilateral filtering* suggests the combination of domain and range filtering. The domain filter averages the nearby pixel values and acts thereby as a low-pass filter. The range filter stands for the nonlinear component and plays an important part in edge preserving. This component averages only the similar pixel values regardless of their position in the filter window. In case the value of a pixel in the filter window diverges from the value of the pixel in the center of the window by a certain amount, the pixel is skipped. The shift-variant filtering operation of the bilateral filter considering Gaussian noise is given by:

$$\bar{\phi}(\mathbf{x}) = \frac{1}{k(\mathbf{x})} \sum_{\mathbf{m} \in F} \phi(\mathbf{m}) \cdot s(\phi(\mathbf{x}), \phi(\mathbf{m})) \cdot c(\mathbf{x}, \mathbf{m}) \,. \qquad (1)$$

The expression $\mathbf{m} = (m, n)$ denotes the pixel coordinates in the image to be filtered, $\mathbf{x} = (x, y)$ stands for the corresponding coordinates in the filtered image. With these notations $\bar{\phi}(\mathbf{x})$ means the gray value of the filtered pixel and $\phi(\mathbf{m})$ identifies the gray value of the spatially neighbouring pixels in the filter window F.

The following expressions (2) and (3) describe the photometric and the geometric components $s(\phi(\mathbf{x}), \phi(\mathbf{m}))$ and $c(\mathbf{x}, \mathbf{m})$ respectively.

$$s(\phi(\mathbf{x}), \phi(\mathbf{m})) = \exp\left(-\frac{1}{2}\left(\frac{\|\phi(\mathbf{x}) - \phi(\mathbf{m})\|}{\sigma_{ph}}\right)^2\right), \quad (2)$$

$$c(\mathbf{x}, \mathbf{m}) = \exp\left(-\frac{1}{2}\left(\frac{\|\mathbf{x} - \mathbf{m}\|}{\sigma_c}\right)^2\right). \quad (3)$$

The photometric component compares the gray value of the center pixel with the gray values of the spatial neighbourhood and computes the weight coefficients depending on the factor σ_{ph}. This factor adjusts the photometric component. The similar gray values are averaged with higher weights than dissimilar values in order to avoid any significant change of the initial gray value of the pixel to be filtered.

The domain filter $c(\mathbf{x}, \mathbf{m})$ acts as a standard low-pass filter in this case, the weights of which are reciprocally proportional to the spatial distance of the center pixel to the neighbourhood. The factor σ_c is selected in a way that blurring caused by low-pass filtering is acceptable.

After normalization with

$$k(\mathbf{x}) = \sum_{\mathbf{m} \in F} s(\phi(\mathbf{x}), \phi(\mathbf{m})) \cdot c(\mathbf{x}, \mathbf{m}) \quad (4)$$

the range of the filtered images does not change significantly due to the filtering.

The detailed description of the FPGA design of the bilateral filter and test results can be found in [4].

III. Image quality evaluation

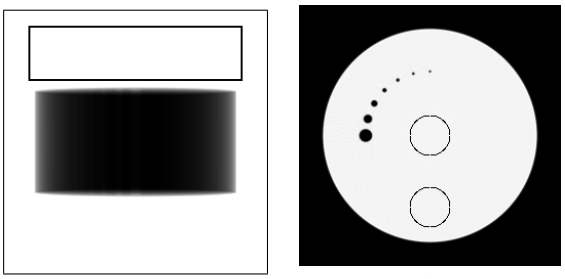
In Fig. 1 (b) one cross section of the volume reconstructed from noiseless projections is shown. As this work is attended to industrial application area the object under test is a carbon fiber-reinforced plastic disk with 10 cm diameter. The end-to-end holes are positioned at a distance of 3 cm from the centre of the disk and have the following diameters: 0.6, 0.4, 0.3, 0.2, 0.16, 0.12, 0.1 cm. The angle between the holes is 15 degrees in each case.

For the evaluation of the image quality of the projections the well known Peak Signal to Noise Ratio PSNR is used. PSNR is defined as follows:

$$\text{MSE} = \frac{1}{MN} \sum_M \sum_N \left[\phi(m,n) - \tilde{\phi}(m,n)\right]^2, \quad (5)$$

$$\text{PSNR [dB]} = 20 \cdot \log_{10}\left(\frac{\phi_{\max}}{\sqrt{\text{MSE}}}\right), \quad (6)$$

where MSE denotes the mean squared error. The M×N image with gray values $\phi(m,n)$ provides the reference for the measurement of MSE. The gray values $\tilde{\phi}(m,n)$ originate from the image to be compared. $\phi_{\max}$ denotes the maximum gray value depending on the word length after the digitalization of the images. Considering the quality of the noise filter



(a) (b)

Fig. 1. (a) Noiseless projection of the carbon disk with the measurement area of the noise statistics in the projections, (b) σ_n measurement areas in the cross section, reconstruction from the noiseless projections

PSNR describes the capability of the filter to suppress noise regardless of the perceived visual quality of the filtered image.

In order to compare the cross sections of the 3D reconstructions the standard deviation of the noise is measured. In Fig. 1 (b) the areas are pointed out where the standard deviation of the noise in the cross section of the volume is measured. As the image noise in the reconstruction varies depending on the distance from the centre, there are two measurement areas. The formula for the calculation of the standard deviation is

$$\sigma_n = \frac{1}{L}\sqrt{\sum_L (\phi_l - \mu_\phi)^2}, \quad (7)$$

where L equals the number of the pixels in the measurement area and μ_ϕ indicates the mean gray value within the measurement area.

IV. Results

For the following tests a projection simulation tool is used, which generates projections with 16 bits. The resolution of the projections is 256×256 pixels, the pixel size is equivalent to 0.8 mm. The x-ray source is simulated as a 100 kV monoenergetic source. No scatter effects are simulated. Over the range of 360 degrees 400 projections are simulated. The reconstructed volume shows 255×255×230 voxels.

In Fig. 1 (a) the noiseless projection of the test object is depicted. For the simulation of the noise signal Gaussian noise with a standard deviation of $\sigma_n = 0.0076$ equivalent to 500 gray values at the digitalization of 16 bits is added to each projection. In case the gray value of a noise affected pixel exceeds the range of 16 bit values [0, 65535], the gray value is clipped. Subsequently each projection is filtered with the bilateral filter with the following settings: $\sigma_{ph} = 5\sigma_n$ and $\sigma_c = 1$.

In the first row of Fig. 2 the noise affected projection and the projection filtered with bilateral filter are shown. In the second row there are according noise statistics calculated for the measurement area shown in Fig. 1 (a). The histograms confirm the reduction of the noise standard deviation after noise filtering by a factor of three. The PSNR of the noise affected projection equals 46.9 dB. After the noise reduction with the bilateral filter the PSNR increases to 51.0 dB. In this case the PSNR is measured across the whole image.

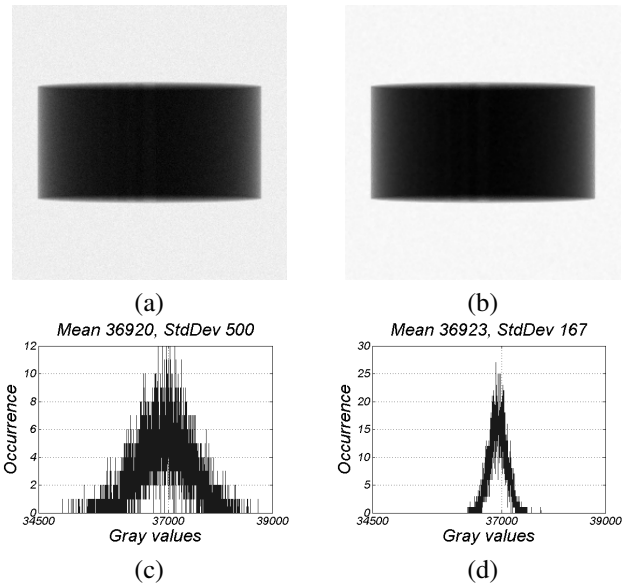


Fig. 2. (a) Noise affected projection, (b) Projection filtered with bilateral filter, (c) Statistics of the gray values in the noise affected projection, (d) Statistics of the gray values in the projection filtered with bilateral filter

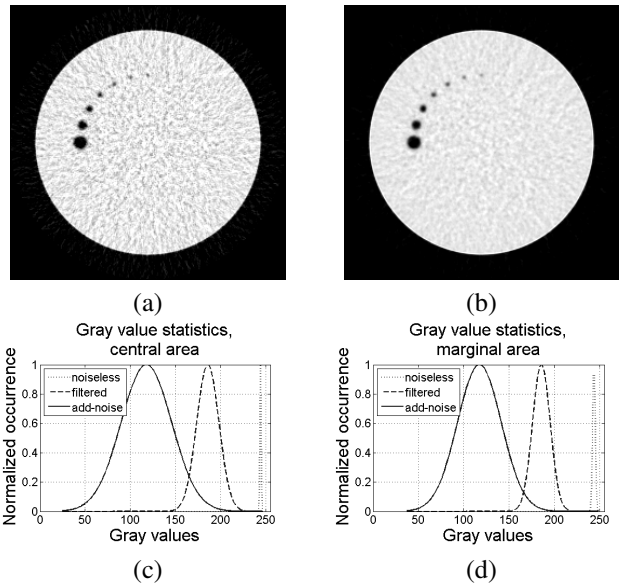


Fig. 3. (a) Cross section of the reconstruction from the noise affected projections, (b) Cross section of the reconstruction from the projections filtered with bilateral filter, (c) Envelopes of the gray value statistics in the central area of the cross sections, (d) Envelopes of the gray value statistics in the marginal area of the cross sections

	μ_ϕ,cent	σ_n,cent	σ_n/μ_ϕ,cent
noiseless	232	0.6	0.0026
add-noise	230	23	0.10
bilateral 5×5	233	11	0.05
median 3×3	232	14	0.06
median 5×5	233	8	0.03

TABLE I
MEASUREMENT RESULTS IN THE CENTRE OF THE CROSS SECTION

	PSNR, dB
add-noise	24.39
bilateral 5×5	30.95
verification	30.84
median 3×3	28.74
median 5×5	31.80

TABLE III
PSNR OF THE CROSS SECTION

In Fig. 3 (a) the cross section of the reconstruction from the noise affected projections can be seen. In Fig. 3 (b) the cross section of the reconstruction from the projections filtered with the bilateral filter is depicted. The reconstruction tool scales the gray values of the voxels to the range of 8 bits. Due to this scaling the following image quality analysis relates to the maximum gray value 255.

The standard deviation of the noise is measured in two areas marked in Fig. 1 (b). The results of this measurements are summarized in the Tables I and II. The second column of these Tables shows the mean gray value of the centered and the marginal area of the cross sections respectively. The third column contains the standard deviation of the noise in the same areas. In both areas the standard deviation of the noise is more than halved by means of the noise filtering of the projections with the bilateral filter. To underline this statement the fourth columns of the Tables I and II show the ratio σ_n/μ_ϕ in both measurement areas respectively. The smaller this ratio the fewer distortion the image shows. It can be seen in the first rows of both Tables, that the standard deviation to mean ratio of the noiseless reconstruction is approximately zero and increases drastically in add-noise reconstruction. After the noise filtering the ratio σ_n/μ_ϕ gets better in both measurement areas of the reconstruction and is only one order of magnitude worse than in the noiseless case. In Fig. 3 (c) and (d) the envelopes of the gray value statistics of both measurement areas are shown. These graphics demonstrate the shift of the mean towards the mean in the noiseless reconstruction on the one hand and the reduction of the standard deviation of the gray values to less than one half on the other hand.

	μ_ϕ,marg	σ_n,marg	σ_n/μ_ϕ,marg
noiseless	231	1.2	0.0052
add-noise	230	20	0.09
bilateral 5×5	232	9	0.04
median 3×3	232	12	0.05
median 5×5	232	7	0.03

TABLE II
MEASUREMENT RESULTS IN THE MARGINAL AREA OF THE CROSS SECTION

For the measurement of PSNR of the cross sections the carbon disk shown in Fig. 3 (a) and (b) are taken. The cross section in Fig. 1 (a) provides the reference. Here the PSNR is measured across the whole cross section. The measurement results are summerized in Table III.

A. Verification

In order to verify the effectiveness of the noise filtering in the projections, different noise data is simulated. The noise

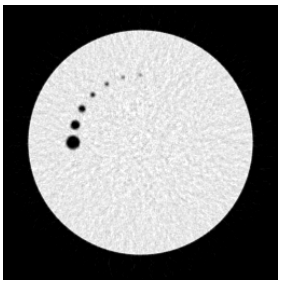
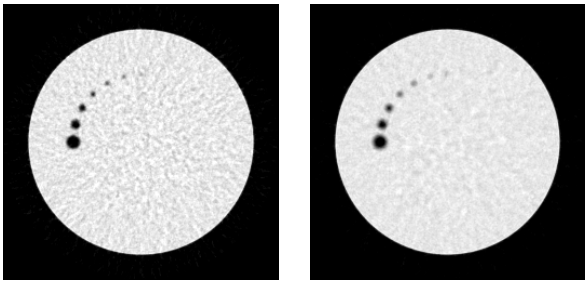


Fig. 4. Cross section of the reconstruction from the noise affected projections with reduced noise standard deviation for the verification



(a) (b)

Fig. 5. (a) Cross section of the reconstruction from the filtered projections with median filter 3×3, (b) Cross section of the reconstruction from the filtered projections with median filter 5×5

standard deviation is chosen in the way, that PSNR of the respective reconstruction reach the PSNR of the reconstruction from the projections filtered with bilateral filter. The cross section of the carbon disk in Fig. 4 exhibits PSNR = 30.84 dB. To achieve these results the noise standard deviation $\sigma_n = 0.003$ equivalent to 196 gray values is simulated. And still the verification results are slightly inferior to the filtering results. Thus via the noise filtering with the bilateral filter the noise in the projections can be reduced to less than one half of the initial value.

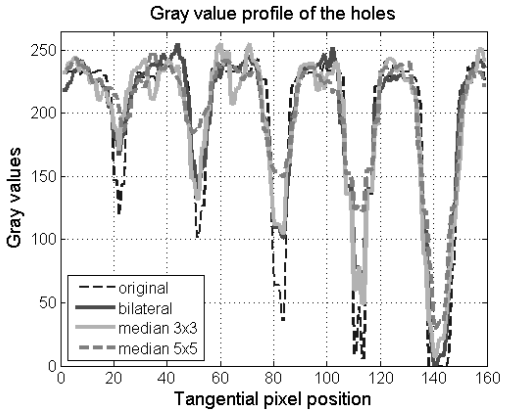


Fig. 6. Gray value profile of the holes in the cross section after the noise filtering

B. Comparison with median filter

The best way to appreciate the performance of the bilateral filter is to compare it to the standard filter, in case of CT to median filter. On the one hand for the sake of fairness the smallest window size 3×3 is chosen for median filter, because the window size 5×5 is the smallest one for bilateral filter. On the other hand it might be interesting to compare both filters with the same window size 5×5. The according cross sections are shown in Fig. 5. The results of the median filtering are shown in Tables I, II and III.

As the bilateral filter preserves details in the image to be filtered, this property propagates into the reconstruction. This is proved by considering the holes in the carbon disk. In Fig. 6 the gray value profile of the holes is depicted. Although the filtering with median filter 5×5 results in the best PSNR, small holes almost disappear in the reconstruction. The smallest hole shows only a third of the initial contrast after filtering with median filter 5×5. The curves representing the gray value profiles for the bilateral filter and the median filter 3×3 are nearly congruent in the hole profile, but they diverge in the homogeneous areas. The increase of PSNR along with the shown contrast resolution demonstrate the suitability of the bilateral filter over the median filter 3×3 and 5×5 for noise filtering in the CT projections.

V. Conclusion

In this work the suitability of the bilateral filter for the noise filtering in the CT projections is shown. The chosen noise standard deviation of 500 gray values is extremely high, but even in this case the application of the bilateral filter for noise suppression in the projections shows satisfying results. The noise standard deviation can be reduced to less than one half. Due to the edge-preserving property of the bilateral filter the quality of the reconstruction is even improved at the same time, compared to the reconstruction from the noise affected projections. In comparison with the median filter the bilateral filter achieves also better results. Using the implementation of the filter on an FPGA [4], real-time processing of the projections is possible, so that there is no negative influence on the reconstruction time.

References

[1] C. Tomasi and P. Manduchi, *Bilateral Filtering for Gray and Color Images*. Proceedings of IEEE Sixth International Conference on Computer Vision, pp.839-846, 1998.

[2] L. Yu, A. Manduca, J. D. Trzasko, N. Khaylova, J. M. Kofler, C. M. McCollough and J. G. Fletcher, *Sinogram Smoothing with Bilateral Filtering for Low-dose CT*. Progress in Biomedical Optocs and Imaging - Proc. of the SPIE, Vol. 6913, 2008.

[3] J. C. R. Giraldo, Z. S. Kelm, L. S. Guimaraes, L. Yu, J. G. Fletcher, B. J. Erickson, and C. M. McCollough, *Comparative Study of Two Image Space Noise Reduction Methods for Computed Tomography: Bilateral Filter and Nonlocal Means*. Proc. of the 31th Annual Conference of the IEEE Engineering in Medicine and Biology Society, pp. 3529-3532, 2009.

[4] A. Gabiger, M. Kube and R. Weigel, *A Synchronous FPGA Design of a Bilateral Filter for Image Processing*. Proc. of the 35th Annual Conference of the IEEE Industrial Electornics Society, pp. 1990-1995, 2009.

Extended Low Contrast Detectability for CT

David P. Rohler[#1], Thomas Toth[#2], Arjun Maniyedath[#3], Steven H. Izen[*4], Thomas Dechant[#5], Michael McNitt-Gray[+6]

[#]*Plexar Associates,*
Cleveland, Ohio, USA.
[1]dprohler@plexar.com
[2]thomas.toth@plexar.com
[3]arjun@plexar.com
[5]dechant@windstream.net

[*]*Department of Mathematics, Case Western Reserve University,*
Cleveland, Ohio, USA.
[4]shi@cwru.edu

[+]*Department of Radiology, UCLA,*
Los Angeles, California, USA.
[6]MMcNittGray@mednet.ucla.edu

Abstract – If dose is to be appropriately managed for CT scanners, then a comprehensive, robust and objective image quality metric is needed to assure that low dose scans provide acceptable quality for the clinical task at hand. Current IQ measurements such as low contrast detectability (LCD), standard deviation, MTF, noise power spectrum, etc. provide technical measures of various specific image characteristics. These individual technical metrics, however, are not suited as a practical image quality gauge for the clinical environment. An image quality metric has been developed to characterize the detectability performance of a scanner over the full range of body sizes and scanner protocols including nonlinear and iterative reconstruction methods.

***Keywords*— Low Contrast Detectability, CT, Image Quality, Contrast Measure, Low Dose**

I. Introduction

While the focus of much of the recent research in CT imaging has been in radiation dose assessment and dose reduction, there has been little research and guidance provided for the assessment of image quality, especially in a clinical diagnostic setting. Obtaining the necessary diagnostic information for the patient is the primary goal of a CT scan. Thus the strategy to reduce dose is to determine the necessary image quality with a relevant quantitative measure and the subsequent dose needed to achieve it. While technologies such as tube current modulation, adaptive collimation and others have been shown to reduce radiation dose, there is little objective evidence that these dose reduction methods have maintained image quality or that they have not affected observer performance in performing diagnostic tasks. Current assessments of image quality (such as low contrast resolution or spatial resolution) are primarily subjectively assessed and while some objective assessment methods have been explored, they are not easily adaptable to clinical use and have not been demonstrated to provide any insight into observer performance. Current assessments are typically performed at only one operating point, with only one phantom size, do not have any way to account for non-linear reconstruction processes and do not provide any information about electronic (system) noise, especially at reduced doses.

In addition, differences in physical performance across scanners and even the lack of standardized descriptions of operating parameters make it difficult to establish uniform imaging protocols in multi-center, multi-scanner trials; this is even more difficult if there is some quantitative imaging measurement (e.g. tumor size, tumor density, etc.). Therefore, the purpose of this research is to fundamentally improve the efficacy of contrast performance measurements for CT scanners and eliminate dependence on human observation. This will have implications for many aspects of both clinical practice and for the use of imaging in clinical trials, especially with respect to standardizing imaging performance across scanners. This may also be able to provide information to guide adjustments for different operating conditions such as patient size and reduced radiation dose levels.

II. Extended Low Contrast detectability (ExLCD)

The ability of a CT system to differentiate a low-contrast object from its background is measured by its low contrast detectability (LCD). LCD is one of the key performance parameters for CT scanners. Currently, the most widely used method for measuring LCD utilizes the low contrast module of the CATphan phantom and a standard head scan protocol to produce a single specification of LCD for the scanner. This method typically measures low contrast performance in only a very limited range of dose and object size parameters. In fact, this method is unable to measure low contrast performance for full-body, low dose protocols. Also, because commercial scanners do not necessarily exhibit quantum noise limited behaviour over their entire range of operation, it cannot be

assumed that the results of the current method can be extrapolated to full-body, low dose protocols.

The low contrast detectability (LCD) performance of a CT system is a critical performance characteristic, providing a measure of the scanner's ability to produce high quality images at the lowest possible x-ray dose. Because it is increasingly important to utilize lower dose protocols in present day CT scanners, it is now critical that LCD be measurable over the entire range of protocols and body sizes.

A. Current LCD Method Definition

The low-contrast resolution of a CT scanner is generally defined as the smallest object that can be visualized at a given contrast level and dose. The contrast level is usually specified as a percentage of the linear attenuation coefficient of water. A sample specification with the current method might be "4 mm at 0.3% contrast for 10mm slice thickness at 30mGy skin dose."

The specification is currently made at a single protocol in one of two ways:

1. Human observation – reconstructed images are viewed by one or more human observers to determine the smallest pin visible;
2. Statistical method – an automated algorithm determines from a flat "water" image which pin size will be visible at 0.3% contrast.

In Table 1, some recently reported measurements from the major CT manufacturers have been collected.

B. ExLCD Contrast Measure

In order to extend the measurement of low contrast detectability, a new contrast measure M is introduced.

$$M = M_0 / (cp). \quad (1)$$

where p is the smallest pin size, measured in millimeters, detectable at contrast level, c, measured in Hounsfield units (HU), where one Hounsfield unit corresponds to 0.1% of water attenuation and M_0 is an arbitrary constant for bringing the measure into a convenient numerical range. In this work, M_0 = 6000 in order to map the best current contrast specification of 2mm at 0.3% to a contrast measure of 1000. For example, the specification from the first row of Table 1, "4 mm at 0.3% contrast for 10mm slice thickness at 10mGy skin dose," would generate a contrast measure of 500,

$$M = 6000 / \big((3)(4)\big) = 500. \quad (2)$$

C. ExLCD Flux Index Definition

Commercial CT scanners typically operate over a wide range of protocols and patient sizes, each of which can have distinct contrast characteristics. Protocol parameters that can impact contrast include (1) scan time, (2) tube current (mA), (3) slice thickness, (4) object (patient) diameter, (5) tube voltage (kVp) and (6) x-ray filter. Also, contrast is significantly impacted by non-linear reconstruction methods as well as the reconstruction pixel size and reconstruction filter.

A relative flux index that incorporates some of these parameters is as follows.

$$FluxIndex = mAs * sliceThick * \frac{e^{-objDiam*attWater}}{e^{-refDiam*attWater}}. \quad (3)$$

For practical combinations of these parameters, the range of FluxIndex is approximately [0.1, 7,000.0]. Refer again to the first row of Table 1 for an example of a current LCD specification: "4mm at 0.3% for 10mm slice at 90mAs." Since this specification relates to the 20 cm CATphan, the FluxIndex would be 900.

TABLE 1: LOW CONTRAST DETECTABILITY SPECIFICATIONS FOR 4 MAJOR SCANNER MANUFACTURERS [8]

Scanner	Contrast	Pin Size	Dose	Slice Thickness	mAs		Contrast Measure	Relative Flux
A	0.3%	4mm	10mGy	10mm	90	Δ	500	900
B	0.3%	5mm	16mGy	10mm	180	Δ	400	1,800
C	0.3%	2mm	40mGy	10mm	350	Δ	1,000	3,500
D	0.3%	5mm	7.3mGy	10mm	105	Δ	400	1,050

III. THE EXLCD METHOD

In order to address the problems with the current LCD methods, the ExLCD contrast performance curve is introduced. It is based on the assumption that equations (1) and (3) apply over the full range of scanning protocols. By scanning a phantom under a selection of protocols, the contrast measure can be determined as a function of relative flux. It can be shown that the contrast performance measurements have a dependence on flux which is directly proportional to the dependence of the signal to noise ratio for the scanner.

D. ExLCD Components

The ExLCD method can be partitioned into four key components: (1) a custom ExLCD phantom containing various contrast/diameter cross-sections, (2) the set of scan protocols and image slices used for ExLCD measurement, (3) the visibility determination, (4) a contrast performance curve generator.

1) Phantom Design: The details of the phantom design will not be discussed here. The design philosophy is initially similar to that of the CATPhan, which has a range of pin sizes at various contrast levels. Differences are the subsets of these pins will have contrast measures uniformly distributed on a log scale and a range of phantom diameters will be used.

2) Scan Protocols: A collection of 20 distinct protocol samples were chosen to be uniformly distributed on the logarithmic relative flux axis

3) Detectability Determination: Detectability determination is the process of analyzing a reconstructed image and determining the smallest pin detectable at a given confidence level for one or more of the contrast sets. The ExLCD analysis method includes one or more of the methods listed

below along with the capability to incorporate true measured contrast.

Single or multiple human observer(s) [1,2]
Statistical method [3]
Rose criterion [4]
Matched filter methods [5,6,7]

4) Curve Generator: As described above, the output of any of the visibility methods applied to an image slice is an ordered pair, [relative flux, contrast measure], corresponding to the smallest pins that are "visible" for any applicable contrast level. The visibility method is applied to images obtained with each sample protocol. The resulting collection of ordered pairs can be plotted on a log-log scale. For example, see Figures 1-2. These plots also illustrate the Contrast Performance Curve generated by a numerical fit to the data.

Sample contrast measurements and the corresponding contrast performance curve are shown in Figure 1.

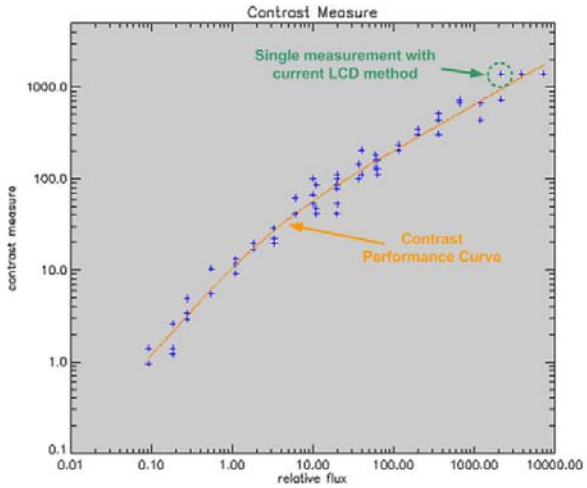


Fig. 1 Blue "+" symbols show contrast measurement values computed by the ExLCD algorithm. Green circle identifies single measurement representing current LCD method. Orange curve illustrates the derived ExLCD Contrast Performance Curve.

E. ExLCD vs. Current LCD Methods

1) Multiple vs. Single Protocol Characterization: A single LCD measurement, as illustrated by the "+" inside the green circle in Figure 1 above, is a very limited characterization of a CT scanner. Extending the demonstration to the comparison of two scanners, the graph, in Figure 2 below, shows ExLCD contrast measurements and Contrast Performance Curves for two different simulated CT scanners. In Figure 2, the contrast measures derived from the data in Table 1 have been located on a graph of contrast measure vs. relative flux. From Figure 2, it's clearly evident that this use of single protocol contrast measurements cannot reliably compare scanners.

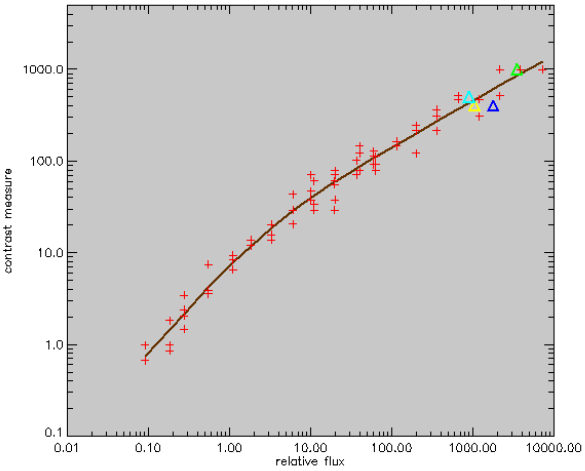


Fig. 2. Low Contrast Detectability specifications for 4 major scanner manufacturers mapped onto an ExLCD graph based on ExLCD Contrast Measure and Relative Flux values.

2) Small Pin Detection: An additional problem is that for conventional detectability methods that are based only on a noise analysis such as the statistical method, noise power spectrum, simple-pixel standard deviation or matched filter standard deviation all can over estimate the performance of a reconstruction process that alters the contrast of the test object. Reconstruction processes that limit spatial bandwidth of both noise and object will not account for changes in the assumed object. For example, assume that a small pin in an LCD test phantom is exactly a cylinder with a 2 mm diameter and a contrast of 0.3 %. If perfectly reconstructed, image pixels within the area of the pin will have an average contrast of 0.3% and all pixels outside this region will be 0%. However the MTF of the system will blur the pin and spread some of its contrast into pixels beyond the original geometric boundary. This results in a reduction in average contrast within the pin region.

3) Human Observer Inconsitency: Human observer variation will produce inconsistencies in the identification of smallest visible pins [11]. This can introduce inaccuracies into the single protocol contrast measurement which are compounded when the contrast performance is extrapolated to other flux ranges.

IV. Numerical Experiments

The key concepts of ExLCD were tested using simulations. In order to assess the value of the ExLCD process, a variety of scanner configurations were simulated. A full-featured fan-beam CT scan simulator was used to generate CT measurement data. The fan-beam simulator included the capability to simulate both quantum and electronic noise. For these experiments, only quantum and electronic noise were varied with the remaining configuration parameters fixed. Seven parameter pairs for quantum and electronic noise were simulated.

For each of seven parameter sets, the CT scan simulator was used to produce raw data for each of 44 applicable phantom cross-sections representing relative flux values in the range [0.1,7000]. Page limitations preclude a complete listing here of the numerical experiments and the display of each of the individual results.

Three detectability methods were applied to the phantom images created by the reconstruction: (1) human observers, (2) Rose and (3) the statistical method. Seven Contrast Performance Curves are produced for each of the detectability methods. The Contrast Performance Curve obtained using a Rose detectability method for a simulated scanner is shown in Figure 3 along with the contrast measures obtained from four human observers.

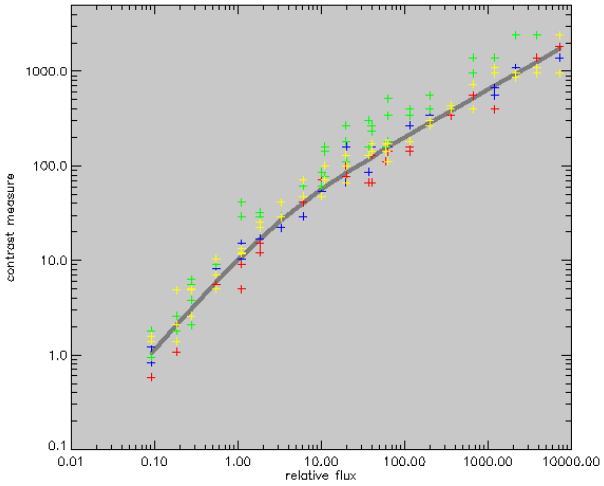


Fig. 3 An ExLCD Contrast Performance Curve for one of the simulated scanners shown with the contrast measures obtained from four human observers. Each color corresponds to an individual observer.

V. Conclusions

The following independent statements are logical conclusions from the work described here:

1) The ExLCD process can successfully characterize the contrast performance of a CT scanner over the entire flux range. It should be adaptable to other radiography applications.

2) A single LCD measurement is inaccurate even in estimating the contrast gain of a scanner in the higher flux regions.

3) A single LCD measurement provides no information about the contrast performance of a scanner in the lower flux regions including (1) body scans at lower dose, (2) scans for a large body, (3) fast scans.

4) With the ExLCD process, human observer visibility determination is much less consistent than either of the automatic methods: statistical and Rose. In fact, observer visibility determination is not accurate enough to differentiate the contrast performance among typical commercial scanners.

5) With the ExLCD process, the Rose method is much closer to the human observer results than the statistical method.

VI. Future Directions

This work is a preliminary exploration of the ExLCD concept. In ongoing work, we will validate and evolve the ExLCD concept. In particular, simulations which include non-linear and iterative reconstruction methods will be performed. The ExLCD concept will be applied to data acquired from real scanners. Finally, the ExLCD detectability predictions will be correlated with clinical image quality studies.

References

[1] Levison M., & Restle, F. Invalid results from the method of constant stimuli. *Perception and Psychophysics*, *4*, 121-12. (1968).

[2] Nicholas Keat, Sue Edyvean, Low contrast detail detectability measurements on multi-slice CT scanners, slide presentation, RSNA2003, http://www.impactscan.org/slides/rsna2003/detaildetectabilityofmsscanners.pdf.

[3] E.H. Chao, T.L. Toth, N.B. Bromberg, E.C. Williams, S.H. Fox, D.A. Carleton, “A Statistical Method of Defining Low Contrast Detectability”, RSNA 2000.

[4] Albert Rose, Vision: Human and Electronic, 1974 Plenum Press.

[5] P. Judy and D. Chakraborty, CT Noise Metrology Symposium, Medical Physics Vol. 31 No. 6, June 2004.

[6] D. Chakraborty and M. Eckert, Quantitative versus subjective evaluation of mammography accreditation phantom images, Medical Physics 22 (2), February 1995.

[7] R. Gagne, B. Gallas, and K. Myers, Toward objective and quantitative evaluation of imaging systems using images of phantoms, Med. Phys. 33 (1), January 2006.

[8] NHS Purchasing and Supply Agency, CEP0807, Buyer’s Guide: Multi-slice CT Scanners, 2009.

[9] T. L. Toth, Z. Ge, M. Daly, “The Influence Patient Size, and Patient Centering on CT Dose and Noise”, Medical Physics, Vol 34, No. 7, July 2007 (3093:3101).

[10] Medical Imaging - The Assessment of Image Quality, The International Commission on Radiation Units and Measurements, Report 54, 1996.

[11] M Brown, D Zhang, T Branham, D Cody, J Kofler, C McCollough, D Pfeiffer, K Strauss, L Yu, and M McNitt-Gray. The Use of a Simple Contrast to Noise Ratio (CNR) Metric to Predict Low Contrast Resolution Performance in CT," Medical Physics 2009, 36(6) p. 2451.

Automated Assessment of Low Contrast Sensitivity for Computed Tomography (CT) Systems Using a Model Observer

W. J. H. Veldkamp[1], I. Hernandez-Giron[2, 3, 4], A. Calzado[2], J. Geleijns[1] ,
(1) Radiology Department, Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, The Netherlands, (2) Departamento de Radiología, Universidad Complutense de Madrid, 28040 Madrid, Spain, (3) Universitat Rovira i Virgili. Facultat de Medicina i Ciències de la Salut, 43201 Reus, Spain (4) Asigma S.A., 30319 Cartagena, Spain

Abstract— **Purpose: Low contrast sensitivity of CT scanners is regularly assessed by (subjective) scoring of low contrast detectability within CT images of a phantom. Subjective rating of low contrast visibility might be biased since low contrast objects are arranged in known fixed patterns. The purpose of this study was to develop a software tool for automated objective low contrast detectability.**

Methods and Materials: Images of the low contrast module of the Catphan 600 phantom were used for the evaluations. The phantom contains three series of low contrast patterns (size 2 -15 mm; contrast 0.3, 0.5, and 1.0%). The software method offered automated determination of low contrast detectability using a NPWE model observer. As a proof of concept, influence of kV, mAs and reconstruction filter on image quality was investigated.

Results: Initial results show that effects of kV, mAs and reconstruction filter on image quality are clearly illustrated by the model observer.

Conclusion: We developed an automated method to investigate image quality objectively. As a first step, low contrast detectability as a function of both aquisition and reconstruction parameter settings was successfully investigated. In future work, this method could play a role in evaluation of image reconstruction algorithms, dose reduction strategies and novel CT technologies.

I. Introduction

Evaluating image quality is essential when investigating new acquisition protocols or new technical developments. A first approach is to look at individual physical properties of the image such as image contrast, resolution and noise. Together these aspects play an important role in detection, classification and estimation tasks in medical imaging.

Following on physical measurements, receiver operating characteristics (ROC) studies involving human observers, are a well known method of evaluating the impact of a particular image manipulation on clinical diagnosis. However, ROC studies may become time consuming and costly because they require a large number of human observations and a significant number of observers. Moreover, the number of possible conditions to investigate can be large.

As an alternative to human observers, computer-model observers can be considered. These are algorithms that attempt to predict human visual performance in noisy images. These models seem very useful in investigating numerous different conditions [1].

To more clearly demonstrate the impact of parameter settings and new technology on the CT image quality we developed an automated method to investigate low contrast detectability using a model observer. The difference between low contrast objects and background is so small that noise, resolution and contrast are significant factors in this test. The model observer is developed in Matlab and currently dedicated for detection of low contrast objects in the Catphan phantom. As a starting point, we chose the non-prewhitening matched filter with eye filter (NPWE model; note that different models may be implemented in future). The NPWE model uses information about the signal but also takes into account the differential human visual sensitivity to different spatial frequencies. The NPWE model therefore uses a template that matches the signal filtered by a human visual response function. The NPWE model has been able to predict human visual performance with respect to different backgrounds (including real anatomic backgrounds) [1, 2].

The method we propose could be helpful in investigating image reconstruction algorithms, dose reduction strategies and novel CT technologies. Also, it could enable fair comparisons between scanners of different vendors especially when image quality as a function of dose could be established.

II. Materials and Methods

A. Catphan Phantom

In this study the Catphan 600 Phantom (Phantom Laboratories, New York) was used. This phantom is constructed from modules that fit into a durable 20cm housing. The CTP515 module (positioned in the Catphan 600 phantom) consists of a series of 40 mm cylindrical rods of various diameters and three contrast levels to measure low contrast performance (see Fig. 1). Also subslice low contrast targets (truncated cylinders) have been included in this module.

In this study the phantom was always aligned with the axis of rotation (z axis). Only the outer 40 mm long objects with

the various diameters (i.e., 2, 3, 4, 5, 6, 7, 8, 9, and 15 mm) were chosen for this study. Our study will involve all three contrast series (i.e. 0.3, 0.5 and 1.0%) but the results shown in this paper are solely based on the highest low contrast series of 1%.

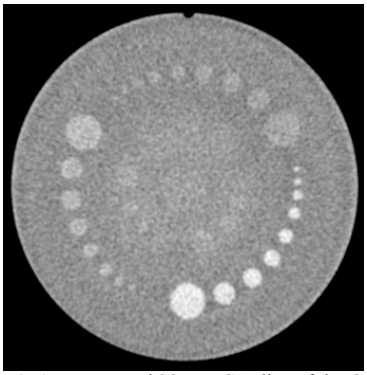


Fig. 1. A constructed 38 mm CT slice of the CTP515 module.

B. Image Acquisition and Reconstruction

Data acquisition concerning the phantom was performed using a 16-detector row CT scanner (Aquilion 16, Toshiba, Japan) in combination with the following parameters: beam collimation of 16x0.5 mm, tube voltage ranging from 80 - 135 kV, tube charge ranging from 25 mAs – 200 mAs. Scan field of view was 400 mm in all
acquisitions. All images were reconstructed at 0.5 mm slice thickness and 0.5 mm reconstruction interval. The images were reconstructed with FC12 (soft convolution kernel; body), FC50 (soft convolution kernel; lung), FC53 (sharp convolution kernel; lung) and FC81 (sharp convolution kernel; bone). Table I gives an overview of the acquisition and reconstruction parameters used.

TABLE I
OVERVIEW OF THE ACQUISITION AND RECONSTRUCTION PARAMETERS

Series	Tube voltage (kV)	Tube current (mAs)	Recon. filter	Slice Thickness (mm)
1	80	100	FC12	0.5
2	100	100	FC12	0.5
3	135	100	FC12	0.5
4	120	200	FC12	0.5
5	120	50	FC12	0.5
6	120	25	FC12	0.5
7	120	100	FC81	0.5
8	120	100	FC53	0.5
9	120	100	FC50	0.5

C. Processing of the Phantom Image Data

Detecting the Phantom and the Low Contrast Module
The phantom was detected by thresholding using a fixed threshold. Detection of the low contrast module within the phantom data was achieved based on knowledge of the CT value of the low contrast module and the presence of low contrast objects in the module in contrast to high contrast inserts in other modules.

Creating the Mask for Low Contrast Objects
The model observer uses templates for correlating the low contrast signals with their expected signal profiles (see section D). Templates with respect to each low contrast object were estimated using a constructed 38 mm slice (with consequently low noise due to the significant slice thickness) of the low contrast module in combination with the knowledge of the diameters of the objects. The highest low contrast series were used for this purpose. The location of the corresponding object templates was estimated by maximizing the signal within each object template (when measured in the corresponding area in the 38 mm slice). In this way a series of templates was obtained corresponding to the highest low contrast series. The remaining two series were obtained by rotating the estimated pattern respectively over an angle of 120° and 240°, resulting in the mask shown in Fig. 2.

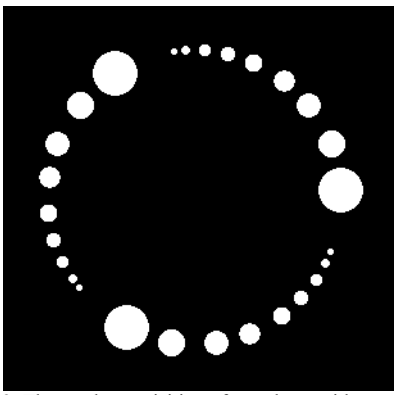


Fig. 2. The mask, consisiting of templates with respect to each low contrast object.

Scaling of the Mask
To apply the mask optimally to the different acquisition and reconstruction series from Table I, the mask was scaled and positioned based on the size and location of the phantom in the images. An automatically detected special high contrast reference object in the outer rim of the phantom was used to estimate the orientation of the phantom in the images. An angle of rotation was determined successively and applied to the mask to optimally fit the mask to the images.

D. NPWE Model

The non-prewhitening matched-filter observer with an eye-filter (NPWE) is a mathematical model that has been shown to be similar to human observers for detection tasks in the presence of low-pass noise [3]. Its strategy consists of correlating the image with the shape of the expected signal profile (of background and signal respectively) filtered by the visual-response function. In our experiments, the location of each template in the mask was separately optimized by shifting the template one pixel position in a 3x3 pixel environment and maximizing the correlation measure (signal location is uncertain case).

In a detection task, the model reaches a decision by comparing test statistics T_1 (test for background) and T_2 (test for signal). The test statistics are obtained by cross-correlation between the expected signal and the image [3]. The eye filter used in the model was $E(f) = fe^{-bf}$, with b chosen such that E(f) peaked at 4 cycles per degree. The eye filter is radially symmetric, and f is spatial frequency. In the experiments, a fixed viewing distance of 500 mm from the monitor was assumed [4].

From the distribution of test statistics, one can compute a discrimination index d′ [3]:

$$d' = \frac{\langle T \rangle_1 - \langle T \rangle_2}{\sqrt{\frac{1}{2}\sigma_1^{\,2} + \frac{1}{2}\sigma_2^{\,2}}}$$

Here <·> is the mean and σ(·) is the standard deviation of the respective distributions. The distributions are derived by performing cross-correlations in 76 consecutive slices of 0.5 mm slice thickness (for each series in Table I). This index can be used as a measure of detection performance. The discrimination index can be determined as a function of the low contrast signal energy (SE), as is described in [3] (i.e. the squared expected signal value integrated over all pixels in the observer template).

III. Results

Results comprising the higher (1%) low contrast series are shown in this section. In Fig. 3 the results with respect to the influence of kV is illustrated. The decrease in kV (with constant mAs) results in decreased low contrast detectability (lower d' values) according to the NPWE model. This can be explained by increased quantum noise using lower kV's. Note that the data points in the graphs correspond with the different object sizes in the highest low contrast series. Note also that signal energy increases with object size resulting in increased object detectability.

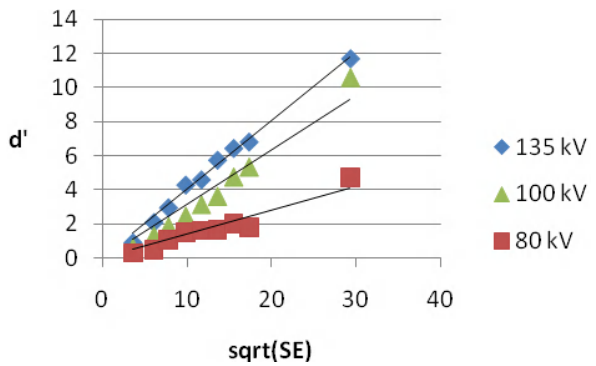


Fig. 3. Discrimination index d' as a function of signal energy for different kV's and constant tube charge (100 mAs).

Fig. 4 illustrates the influence of mAs (with constant kV). As can be expected, lower mAs values deteriorate the low contrast detecability due to decreasing signal-to-noise ratios.

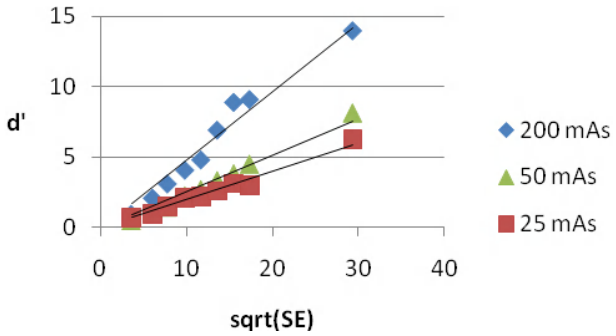


Fig. 4. Discrimination index d' as a function of signal energy for different tube charges and 120 kV tube voltage.

Fig. 5 finally, gives an impression of reconstruction filter influence. The soft filter gives the best performance. This filter suppresses the high frequencies (noise in these images) which strengthens the visibility of low contrast objects.

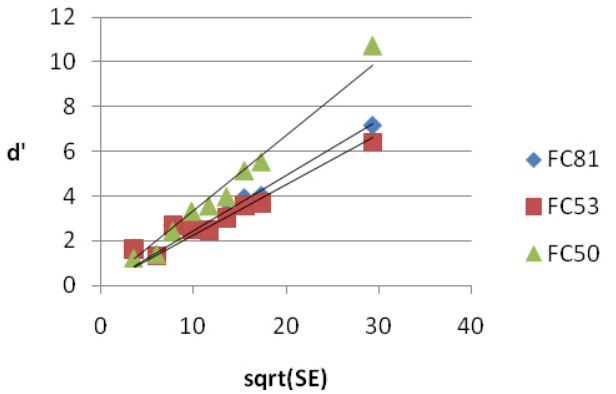


Fig. 5. Discrimination index d' as a function of signal energy for different reconstruction filters (120 kV and 100 mAs in all cases).

IV. Conclusion

In conclusion, we have validated a method for investigating low contrast detectability in CT images The proposed method can be considered a reasonable aid for investigating trends regarding diagnostic image quality as a function of dose reduction, acquisition and reconstruction parameter settings and new CT technologies.

Acknowledgment

References

[1] Handbook of Medical Imaging, Vol 1. Physics and Psychophysics, Jacob Beutel, Harold L. Kundel, Richard L. Van Metter, SPIE, 2000
[2] A. E. Burgess. Statistically defined backgrounds: performance of a modified nonprewhitening observer model, JOSA A, Vol. 11, Issue 4, pp. 1237-1242
[3] Reiser I, Nishikawa RM. Identification of simulated microcalcifications in white noise and mammographic backgrounds. Med Phys. 2006 Aug;33(8):2905-11.
[4] Veldkamp WJ, Kroft LJ, van Delft JP, Geleijns J. A technique for simulating the effect of dose reduction on image quality in digital chest radiography. J Digit Imaging. 2009 Apr;22(2):114-25.

GPU IMPLEMENTATION OF A 3D BAYESIAN CT ALGORITHM AND ITS APPLICATION ON REAL FOAM RECONSTRUCTION

Nicolas Gac[1]*, Alexandre Vabre*[2]*, Ali Mohammad Djafari*[1]*, Asier Rabanal*[1] *and Fanny Buyens*[2]

[1] L2S, Laboratoire des Signaux et Systemes (CNRS-SUPELEC-UPS), F-91191 Gif sur Yvette, France
[2]CEA, LIST, Laboratoire Images et Dynamique, , F-91191 Gif sur Yvette, France

ABSTRACT

A great number of image reconstruction algorithms, based on analytical filtered backprojection, are implemented for X-ray Computed Tomography (CT) [1, 3]. The limits of these methods appear when the number of projections is small, and/or not equidistributed around the object. In this specific context, iterative algebraic methods are implemented. A great number of them are mainly based on least square criterion. Recently, we proposed a regularized version based on Bayesian estimation approach. The main problem that appears when using such methods as well as any iterative algebraic methods is the computation time and especially for projection and backprojection steps. In this paper, first we show how we implemented some main steps of such algorithems which are the forward projection and backward backprojection steps on GPU hardware, and then we show some results on real application of the 3D tomographic reconstruction of metallic foams from a small number of projections. Through this application, we also show the good quality of results as well as a significant speed up of the computation with GPU implementation.

Index Terms— Computed Tomograhy (CT), Iterative 3D reconstruction, Bayesian estimation, GPU implementation

1. INTRODUCTION

The inverse problem we solve is to reconstruct the object $\boldsymbol{f}$ from the projection data $\boldsymbol{g}$ collected by a cone beam 3D CT. The link between $\boldsymbol{f}$ and $\boldsymbol{g}$ can be expressed as :

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon} \tag{1}$$

where $\boldsymbol{H}$ is the forward projection matrix operator modeling the acquisition system and $\boldsymbol{\epsilon}$ represents all the errors (modeling and measurement noise). The element H_{ij} represents the participation of the j pixel in the i data point.

In this discretized presentation of the CT forward problem, the backprojection (BP) solution can be expressed as $\widehat{\boldsymbol{f}}_{BP} = \boldsymbol{H}^t\boldsymbol{g}$ where $\boldsymbol{H}^t$ is the transpose of $\boldsymbol{H}$ and the filtered backprojection (FBP) method which is also equivalent to the Least squares (LS) solution can be expressed as $\widehat{\boldsymbol{f}}_{FBP} = (\boldsymbol{H}^t\boldsymbol{H})^{-1}\boldsymbol{H}^t\boldsymbol{g}$. The LS solution $\widehat{\boldsymbol{f}}_{LS} = \arg\min_{\boldsymbol{f}}\left\{Q(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2\right\}$ as well as the quadratic regularization (QR) solution

$$\widehat{\boldsymbol{f}}_{QR} = \arg\min_{\boldsymbol{f}}\left\{J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + \lambda\|\boldsymbol{D}\boldsymbol{f}\|^2\right\} \tag{2}$$

can be obtained by a gradient based optimization algorithm which can be described as follows:

$$\left\{\begin{array}{l} \boldsymbol{f}^{(0)} = \boldsymbol{H}^t\boldsymbol{g} \\ \boldsymbol{f}^{(i+1)} = \boldsymbol{f}^{(i)} + \alpha\left[\boldsymbol{H}^t(\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}^{(i)}) + \lambda\boldsymbol{D}^t\boldsymbol{D}\boldsymbol{f}^{(i)}\right] \end{array}\right. \tag{3}$$

where α is a fixed, variable or computed optimally step size and (i) is the iteration number. Looking at this iterative algorithm, we can distinguish, at each iteration the following operations:

1. Forward projection operation: $\widehat{\boldsymbol{g}} = \boldsymbol{H}\widehat{\boldsymbol{f}}$
2. Computation of the residuals: $\delta\boldsymbol{g} = \boldsymbol{g} - \widehat{\boldsymbol{g}}$
3. Backprojection operation of the residual: $\delta\boldsymbol{f}_1 = \boldsymbol{H}^t\delta\boldsymbol{g}$
4. Computation of the regularization or a priori term: $\delta\boldsymbol{f}_2 = \lambda\boldsymbol{D}^t\boldsymbol{D}\widehat{\boldsymbol{f}}$
5. Updating of the solution for the next iteration: $\boldsymbol{f}^{(i+1)} = \boldsymbol{f}^{(i)} + \alpha(\delta\boldsymbol{f}_1 + \delta\boldsymbol{f}_2)$

As we can see the implementation of such iterative algorithm as well as any other more sophisticated algorithm such as the Bayesian estimation approach we propose needs these operations. The two main steps are the steps 1 and 3. As we will see later, we implemented these two steps using GPU. So, one of the main contribution of this paper is the presentation of this implementations and their relative performances. The second contribution of this paper is adaptation of a particular Bayesian estimation approach with appropriate prior modelling which is particularly adapted for our application which is related to Non Destructive Testing (NDT) application.

E-mail : nicolas.gac@lss.supelec.fr, alexandre.vabre@cea.fr, djafari@lss.supelec.fr

In the following, first, we present the basic ideas of our Bayesian estimation approach and in particular the prior model we proposed and used. Then, we present the main steps of the resulting Joint Reconstruction-Segmentation-Characterization Algorithm (JRSCA) we developped. Then, we detail as much as possible the implementation on GPU parts and their performances, and finally, we show the 3D reconstruction results obtained for our application and we conclude on this paper.

1.1. Bayesian method

The proposed Bayesian method lies on a prior model for the object $\boldsymbol{f} = \{f(\boldsymbol{r}), \boldsymbol{r} \in \mathcal{R}\}$ where $\boldsymbol{r} = (x, y, z)$ represents a voxel position. This model considers that the object $f(\boldsymbol{r})$ is composed of a finite number K of materials; all voxels of the same material are grouped in compact regions $\mathcal{R}_k$, labeled by a hidden variable $z(\boldsymbol{r}) = k,\ k = 1, \cdots, K$. We then have $\mathcal{R}_k = \{\boldsymbol{r} : z(\boldsymbol{r}) = k\}$. To translate the homogeneity in each class of matierial, we use:

$$p(f(\boldsymbol{r})|z(\boldsymbol{r}) = k, m_k, v_k) = \mathcal{N}(m_k, v_k) \tag{4}$$

and to translate the desir that all the voxels in a given class be grouped in compact regions, we use a Potts-Markov model for $\boldsymbol{z} = \{z(\boldsymbol{r}), \boldsymbol{r} \in \mathcal{R}$:

$$p(\boldsymbol{z}) \propto \exp\left[\sum_{\boldsymbol{r}\in\mathcal{R}}\sum_{k}\alpha_k\delta(z(\boldsymbol{r}) - k) + \gamma\sum_{\boldsymbol{r}'\in\mathcal{V}(\boldsymbol{r})}\delta(z(\boldsymbol{r}) - z(\boldsymbol{r}'))\right] \tag{5}$$

where $\mathcal{V}(\boldsymbol{r})$ means the neighborhood of $\boldsymbol{r}$ and $\{\alpha_k, k = 1, \cdots, K\}$ and γ are the Potts model parameters. The parameters m_k, v_k and also standard variation of the noise v_ϵ are called the hyperparameters $\theta = \{(m_k, v_k, \alpha_k), k = 1, \cdots, K; v_\epsilon\}$. With this prior model and a centered uncorrelated Gaussian model for the noise, we can obtain the expression of all the probability laws $p(\boldsymbol{g}|\boldsymbol{f}, v_\epsilon)$, $p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{\alpha}, \boldsymbol{v})$, $p(\boldsymbol{z}|\gamma, \boldsymbol{\alpha})$ and the joint a posteriori law $p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\boldsymbol{g})$ and all the conditionals $p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{\theta}, \boldsymbol{g})$, $p(\boldsymbol{z}|\boldsymbol{f}, \boldsymbol{\theta}, \boldsymbol{g})$ and $p(\boldsymbol{\theta}|\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{g})$ which are needed to estimate jointly the object $\boldsymbol{f}$, the image of $\boldsymbol{z}$ which will show the segmented and classified volume and the parameters $\boldsymbol{\theta}$ which charcterize all the classes.

The iterative algorithm structure is then constituted of three main steps, as follows:

- Reconstruction step: Updating $\boldsymbol{f}$ by computing $\widehat{\boldsymbol{f}}^{(i+1)} = \arg\max_{\boldsymbol{f}}\{p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{\theta}, \boldsymbol{g})\}$. This is done by using a gradient type optimization algorithm.
- Segmentation step: Updating $\boldsymbol{z}$ by generating a sample from $p(\boldsymbol{z}|\boldsymbol{f}, \boldsymbol{\theta}, \boldsymbol{g})$. This needs a sampling algorithm from a Potts-Markov model.
- Characterization step: Updating the hyperparameters using $p(\boldsymbol{\theta}|\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{g})$. This step can be done either analytically or by sampling from known probability laws such as Gaussians or Inverse Gamma.

More details about this method can be found in [5].

2. ALGORITHM SPEED UP

2.1. Introduction

In this paper, we focus more on a hardware speed up of some of these steps. A preliminary study has been conducted to estimate which hardware architecture is the more appropriate to each calculation step: Cell [6], FPGA, CPU, cluster of PC's, graphic processing units [7, 8]. And so from the literature [9], for gradient descent (95 % of the calculation time), graphic processors such as GPU seem well adapted. The convergence of the algorithm has to be warranted for the different chosen parameters, such as: N (local number of iteration for gradient descent), M (segmentation number of iterations) and I (global number of iterations). The proposed method includes not only a reconstruction of the CT data but also a segmentation of the volume into classes. Recent works have been carried out on similar approaches for binary cases using discrete tomography [10]. Our approach allows to have any number of materials that needed, and also we associate a probability law to belong to a given class [5]. Priors are also introduced on the voxel class estimation according to their neighborhood.

2.2. Implementation of projector and backprojector

For the iterative step of gradient descent, the two main consuming time operations are projection ($\boldsymbol{Hf}$) and backprojection ($\boldsymbol{H}^t\delta\boldsymbol{g}$) which are used to estimate a convergence criterion and its gradient. These two operations represent 95 % of the computing time.

The follow up of the work aims at speeding up these two steps. GPU hardware, since 2006 is one of the most used tool inside research community. Both simplicity in implementation and performance improvements have imposed scientic community to migrate to such a tool. Recent improvements from NVidia have allowed to dispose of CUDA, this developing environment allows to design operating software with high computing performances.

In order to compute the two matrix operations ($\boldsymbol{Hf}$ and $\boldsymbol{H}^t\delta g$) without the too expensive memory use of H=(h_{ij}) (1 To is needed to store H for a 2048^3 reconstruction), projection and backprojection geometric operators are widely used. This operators compute in line the coefficient h_{ij}, instead of reading a matrix H stored in memory. Differents kinds of projection and backprojection algorithms can be used [11, 12].

For each operator, we choose the one which enables the best implementation on Nvidia GPUs with CUDA. As a consequence, our projection/backprojection pair is unmatched. Thus each operator defines a different matrix H: H_p for projection and H_{bp} for backprojection. Use of unmatched backprojection/projection pairs is widely used. Indeed, effect on convergence is in pratice not penalising during the first iterations [13]. Main difference on backprojection and projection

algorithm is the main loop of computation : for backprojection, the loop is on voxels (voxel-driven) and for projection it is on X rays (ray-driven).

2.3. Backprojection

Backprojection algorithm used is a voxel-driven (main computation loop on voxels) with a bi-linear interpolation done on detector pixels. Loops are ordered in manner to exploit as much as possible the spatial and temporal locality of memory access as described in [9]. In CUDA parallelization scheme, one thread is responsible to one voxel reconstruction. Memory accesses to the 2D projection of the volume is done via the 2D texture available on GPUs which allows a cache access to global memory and a hardwired bi-linear interpolation. Standard software optimizations techniques have been carefully used : pre-computation stored on constant cache-memory, incremental computation used as much as possible and loop unrolling.

2.4. Projection

Projection algorithm used is a ray-driven (main computation loop on rays) with a tri-linear interpolation done on volume voxels. In CUDA parallelization scheme, one thread is responsible to integrate the 3D volume along one X-ray. The volume integration for a ray is done simply by sampling regularly the volume along the ray. Memory accesses to the 3D volume is done via the 3D texture available on GPUs which allows a cache access to global memory and a hardwired tri-linear interpolation. Standard software optimization techniques have been used for projection as well.

3. REAL DATA RECONSTRUCTION

3.1. Metallic foams

Solid foams are a class of materials with a complex behavior related to the properties of the constitutive material, the geometry and the topology of the material distribution. These materials present a very high porosity, and are thus very light, but nevertheless very resistant due to a good distribution and architecture of matter. The most known examples of such materials are bone and wood, or also coral and sponge.

Metallic foams are very recent materials. The application field of these materials is very large: they can be used as deformation absorbers in mechanical engineering or fluid distributors for many applications such as thermal exchangers, fuel cells and electrolysers. A strong need in modeling tools as reliable as possible is necessary to make clearer the behavior of these materials and to design optimal foams for desired application [14]. It is necessary to estimate the mechanisms that control their deformations, their durability versus time or stresses to employ them. It is also necessary to study their behavior versus mass and thermal transfers to address fluid flow applications. In this context, our work is focused on collecting basic knowledge on fluid two-phase flows in metallic foams [15]. A scientific community works on flows in porous media for geology or oil extraction. Our idea is to implement the modelling methods developed in the context of fractured geologic medias and adapt them to the metallic foam structures [16].

However, in order to obtain reliable results from these modelling methods, it is necessary to obtain of a thin topology and geometry foam structures. For topology characterization, the pore size distribution and the specific surfaces are fundamental parameters, i.e. the normalized surface of the foam. A high spatial resolution of three-dimensional structure of the foam (in the magnitude of 5 m) is required for geometry characterization [17]. In the follow-up, we present our studies on water kinetics in open-cell nickel foams using x-ray microtomography. The experiments are conducted on a small sample size (1 mm^3 foam) to estimate the thin geometry and model the water behavior at a scale of few pores.

Data set is made of 96 projections on the 256^2 plane detector. The volume is reconstructed inside the cylindric field of view of the X ray tomograph. Thus, we reconstruct 256 (z dimension) * $\Pi \cdot 128^2$ (x,y dimensions) voxels.

3.2. Reconstruction time

We have used a Nvidia GTX 295 to reconstruct the metallic foam. Only one GPU is used here, no multi-GPU implementation has been done. Reconstruction time are greatly penalised by memory transfer between CPU and GPU.

The purpose of this work is not to evaluate the acceleration factor obtained on GPU (see [9, 6] for time comparaison on CPU and GPU). But compared to the former reconstruction "C++" software used in the CEA lab , we reach about 100 acceleration factor. Previously a 100 iteration reconstruction took days and now it takes hours.

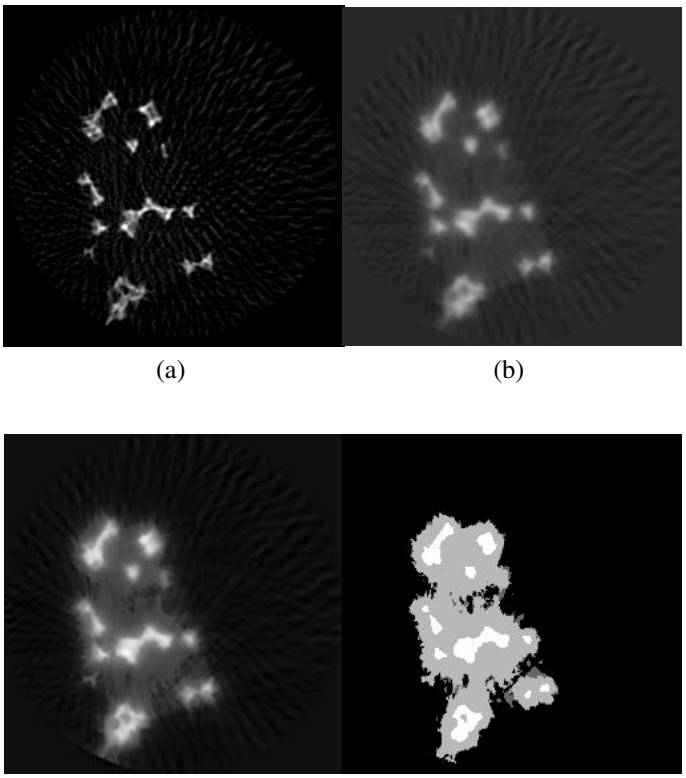
Operator	Time
Projector	755 ms (128 ms for memory transfer)
Backprojector	234 ms (133 ms for memory transfer)

Table 1. Reconstruction time on a GTX 295 ($96 * 256^2$ data)

3.3. Foams reconstructed

As first results, we present here the foam reconstructed with a non Bayesian iterative algorithm and with our Bayesian iterative algorithm. As we can observe on figure 1, while standard algorithm like FDK (a) or an iterative quadratic regularization method (b) does not suceed to reconstruct the water inside the metallic foam, our method suceeds to reconstrut it (c), and provide a segmentation image (d) . The used prior model which suppose that the reconstructed object is constituted of

N compact regions $\mathcal{R}_k$,, is well adapted to this context of data set.



(a) (b)

(c) (d)

Fig. 1. Foam reconstructed : (a) Slice reconstructed with a FDK method (b) Slice reconstructed with a non bayesian method (standard gradient descent after 50 iterations); (c) Slice recontructed with our method (after 50 iterations); (d) Segmentation obtained during iterative reconstruction

4. CONCLUSION AND PERSPECTIVES

We have presented an original method based on a Bayesian statistical method for 3D tomographic reconstructions. The main interest is to apply it to a context of non-consistent data sets, for example with a small number of projections. We have shown a good quality of our first results on an experimental data set with low contrasted regions (air/water as compared to nickel) acquired on the CT set-up of our lab and a signicant speed up of the calculation with GPU implementation.

Both backprojection and projection steps were implemented on GPU. The obtained performance for the global reconstruction time is in the magnitude of 100. Hovewer, performance of our projector can be still improved. In this goal, a Joseph projector implemented also on GPU, would be compared in term of time and qualiy of reconstruction. Our futur work will be focusing on the study of the effect of the unmatched projector/backprojector pair on the reconstruction process. For this purpose, a matched voxel-driven projector has been implemented on CPU.

Our futur goal is to reconstruct 1024^3 real data acquired on the new CT set-up of the lab (1024^2 detector pixels). We are currently working on a multi-GPU implementation in order to handle such large data sets. GPU implementations of computation costly steps as 3D convolution, segmentation step would help to speed-up even more the reconstruction process.

5. REFERENCES

[1] K.J. Batenburg and J. Sijbers, "DART: a fast heuristic algebraic reconstruction algorithm for discrete tomography," 2007, vol. IV, pp. 133–136, IEEE Conference on Image Processing.

[2] Fang Xu and Klaus Mueller, "Real-time 3d computed tomographic reconstruction using commodity graphics hardware," *Physics in Medicine and Biology*, vol. 52, no. 12, pp. 3405–3419, 2007.

[3] Steckmann S. Marone F. Kachelrie M., Knaup M. and Stampanoni M., "Hyperfast o(2048**4) image reconstruction for synchrotron?based x?ray tomographic microscopy," in *MIC proceedings*, 2008.

[4] G. Pratx, G. Chinn, P. D. Olcott, and C. S. Levin, "Fast, accurate and shift-varying line projections for iterative reconstruction using the GPU," *Medical Imaging, IEEE Transactions on*, vol. 28, no. 3, pp. 435–445, March 2009.

[5] S. Fékih-Salem, A. Vabre, and A. Mohammad-Djafari, "Bayesian tomographic reconstruction of microsystems," in *Bayesian Inference and Maximum Entropy Methods, AIP Conf. Proc. 954*, K. et al. Knuth, Ed. MaxEnt Workshops, July 2007, pp. 372–380, American Institute of Physics.

[6] Marc Kachelriess, Michael Knaup, and Olivier Bockenbach, "Hyperfast parallel-beam and cone-beam backprojection using the cell general purpose hardware," *Medical Physics*, vol. 34, no. 4, pp. 1474–1486, April 2007.

[7] W. Xu and K. Mueller, "A performance-driven study of regularization methods for GPU-accelerated iterative CT," in *Workshop on High Performance Image Reconstruction (HPIR)*, 2009.

[8] W. Xu and K. Mueller, "Learning effective parameter settings for iterative CT reconstruction algorithms," in *Workshop on High Performance Image Reconstruction (HPIR)*, 2009.

[9] N. Gac, S. Mancini, M. Desvignes, and D. Houzet, "High speed 3D tomography on CPU, GPU and FPGA," *EURASIP Journal on Embedded Systems, special issue on "Design and Architectures for Signal Image Processing"*, vol. Volume 2008 (2008), pp. Article ID 930250, 2008.

[10] K.J. Batenburg and J. Sijbers, "Automatic multiple threshold scheme for segmentation of tomograms," 2007, vol. 6512.

[11] Peter M. Joseph, "An improved algorithm for reprojecting rays through pixel images," vol. 1, no. 3, pp. 192–196, Nov. 1982.

[12] R.L. Sidon, "Fast calculation of the exact radiological path for a three-dimensional CT array," *Med. Phys.*, vol. 12, no. 2, pp. 252–255, 1985.

[13] G.L. Zeng and G.T. Gullberg, "Unmatched projector/backprojector pairs in an iterative reconstruction algorithm," *Medical Imaging, IEEE Transactions on*, vol. 19, no. 5, pp. 548–555, May 2000.

[14] O. Gerbaux, S. Crosnier, and M. Dubruel, "Experimental and numerical predictions of pressure drops for gas flow through isotropic metallic foams," in *France-Deutschland Fuel Cell Conference, Belfort*, 2004.

[15] M.L. Turner, C.H. Arns L. Knfing, A. Sakellariou, T.J. Senden, A.P. Sheppard, R.M. Sok, A. Limaye, W.V. Pinczewski, and M.A. Knackstedt, "Three-dimensional imaging of multiphase flow in porous media," *Physica A*, vol. 339, pp. 166–172, 2004.

[16] S. Bekri and P.M. Adler, "Dispersion in multiphase flow through porous media," *International Journal of Multiphase Flow*, vol. 28, pp. 665–697, 2002.

[17] O. Brunke, A. Hamann, S.J. Cox, and S. Odenbach, "Experimental and numerical analysis of the drainage of aluminium foams," *Journal of Physics: Condensed Matter*, vol. 17, pp. 6353–6362, 2005.

Image Reconstruction from Sparse Data in Synchrotron-based Micro-tomography of Biomedical Samples

Xianghui Xiao[1], Dan Xia[2], Junguo Bian[3], Xiao Han[3], Emil Sidky[3], Francesco De Carlo[1], and Xiaochuan Pan[3]
[1]Advanced Photon Source, Argonne National Laboratory, Chicago, IL, U.S.A.
[2]Paul C. Lauterbur Research Center for Biomedical Imaging, Institute of Biomedical and Health Engineering, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, China
[3]Department of Radiology, The University of Chicago, Chicago, IL, U.S.A.

***Abstract*—Synchrotron radiation micro-CT (SR-μCT) is a useful tool for specimen studies. However, radiation damage to biology specimens can be an obstacle for obtaining high-quality data. The radiation damage can be minimized through the use of a considerably reduced number of projection views. Using experimental data collected for a number of specimens, including a sea urchin spine specimen, we investigate image reconstruction from sparse projection data by use of carefully designed algorithms, including the total-variation (TV) algorithm. Results of our studies indicate that, from much reduced data, the TV algorithm can yield images that comparable to those obtained from full data by use of the existing algorithm in current SR-μCT studies, suggesting that the TV algorithm may find broad applications in SR-μCT.**

I Introduction

Synchrotron radiation micro-CT (SR-μCT) is a useful tool in study of a specimen's structure preserved in its natural state, thus preventing structure deformation or destruction otherwise encountered in its histologic fixation and sectioning. In comparison with other non-destructive inspection techniques, such as magnetic resonance imaging (MRI), micro-CT, and confocal microscopy, SR-μCT is capable of imaging millimeter-sized samples with micron-scale resolution. In SR-μCT, a synchrotron source delivers intense X-ray over a wide energy spectrum, ranging from a few eV to a few-hundred keV. With an X-ray monochromator, X-ray with a narrow energy band can be selected for illuminating the imaged specimen. Because SR-μCT uses virtually monochromatic X-ray, it is free of beam hardening effect.

Despite the fact that SR-μCT possesses unique, desirable properties such as monochromatic X-ray energy, issues exist that would affect its adequate applications to some imaging tasks. An issue concerns the large number of projection views that are required for yielding images without significant artifacts reconstructed by use of existing algorithms, such as the filtered-back projection (FBP) algorithm. Such a large number of projection measurements not only increases the imaging time but also, more importantly, may induce serious radiation damage to the specimen. This is a particularly important issue when a wet biological specimen is imaged, because the high flux of synchrotron X-ray radiation can induce significant internal structure deformation within the wet specimen. One possible approach to reducing imaging time and specimen structural deformation in SR-μCT imaging is to decrease the number of projection views.

In this work, we investigate image reconstruction in SR-μCT and demonstrate images without significant artifacts can be reconstructed by use of carefully designed algorithms from data acquired at a number of projection views significantly lower than that currently used in a typical SR-μCT imaging experiment. Also, it is not uncommon that, because of physical and hardware constraints, data can be acquired over a limited angular range in SR-μCT imaging experiments. Therefore, we will also investigate image reconstruction from SR-μCT data collected over an angular range less that 180°.

II SR-μCT system and imaging experiments

In the work, we use the SR-μCT system at the Advanced Photon Source (APS) at Argonne National Laboratory (ANL) to collect data from a sample of sea urchin spine. The SR-μCT system, as shown in 1, has seven degrees of motion freedom that allow one to control the orientation of the rotation axis and the sample position. The detector component is composed of a scintillator, a microscope lens, and a Coolsnap 2K×2K CCD camera with a pixel size of 7.5 μm. The scintillator converts X-ray into visible lights that are subsequently picked up by the microscope lens and then recorded as a magnified projection image on the CCD camera. Depending upon the magnification of the microscope lens, the size of a pixel can vary between 0.75 μm and 6 μm in the image space. The detector component is placed on a translation stage along the beam direction so that the distance between detector and the imaged sample can be adjusted.

Sea urchin spine sample is highly fenestrated single crystals of high magnesium calcite and contains about 50vol% open space. Yet, it has a remarkable strength and flexibility. Because of its unique, interesting structural and mechanical properties of high practical interest, sea urchin spine is used often as a model system in bionic engineering. The sea urchin spine sample used in the work is an *A. radiata* spine with a diameter of 0.6 mm that was col-

lected at Miyako Island, Japan. We used a 2.5× Zeiss microscope lens that allows a detector pixel to sample the image space with a 3μm resolution. The projection data were collected at a total of 400 views evenly distributed over 180° with a 0.36° angular step. In our experiment, the X-ray energy was 20 KeV, and whereas the detector-sample distance was 5 mm At such a short distance, the phase-contrast effect due to X-ray propagation is insignificant. As such, the imaging mechanism can be considered absorption-contrast dominated, and the projection data can thus be modeled adequately as the 2D Radon transforms of the linear attenuation coefficient distribution within the sample.

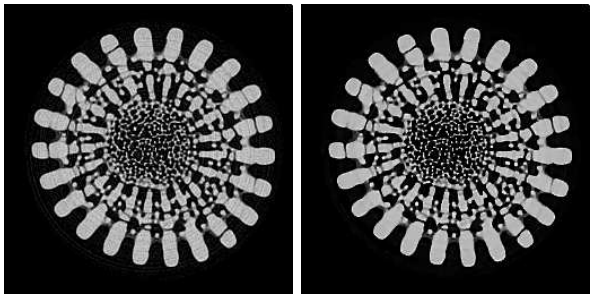
III Reconstruction algorithms

In synchrotron X-ray imaging, the FBP algorithm has been used as the primary algorithm for yielding images. It can reconstruct images with high quality from high quality data collected at a large number of projection views. However, when the number of views decreases, the FBP reconstruction often suffers from strong artifacts due to the lack of sufficient angular samples. In this work, we focus on the investigation of using advanced algorithms for image reconstruction from data acquired at a number of views substantially below that used currently in synchrotron X-ray imaging.

In an attempt to address image reconstruction from sparse projection data, we have recently developed a constrained total-variation (TV)-minimization algorithm [1, 2], which is referred to as the TV algorithm in the work. In the TV algorithm, an image is reconstructed through the minimization of its TV under data and image-positivity constraints. Existing studies seem to suggest that the TV algorithm has the potential to yield images from highly sparse data, including data acquired at the small number of views. Therefore, in the work, we adjust and apply the TV algorithm to reconstructing images specifically from data acquired with the SR-μCT system described above. Detailed information above the TV algorithm and its implementation was provided in Refs. [1, 2]. However, we will report at the conference additional adjustments involved in the application of the TV algorithm to reconstructing images in synchrotron X-ray CT experiments. Also, in an attempt to evaluate algorithm performance, we have also implemented some existing algorithms such as projection-onto-convex-set (POCS) [3] and expectation-maximization (EM) algorithms [4, 5] and applied them to reconstructing from the same data sets. Their performance will be compared with that of the TV algorithm in the imaging task described above.

IV Results

In the experiment of imaging a sea urchin spine sample, we have collected a set of data at 400 projection views evenly distributed over 180°, which we referred to as the full data set. Clearly, the "true structure" of the sea urchin spine is unknown. Therefore, we reconstruct images of the sample, which are shown in 2, from the full data set by use of the FBP and TV algorithms and will use them as the FBP-reference and TV-reference images (i.e., as the "gold standard" images) in evaluation studies.



a b

Fig. 2. Images reconstructed from the full data set by use of (a) the FBP and (b) TV algorithms. Display window: [0.0, 0.5].

A *Image reconstruction from SR-μCT data collected at a reduced number of views*

From the full data set, we extracted a subset of data at 80 projection views uniformly distributed over 180°. From this data subset, we use the FBP, POCS, EM, and TV algorithms to reconstruct images, and the reconstruction results are displayed in the upper row of Fig. 3. It can be observed from the results that the TV algorithm yields reconstruction with less artifacts than those obtained with the FBP, EM, and POCS algorithms. For comparison, we also display in Fig. 3 the absolute difference images between the reconstructed images and the FBP-reference image (lower row). As the result in the lower row of Fig. 3 shows the difference between images reconstructed with the TV algorithm seems to be lower than those obtained with the FBP, POCS, and EM algorithms.

B *Image reconstruction from SR-μCT data collected over a limited angular range*

Due to certain physical and hardware constraints, data may be collected only over a limited angular range in SR-μCT imaging. For example, the sample structure deformation often gradually develops during data acquisition, and the deformation in measurements taken at earlier views is likely to be less noticeable than those taken at later views. Consequently, the portion of data acquired in the very late part of the measurement process may not be usable due to the serious structural deformation, and the relatively consistent data are available only over a limited angular range. In attempt to simulate the case, we extracted a limited-angular-range subset of data at 160 projection views that uniformly distributed over an an-

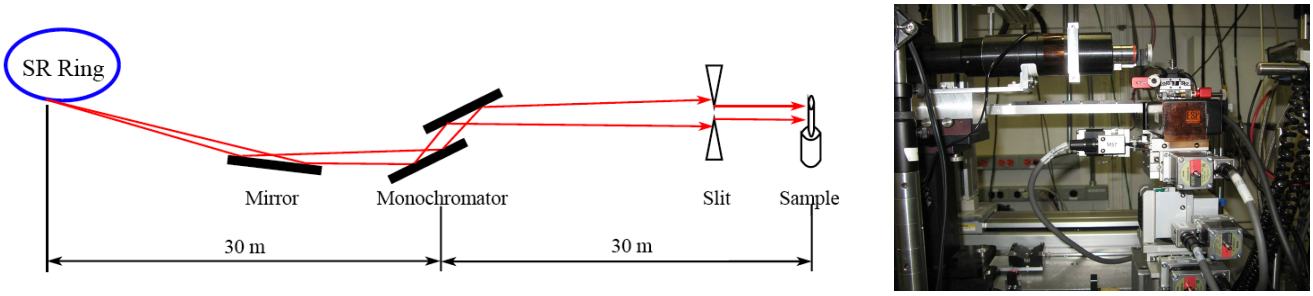
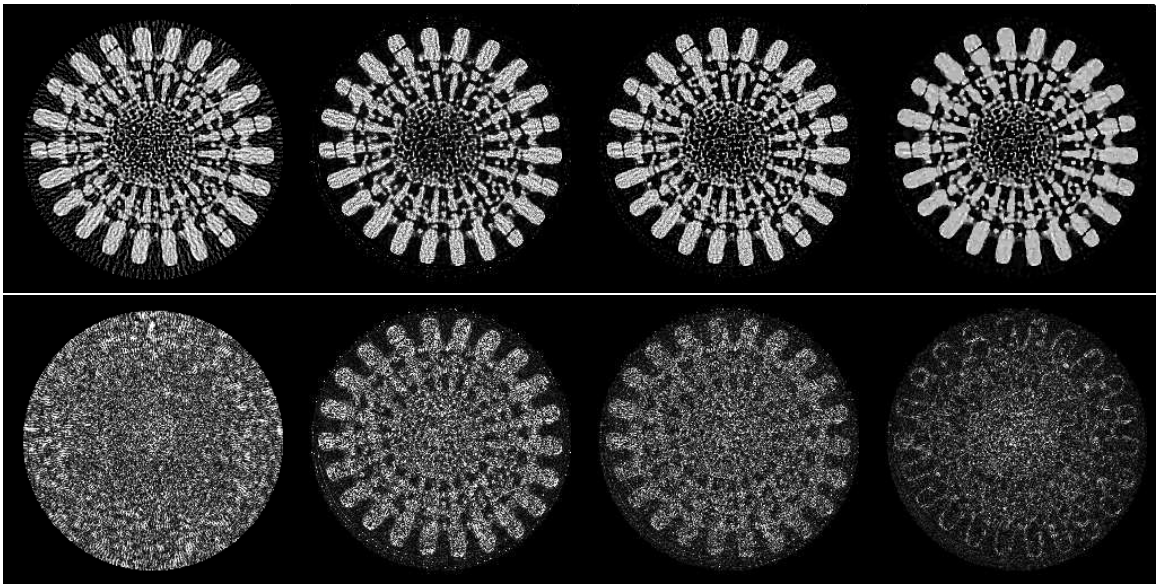


Fig. 1. Schematic (left) and photo (right) of the SR-μCT system at beamline 2BM of the Advanced Photon Source at Argonne National Laboratory. The specimen data used in the work were collected with this system.



a b c d

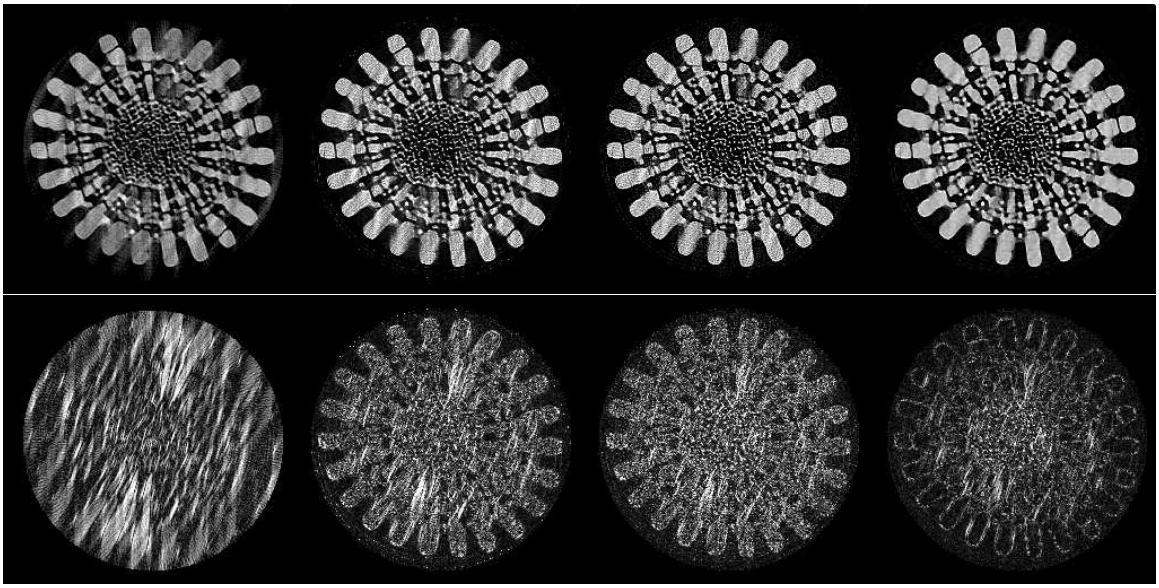
Fig. 3. Upper row: images reconstructed from 80-view data by use of (a) FBP, (b) EM, (c) POCS, and (d) TV algorithms. Display window: [0.0, 0.5]. Lower row: absolute differences of between images in the top row and the FBP-reference image. Display window: [0.0, 0.2].

gular range of 144^o. From this subset, we use the FBP, POCS, EM, and TV algorithms to reconstruct images, and the reconstruction results are displayed in the upper row of Fig. 4. It can be observed from the results that the TV algorithm yields reconstructions with fewer artifacts than those obtained with the FBP, EM, and POCS algorithms. Again, for comparison, we display in Fig. 4 the absolute difference images between the reconstructed images and the FBP-reference image (lower row). The difference between images reconstructed with the TV algorithm seems to be lower than those obtained with the FBP, POCS, and EM algorithms.

C Evaluation studies

We have also performed quantitative evaluation studies of reconstruction quality using a number of image quality metrics, including the mutual information (MI) [6] and the universal quality index (UQI) [7]. The index MI is used to describe similarity of the image intensities of corresponding pixels within selected regions of interest (ROIs) between an image and a reference image, whereas the index UQI is used for describing the structural similarity within selected ROIs between an image and a reference image [7]. The closer an image to the reference image, the higher the values of MI and UQI. In particular, UQI= 1.0 the image is identical to the reference image.

For an image under evaluation, we selected within the image seven circular ROIs radii that are 30%, 40%, 50%, 60%, 70%, 80%, and 90%, respectively, of the radius of the field of view. Using the corresponding ROIs within a reference image, we calculated MIs and UQIs for the seven ROIs. In Fig. 5 and 6, we displayed MIs and UQIs calculated for the seven ROIs in images reconstructed displayed in the top rows of Figs. 3 and 4, respectively, when the FBP-reference image is used. Similarly, in Fig. 7 and 8, we displayed MIs and UQIs calculated for the seven ROIs in images reconstructed displayed in the top rows



a b c d

Fig. 4. Upper row: images reconstructed from data acquired over a limited angular range of 144°, by use of (a) FBP, (b) EM, (c) POCS, and (d) TV algorithms, respectively. Display window: [0.0, 0.5]. Lower row: absolute differences of between images in the top row and the FBP-reference image. Display window: [0.0, 0.2].

of Figs. 3 and 4, respectively, when the TV-reference image is used. Results of the quantitative studies in Figs. 5-8 demonstrate that the MI and UQI values of the TV reconstructions are generally higher than those of other algorithms, even when the FBP-reference image is used, indicating that the TV algorithm yields images of quality higher than do the other algorithms studied.

V Conclusions

In this work, we demonstrate that the SR-μCT data is suitable to the novel TV algorithm. With TV algorithm, few-view and limited-few-view projection data can be successfully reconstructed. Our quantitative studies show that, compared to FBP, EM, POCS algorithms, TV algorithms can yield the better reconstructions in terms of MI and UQI even if different references are used. Taking fewer projection views is critical to prevent the internal structure deformation in biomedical specimens in SR-μCT. It is expected that TV-minimization algorithm will have broad applications in SR-μCT. The results of additional, detailed studies about TV reconstruction performance of TV algorithm for sparse SR-μCT data will be report at the conference.

References

[1] E. Y. Sidky, K.-M. Kao, and X. Pan, "Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT," *Journal of X-Ray Science and Technology*, vol. 14, pp. 119–139, 2006.

[2] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Physics in Medicine and Biology*, vol. 53, pp. 4777–4807, 2008.

[3] R. Gordon, R. Bender, and G. T. Herman, "Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography," *J. Theor. Biol.*, vol. 29, pp. 471–481, 1970.

[4] A.P. Dempster, N.M. Laird, D.B. Rubin, et al., "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.

[5] K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography," *J. Comput. Assisted Tomogr.*, vol. 8, pp. 306–316, 1984.

[6] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multi-modality image registration maximization of mutual information," in *Proceedings of MMBIA*, 1996, pp. 14–22.

[7] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE. Trans. Image Proc.*, vol. 13, pp. 600–612, 2004.

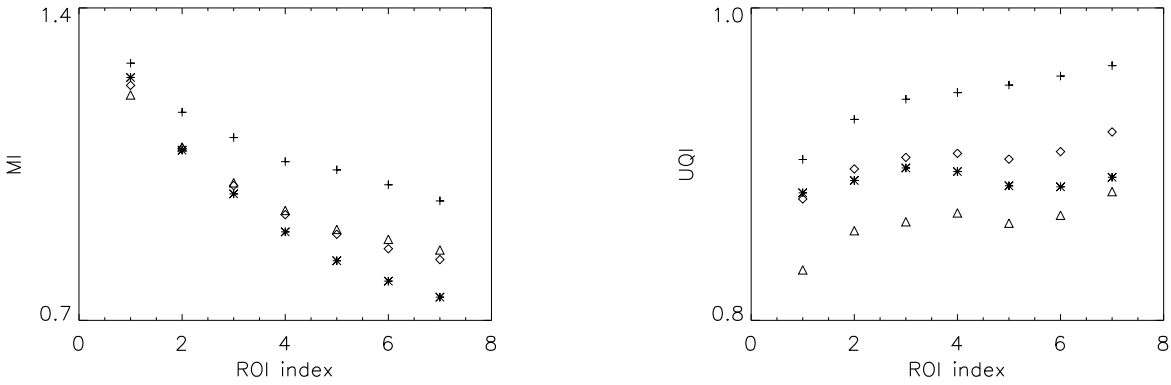


Fig. 5. MIs (left) and UQIs (right) calculated from images displayed in the upper row of Fig. 3, which are obtained from 80-view data by use of FBP ("⋆"), EM ("△"), POCS ("◇") and TV ("+") algorithms, with respect to the FBP-reference image.

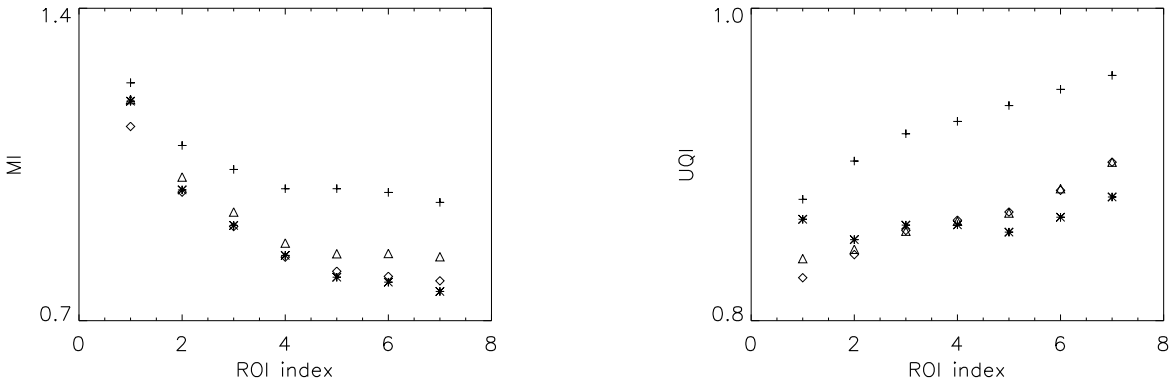


Fig. 6. MIs (left) and UQIs (right) calculated from images displayed in the top row of Fig. 4, which are obtained from 160-view data by use of FBP ("⋆"), EM ("△"), POCS ("◇") and TV ("+") algorithms, with respect to the FBP-reference image.

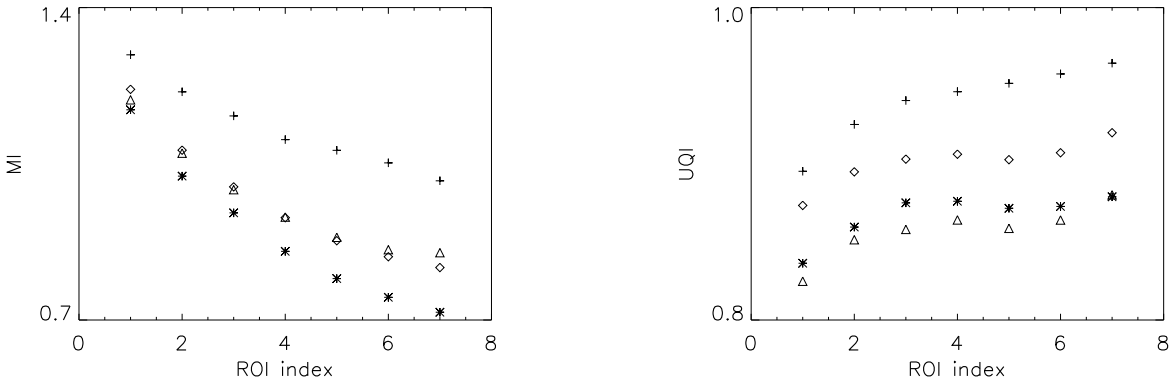


Fig. 7. MIs (left) and UQIs (right) calculated from images displayed in the top row of Fig. 3, which are obtained from 80-view data by use of FBP ("⋆"), EM ("△"), POCS ("◇") and TV ("+") algorithms, with respect to the TV-reference image.

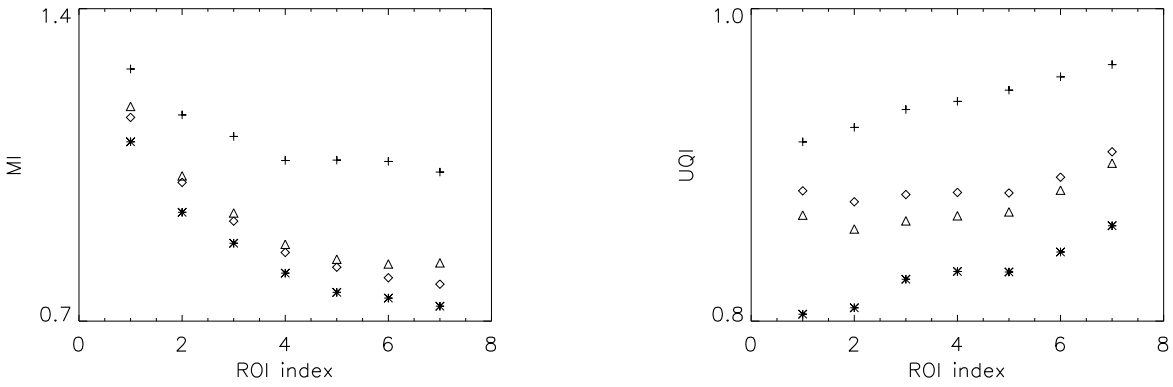


Fig. 8. MIs (left) and UQIs (right) calculated from images displayed in the top row of Fig. 4 with respect to the TV-reference image.

Motion compensated filtered backprojection for non-rigid deformation

J. Cammin*, K. Taguchi, *Members, IEEE*

***Abstract*—Deforming organs pose an ongoing problem for tomographic image reconstruction. Time dependent deformations within a scanned body can lead to significant artifacts and obscure diagnostic information. In this paper we evaluate a motion-compensated filtered backprojection algorithm suited for non-rigid motion. The algorithm is applied to projections from a 4-D computer phantom with realistic heart motion. The motion vector field describing the true deformations is provided to the reconstruction algorithm. The motion compensation makes it possible to extend the scan angle significantly without being affected by motion artifacts. The reconstructed images are evaluated in terms of sharpness, uniformity, and noise.**

***Index Terms*—Computed tomography, filtered back-projection, motion compensation, non-rigid deformation.**

I. Introduction

IMAGING deforming objects such as the heart in diagnostic cardiac computed tomography (CT) is the subject of active research in computed tomography. This four-dimensional (3-D plus time) imaging problem can be divided into two sub-problems—motion estimation and motion compensation. The 4-D images can be reconstructed by solving the two sub-problems either sequentially or iteratively. For the motion compensation, the integration of motion information into the image reconstruction process has the potential to improve the quality of tomographic images of rapidly deforming objects.

While it is desirable to develop a mathematically exact, 4-D reconstruction algorithm for arbitrary non-rigid deformations, achieving this goal is unlikely to be possible, since the motion estimation is already a challenging, ill-posed problem. Therefore, the aim of our project is to develop an approximate 4-D reconstruction algorithm that provides clinically acceptable image quality. We plan to achieve this goal by developing an iterative algorithm that estimates the time-dependent motion vector field of the heart from the measured projection data and integrates it into the image reconstruction process. This paper concerns the motion compensated image reconstruction method.

A number of motion compensating algorithms for dynamically deforming objects have been proposed in the literature. However, usually a compromise needs to be made and the algorithm is either an approximation of the exact solution, or the allowed deformations are limited to a small subset of possible transformations, or both.

Exact algorithms can compensate mathematically exactly for affine transformations [1, 2] or deformations that maintain acquisition rays as straight lines [3]. In order to apply such exact methods for non-rigid deformations, one needs to approximate the given arbitrary deformation of a small region-of-interest (ROI) by a corresponding transformation formula, for example, by affine transformations. It is desirable, however, to directly use a motion vector field (MVF) which any motion estimation method can provide.

An empirical, approximate method proposed by Schäfer [4] does not use transformation models but instead tracks the deformation of the object during the filtered backprojection, using a given MVF directly. It has been shown that Schäfer's method with fan-beam geometry is a very good approximation of an exact method for affine transformations that are limited to isotropic scaling, rotation, and translation. Schäfer's method was later extended to fan parallel-beam geometry for diagnostic, cardiac helical CT by Stevendaal [5, 6]. Clinical cardiac images reconstructed by the extended Schäfer method with a gating window width of 40% of the R-R interval exhibited fewer motion artifacts than those reconstructed with no motion information with a gating window width of 22% of the R-R interval. The authors argued [6] that the signal-to-noise ratio of images will further improve if images are reconstructed with a gating window width of 100% of the R-R interval and with accurate MVFs over one heart beat. However, it was impossible to obtain accurate MVFs of the entire heart for one complete heart beat for clinical cases. It remains to be shown that Schäfer's method can maintain good image quality in case of large deformations if accurate MVFs are available.

The aim of this study was to assess the quality of images reconstructed by the original Schäfer algorithm with known true MVFs.

Manuscript received February 5, 2010. This work was supported in part by NIH grant R01 HL087918 and AHA grant BGIA 0865315E.

J.C. and K.T. are with the Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins School of Medicine, Baltimore, MD 21287, USA (phone: 410-502-3774; fax: 410-614-1060; e-mail: jcammin1@jhmi.edu).

II. Materials and Methods

A. Acquisition of the data sample

The dynamic 4-D XCAT phantom [7] was used to simulate a human torso, including a beating heart. The presence of iodine contrast agent was simulated in the ventricles and atria. The heart rate was set to 60 bpm and the start time $t_0 = 0$ was set to the end-diastole phase (start of the R-R phase). We restricted this study to the reconstruction of one 2-D slice and changed the XCAT program to allow for heart motion only in the x and y direction (perpendicular to the body axis). The *CT project* [7] program was used to generate a total of 2880 projections over four rotations with a rotation time of 1/3 s. The detector geometry was as follows: fan-beam projections onto a cylindrical detector with 768 channels and half fan-angle of $26.014°$. The distance from the x-ray source to the center of the phantom was 570 mm and the distance from the center of the phantom to the detector was 470 mm. The simulated x-ray spectrum was monochromatic with an energy of 80 keV. The choice for a single-energy spectrum was made so that the evaluation of the image quality of the algorithm under study is not impacted by artifacts from beam-hardening.

B. Motion vector field

The true MVF generated by the XCAT phantom was used as input to the reconstruction algorithm. The MVF described the displacement for each pixel within the heart boundary from the starting time t_0 to an arbitrary time t. A total of 100 frames with temporal spacing of 10 ms was produced to cover the duration of one complete cardiac cycle. The spatial distance of the vector grid in the x, y plane was 0.4 mm. Linear interpolation in space and time was used to obtain the motion vector for arbitrary times and spatial coordinates within the heart boundary. In future work, the MVF will be replaced by the estimated motion obtained from non-rigid image registration. An example for the applied MVF is shown in Fig. 1. The MVF in the XCAT phantom is defined on the myocardium and interpolated into the blood pool to obtain spatially smooth deformation information over the entire heart.

C. Time-dependent deforming object and scan

The time-dependent deforming object can be described by

$$f_t(\underline{x}_t) = f_t\left(\underline{x}_0 + \underline{v}_t(\underline{x}_0)\right) = f_0(\underline{x}_0), \qquad (1)$$

$$\underline{x}_t = \underline{x}_0 + \underline{v}_t(\underline{x}_0),$$

where $\underline{v}_t(\underline{x}_0)$ is the displacement vector from $\underline{x}_0$ to $\underline{x}_t$.

The circular source trajectory and fan-beam projections can be expressed as:

$$\underline{s}_t(\lambda) = \left(-R\sin\lambda, R\cos\lambda\right)^T, \qquad (2)$$

$$g_t(\lambda, \underline{\alpha}_t) = \int_0^\infty f_t\left(\underline{s}_t(\lambda) + l\underline{\alpha}_t\right) dl, \qquad (3)$$

where $\underline{\alpha}_t = (-\sin\alpha, \cos\alpha)^T$ is a unit vector along a ray and R is the distance from the source to the rotation axis. The source parameter is a monotonically increasing function of time, for example, $\lambda = at$.

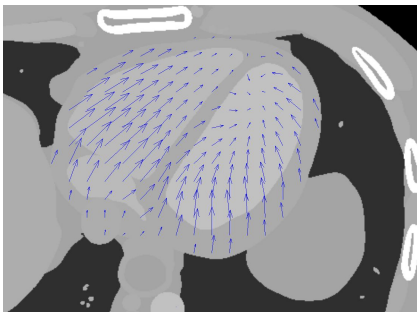


Fig. 1: Visualization of the XCAT phantom and the motion vector field. The image was generated at the end of diastole (t=0 ms). The vectors indicate the motion to a frame 0.5 seconds later. The duration of the heart cycle was 1 s.

D. Motion-compensated reconstruction

1) Algorithm

We used Schäfer's algorithm with a redundancy weighting as re-defined in [8].

$$f_0(\underline{x}_0) = \frac{1}{2}\int_{\Lambda_t} \frac{\left\|\underline{s}_t'(\lambda)\right\|}{\left\|\underline{x}_t - \underline{s}_t(\lambda)\right\|^2} \times \int_{-\infty}^{\infty} w(\lambda, \gamma_t)\cos\gamma_t' g_t\left(\lambda, \gamma_t'\right) h_R\left(\gamma_t - \gamma_t'\right) d\gamma_t' d\lambda \qquad (4)$$

where $w(\lambda, \gamma_t)$ is the redundancy weight (see next paragraph), $\cos\gamma_t'$ is the cosine weight, $g_t(\lambda, \gamma_{t'})$ are the projection data, and $h_R(\gamma_t - \gamma_t')$ denotes the ramp filtering.

2) Redundancy weight function

The performance of the reconstruction algorithm with non-rigid motion compensation was evaluated using two different weighting schemes: 1. Parker weight, which is currently the standard in most clinical protocols for cardiac imaging and uses projection data from a half-scan plus fan-angle; 2. overscan weight, which utilizes projections from a larger angular range. The weighting function is characterized by a symmetric trapezoidal shape and defined by λ_0, the starting view angle, λ_c, the center of the range used for reconstruction, λ_m, half of the width of the reconstruction window at half maximum, and λ_f, the "feathering" range at the edges of the trapezoid. For an illustration see Fig. 2.

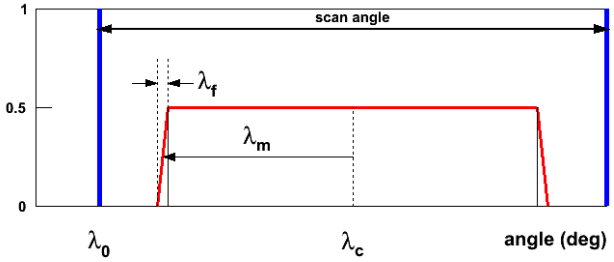


Fig. 2: Definition of the redundancy weight function.

III. Evaluation of the Image Quality

The image quality was evaluated in terms of sharpness of edges, uniformity, and noise using the following procedures.

A. Sharpness of edges

The generated phantom has uniform intensity values within a given tissue type and sharp boundaries between different tissues and organs. Therefore, a profile of intensity values perpendicular to the border between two different tissue types should follow a step function. Blurring of edges during the reconstruction due to motion results in a less steep step function (segment B in Fig. 3).

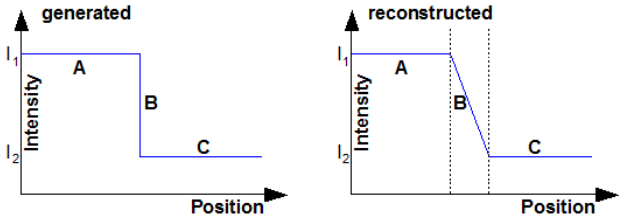


Fig. 3: Idealized intensity profile across two different tissue types. The generated image has very sharp edges. Edges in the reconstructed image are blurred.

The width of segment B of the step function is a measure of the sharpness of edges. To estimate the width, the three segments where fitted separately: Segment A and C with constant functions and segment B with a linear function. The intersections of the three fitted segments yield the width of segment B (for an example see Fig. 6).

B. Uniformity

The uniformity of the reconstructed images was measured as the mean value of pixel values within seven ROIs within the same tissue type (myocardium around the left ventricle). Then the normalized intensity difference Δ_I between the maximum and minimum intensity of the seven ROIs was calculated as $\Delta_I = (I_{max} - I_{min}) / \bar{I}$, where $\bar{I}$ denotes the average of the seven intensities. Δ_I is a measure of the image uniformity.

C. Noise

In order to compare the noise, Poisson noise was added to the projection data corresponding to 100,000 photons per projection ray. Then the mean μ and standard deviation σ of intensity values were calculated in several ROIs, and for each ROI the relative standard deviation σ / μ was taken as a measure of the noise.

IV. Results

A. Reconstructed images

Images were reconstructed in four ways: with and without motion compensation and with Parker weight and with overscan weight. The parameters for the overscan weight are $\lambda_0 = 90°$, $\lambda_c = 630°$, $\lambda_m = 360°$, and $\lambda_f = 30°$. The effective cardiac gating window width is therefore 17% for the Parker weight and 67% for the overscan weight. Fig. 4 shows the results for the images reconstructed with Parker weight. The motion compensation correctly reconstructs the phase at $t = t_0$ (compare to Fig. 1) whereas the non-compensated image exhibits strong motion artifacts. With the overscan weight (Fig. 5) the non-compensated image shows even larger blurring of the edges (in particular in the left ventricle). Again, the motion-compensation is able to restore the phase at time t_0 better.

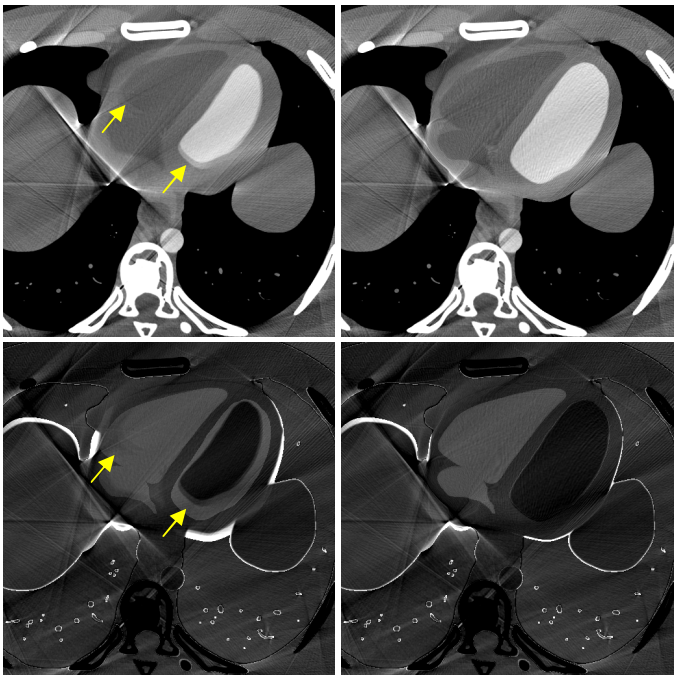


Fig. 4: Parker weight. Top: reconstructed images without (left) and with (right) motion compensation. Bottom: difference images between the generated phantom and the reconstructed images. Without (left) and with (right) motion compensation. The window-level for the display of the top row is W=0.1, L=0.2 (in Hounsfield units: W=590HU, L=180HU). The window-level for the subtracted images is W=0.05, W=0.15.

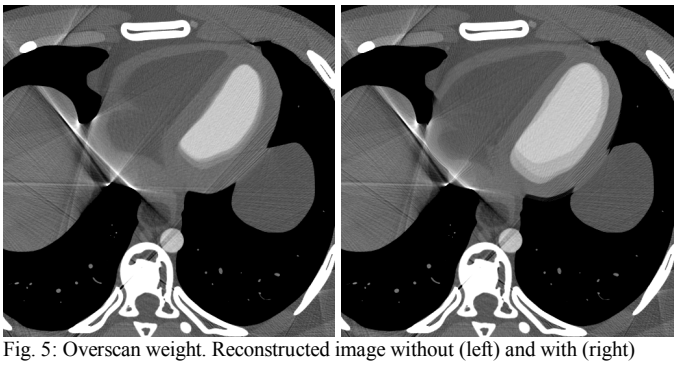


Fig. 5: Overscan weight. Reconstructed image without (left) and with (right) motion compensation.

B. Sharpness of edges

The sharpness of edges was evaluated from the profiles of two rectangular ROIs. The profile was calculated along the longer side of each rectangle. Table I shows that edge blurring with the overscan weight is 1.5 to 1.9 times smaller than with the Parker weight.

Table I: Step width of profiles for two regions of interest.

Width	non compensated	motion comp.	ratio
ROI 1	5.4	3.7	1.5
ROI 2	7.8	4.2	1.9

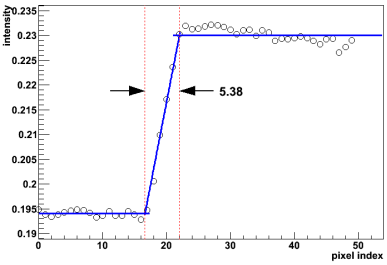


Fig. 6: An example for the fit to a profile to determine the sharpness of edges.

C. *Uniformity*

The calculated values of Δ_I obtained from seven ROIs are detailed in Table II. With the Parker weight, the uniformity of the compensated and non-compensated reconstruction agrees within 8%. Using the overscan weight, the motion compensated algorithm produces an image with a uniformity that is 39% better than without motion compensation. The uniformity using the overscan weight is on average 4.5 times better than the uniformity using the Parker weight.

Table II: Results of the uniformity measurement.

Δ_I	non compensated	motion comp.	ratio
Parker	6.6%	7.2%	0.92
overscan	1.8%	1.3%	1.39

D. *Noise*

The relative noise was measured in four regions-of-interest: ROI1in the left ventricle, ROI2 in the right ventricle, ROI3 in the myocardium between the left and right ventricle, and ROI4 in the myocardium of the outer side of the left ventricle. The results in Table III show that the relative noise decreases when the overscan weight is used. No significant differences are observed between compensated and non-compensated reconstruction. On average the noise using the overscan weight is about 30% smaller compared to the Parker weight.

Table III: Results of the noise measurement.

σ / μ	ROI1	ROI2	ROI3	ROI4
Parker, no mc	0.051	0.058	0.065	0.055
Parker, mc	0.053	0.057	0.063	0.051
Overscan, no mc	0.033	0.041	0.049	0.040
Overscan, mc	0.036	0.042	0.040	0.039

V. Conclusion

We evaluated a motion-compensated, filtered backprojection algorithm based on Schäfer's method using the dynamic XCAT phantom of a human torso with a beating heart. Applying the true motion vector field from the phantom, the algorithm reconstructed the image correctly at the target heart phase.

Edges were preserved better with the motion-compensated algorithm while the image uniformity is preserved (Parker weight) or even enhanced (overscan weight). There was no significant difference in the image noise.

Limitations: The algorithm was tested in 2-D+t with the true motion information and a monochromatic x-ray source. In the future, we intend to repeat the study on a 3-D+t volume. We will also replace the ground-truth MVF with deformation information obtained by image registration, and use a realistic x-ray spectrum.

VI. References

1. Roux, S., et al., *Exact reconstruction in 2D dynamic CT: compensation of time-dependent affine deformations.* Physics in Medicine and Biology, 2004. **49**(11): p. 2169-2182.
2. Taguchi, K. and H. Kudo, *Motion Compensated Fan-Beam Reconstruction for Nonrigid Transformation.* Medical Imaging, IEEE Transactions on, 2008. **27**(7): p. 907-917.
3. Desbat, L., S. Roux, and P. Grangeat, *Compensation of Some Time Dependent Deformations in Tomography.* Medical Imaging, IEEE Transactions on, 2007. **26**(2): p. 261-269.
4. Schafer, D., et al., *Motion-compensated and gated cone beam filtered back-projection for 3-D rotational X-ray angiography.* Medical Imaging, IEEE Transactions on, 2006. **25**(7): p. 898-906.
5. van Stevendaal, U., et al., *Motion-compensated reconstruction in helical cardiac CT*, in *the 9th international conference on fully three-dimensional reconstruction in radiology and nuclear medicine*, M. Kachelriess and F. Beekman, Editors. 2007: Lindau, Germany. p. 437-440.
6. van Stevendaal, U., et al., *A motion-compensated scheme for helical cone-beam reconstruction in cardiac CT angiography.* Medical Physics, 2008. **35**(7): p. 3239-3251.
7. Segars, W.P., et al., *Realistic CT simulation using the 4D XCAT phantom.* Medical Physics, 2008. **35**(8): p. 3800-3808.
8. Taguchi, K. and H. Kudo, *A simple motion tracking backprojection for a class of affine transformation.* Medical Imaging 2008: Physics of Medical Imaging, 2008. **6913**: p. 66.

Spectroscopic X-ray CT for Improved Explosives Detection

Timothy A. White[1], Frédéric Noo[2]
[1]Pacific Northwest National Lab, Richland, WA
[2]University of Utah, Salt Lake City, UT

***Abstract*—Pixelated photon-counting detectors have been investigated for improved tissue discrimination in medical-imaging applications. In this work, we investigate the challenges and potentials of using spectroscopic detectors for discrimination of benign and threat (explosive) material in security-screening applications. We assume a Bremsstrahlung source and an idealized detector with uniform energy windows and unbounded maximum count rate. The simulated object consists of materials commonly found in security-screening applications, which differ significantly in Z_{eff} and density from medical-imaging applications. A decomposition of sinogram data using mass attenuation coefficients as basis functions is performed. It is shown that judicious choice of two to three basis functions provide useful material discrimination, but using more basis functions will be challenging because of the non-linearity of the decomposition process.**

***Index Terms*—computed tomography, spectroscopic detector, photon counting, security screening, explosive detection**

I. Introduction

The Transportation Security Administration (TSA) inspects over 500 million pieces of luggage each year [1]. Baggage inspection is automated and conducted by explosive detection systems (EDS), primarily computed-tomography (CT) systems tasked with distinguishing explosives from benign material. Security screening differs from medical imaging in both the composition of the scanned object and the threat-detection task, e.g., contraband materials versus tumors. In one respect the imaging problem is similar in both cases: material discrimination – determining whether the gray level in an image represents an explosive or benign material (a tumor or healthy tissue) – is hampered by the broad energy spectrum of a Bremsstrahlung x-ray beam that is used in contemporary CT systems.

Dual-energy computed tomography has been used in both medical and security applications to enhance this discrimination task [2, 3]. In this approach, a source (or sources) with two endpoint energies, or a detector sensitive to different energy ranges collect two sets of sinograms. These data are used to decouple physical parameters of the attenuation coefficient, for instance, density and effective atomic number (Z_{eff}), in the final images. The high flux from a Bremsstrahlung source, coupled with the requirement that imaging be performed rapidly, requires the use of current-integration detectors with which any spectral information is lost. ("Current-integration detectors" are defined here to be radiation detectors in which the output signal is proportional to the total energy deposited within a (short) integration window.) Recent work in the medical-imaging community has asked whether multiple energy windows would assist in diagnostic utility [4-6]. Detectors that would enable this mode of detection are commonly called *photon-counting detectors*. Detection rates in room-temperature semiconductors (CdTe) of up to $5\text{x}10^6$/sec/mm^2 have been reported. This rate may be too low for conventional CT, but it is assumed that the capabilities of these detectors will continue to increase. A small-animal imaging system based on CdTe detectors has already been demonstrated [7].

The purpose of this paper is to examine the potential offered by spectral information in security CT screening. A simulated fan-beam CT geometry using a Bremsstrahlung source is coupled to an idealized detector with an unbounded maximum count rate and uniform energy. The question is how to use the information from such a detector to better discriminate threat from benign material in a security CT scan.

II. Method

A. Data Model

Projection data is modeled as follows. Let $S(E)$ represent the normalized incident (Bremsstrahlung) spectrum and $N_{emitted}$ represent the incident number of photons. Then the number of photons that pass through the object unattenuated can be written as

$$N_{out}(E) = N_{emitted} S(E) e^{-\int_L \mu(\vec{x},E)dl} \quad (1)$$

where L is the pathlength through the object along the line from the source position to a detector pixel and $\mu(\vec{x},E)$ is the position and energy dependent linear attenuation coefficient of the material in the object. The detector response to the input flux can be modeled as an efficiency function, $D(E)$, and an energy-weighting function, $R_i(E)$, where the subscript i is the index to each energy bin. The energy-weighting function is the integral of the energy response of the detector over an

We gratefully acknowledge the Laboratory Directed Research and Development program for funding this research through the Initiative for Explosives Detection at Pacific Northwest National Laboratory. Pacific Northwest National Laboratory is a multiprogram national laboratory operated by Battelle Memorial Institute for the US Department of Energy under Contract DE-AC05-76RL01830.

energy window set by upper and lower limits and will be approximated by a Gaussian in this work. The number of detected photons in each energy bin can then be written as

$$N_i = \int_0^{E_{max}} R_i(E) D(E) N_{out}(E) dE. \quad (2)$$

The log-corrected data can be written as

$$g_i = -\ln\left(\int_0^{E_{max}} w_i(E) \frac{N_{out}(E)}{N_{emitted} S(E)} dE\right) = -\ln\left(\int_0^{E_{max}} w_i(E) e^{-\int_L \mu(\vec{x},E) dl} dE\right), \quad (3)$$

where

$$w_i(E) = \frac{R_i(E) D(E) S(E)}{\int_0^{E_{max}} R_i(E) D(E) S(E) dE} \quad (4)$$

can be considered to represent the relative weights of each energy bin and is noise free.

B. Multiple-Energy CT Imaging Methodology

Dual- and multiple-energy CT inversion methods decompose the linear attenuation coefficient for a material into independent functions of space and energy: $\mu(\vec{x},E) = \sum_j a_j(\vec{x}) f_j(E)$.

This can be considered to be an expansion of the linear attenuation coefficient into a set of energy-dependent basis functions with spatially dependent coefficients. Substituting this expression into the log-corrected data equation, it can be seen that

$$g_i = -\ln\left(\int_0^{E_{max}} w_i(E) e^{-\sum_j A_j f_j(E)} dE\right), \quad (5)$$

where $A_j = \int_L a_j(\vec{x}) dl$ are energy-independent sinograms. Given a set of $i = 1 \ldots N$ measurements and $j = 1 \ldots M$ basis functions, the game is to invert the set of nonlinear equations defined by (4) to solve for the coefficients (energy-independent sinograms); this step is referred to as the decomposition into basis functions. A CT reconstruction is then performed on line integrals. This results in M reconstructed images, whose interpretation depends on the choice of basis functions.

For imaging biological tissue for which there are no materials with atomic number greater than 20, there is an argument that the basis functions should be based on the dominant attenuation mechanisms: photoelectric absorption (PE) and Compton scatter. Both coefficients can then be described as functions of electron density, and the coefficient for the photoelectric effect is a function of Z_{eff} as well. This is often referred to as the physics-based decomposition approach. An alternative is to use the mass attenuation coefficients (MACs) of two dissimilar materials; in this case, the coefficients will be functions of the relative density. It can be shown [7] that a set of two MAC-based basis functions can be derived from the physics-based (PE and Compton) approach.

For an imaging system that has more than two available energy bins, it is possible to consider using more than two basis functions. At first glance, it would seem that more basis functions would be useful for better discrimination, however attenuation coefficients are slowly varying as a function of element (Z), and so it may be difficult to find many orthogonal (or nearly orthogonal) basis functions for the energy range used in security screening of baggage. Indeed, some have argued that the principle attenuation mechanisms (PE and Compton scatter) limit the number of orthogonal basis functions to two. Schlomka, *et al.* [8], have used a mix of physics-based and MAC-based basis functions that exploits the K-edges of iodine and gadolinium to develop a four-function decomposition scheme for a small animal imager. Choice of the basis functions for the security screening will be a challenge.

The reconstructions from the basis-function coefficients are not weighted averages of the linear attenuation coefficients, but are proportional to the material density (if MAC basis functions are used) or material density and Z_{eff} for the physics-based cases. In an additional image-processing step, the energy dependent attenuation coefficients can be "reconstructed" as a function of position in the reconstructed image using the reconstructed basis-function coefficients (in the basis-function decomposition equation).

C. Materials and Basis Functions

For the security-screening experiment, a set of materials was simulated that spans the range of material types that could be found in baggage. These objects were developed in an earlier project focused on air-cargo screening [9]. It is important to note that these materials span a larger range of density and Z_{eff} than in medical imaging. The materials used are summarized in Table 1 and the mass attenuation coefficients of these compounds using elemental cross sections from the XCOM database are plotted in Fig. 1.

Material	*Density [g/cm^3]*	*Z_{eff}*
Explosives (EXP)	1.76	7.44
Wearing Apparel (WA)	0.16	22.06
Water (H_2O)	1.0	7.42
Machine Parts (MP)	0.08	25.89
Electronics (EE)	0.24	34.64
Alcohol (CH)	0.79	6.61
Printed Material (PM)	0.8	7.38

Table 1 Material descriptions for the simulation studies.

As seen in Fig. 1, the low-Z materials have similar, almost scaled, attenuation plots. Although it is difficult to see, the alcohol (chemicals) curve crosses over the other three low-Z curves in the range 34-100keV. The higher-Z materials have a stronger photoelectric contribution in this energy range and have a few characteristic K-edges: a Br line at 13keV in wearing apparel (bromine is found in polyester) and a Pb line at 84keV in electronics. The density that is assumed for the high-Z materials is a volume-averaged density; in the initial simulations, rather than detailed simulation of actual objects, larger, lower-density objects are being used. The resulting reconstructions should scale with density and the effects on discrimination will be discussed below. A simple object

containing all 7 materials was used to test the material decomposition.

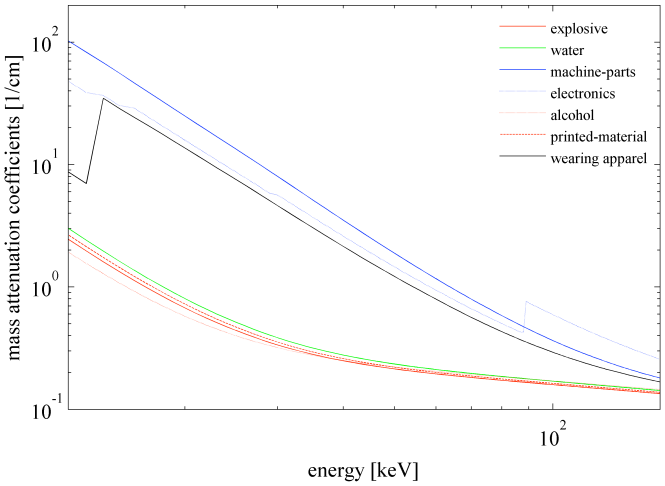


Figure 1 Mass attenuation coefficients of materials in Table 1. [Color versions of the plots are available in the online version.]

D. *Experimental Setup*

A single-slice fan-beam system using a flat detector, a 1m source-to-detector distance and a nominal magnification of two was simulated. The detector was simulated to have 400 pixels on a 0.15mm pitch, and an energy efficiency $D(E)=1$. Five hundred equally spaced projections on a circle were simulated. Energy-weighting functions were modeled as a set of Gaussian with mean energies at 21, 51, 81, 111, and 141keV with a standard deviation of 10keV. The energy spectrum, energy windows, and a normalized mass attenuation curve (for explosive) are shown in figure 2.

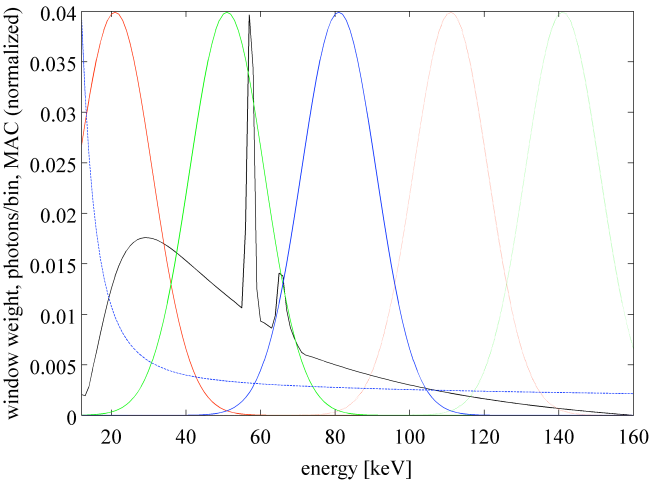


Figure 2 The energy spectrum and energy windows used in the source and detector model in the simulations. The mass attenuation coefficient of the explosive material is also shown.

A ray-tracing simulation was used to calculate the line integrals through the object in 1keV energy steps [$S(E)$ and $\mu(E)$ were discretized into 1keV bins]. Poisson noise was added to the spectra that passed through the object, $N_{out}(E)$, before summation into the energy windows, g_i. The number of emitted photons, $N_{emitted}(E)$, were scaled to represent a total photon flux at the detector face (assumed to be 40k photons in all of the simulations below), and noise was added to each 1keV wide bin independently using $N_{out}(E_k)$ as the mean. The decomposition of the windowed data, g_i, into line integrals of the basis-function coefficients, A_j, was performed as a least-squares optimization using $w_i(E)$ and $f_j(E)$ as prior information. Mass attenuation coefficients were used as the basis functions. Images of the basis-function coefficients were reconstructed using a filtered backprojection algorithm.

The simulated object was a uniform disk (24cm diameter) of the wearing-apparel material, meant to simulate the clothing "background" in an image. Inserted in this uniform disk were six uniform cylinders (2cm diameter) consisting of the other six materials in Table 1. The basis functions used are the mass attenuation coefficients of some subset of these seven materials. The imaging task was to be able to distinguish the explosive material (threat) from the other (benign) materials.

The hypothesis is that the inversion process from g_i into A_j separates the object contents into bins according to materials. For instance, if the object consisted of two materials (e.g., wearing apparel and explosive) and MACs for those materials were used as the basis functions, the reconstructed values of the coefficients for the explosive would be equal to the density of explosive in the region of the explosive and zero everywhere else and the reconstruction of the wearing-apparel coefficients would be the opposite. In this context, a third different material in the object would end up correlated with the basis function with the most similar MAC, or "shared" between basis functions. Then discrimination between materials would have to be based on relative intensity of the reconstructed pixels (equal to density); for example the image corresponding to the explosive basis function would be non-zero for regions of the object that did not contain explosive material, but the density would be incorrect. As more materials are added to the object, there will be more sharing between basis functions and more potential that the reconstructed densities will confuse the discrimination process.

III. Experimental Results

Reconstructed images using two basis functions (explosive and wearing apparel) are shown in Fig. 3. The average and standard deviation of the pixel intensities in regions of interest within each material are summarized in Table 2.

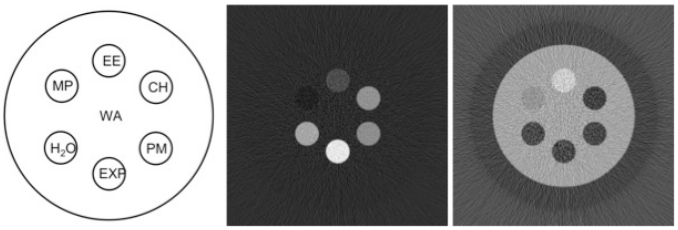


Figure 3 Reconstructed images of the phantom using two the mass attenuation coefficients of explosive (EXP) and wearing apparel (WA) as the basis functions. The drawing on the left indicates the material distribution. The image in the middle is the reconstruction using explosive as the basis function. The image on the right is the reconstruction of the wearing apparel basis function. The images have been scaled to maximum and minimum values independently.

	Density		
		Reconstructed	
Material	*Simulated [g/cm³]*	*EXP Basis Function*	*WA Basis Function*
Explosives (EXP)	1.76	1.75 ± 0.09	0.01 ± 0.03
Wearing Apparel (WA)	0.16	-0.01 ± 0.08	0.16 ± 0.03
Water (H_2O)	1.0	1.05 ± 0.07	0.01 ± 0.03
Machine Parts (MP)	0.08	-0.07 ± 0.05	0.13 ± 0.02
Electronics (EE)	0.24	0.23 ± 0.07	0.24 ± 0.03
Alcohol (CH)	0.79	0.85 ± 0.07	0.00 ± 0.02
Printed Material (PM)	0.8	0.82 ± 0.06	0.00 ± 0.02

Table 2 Simulated and reconstructed densities of the materials in the phantom using the two-basis-function material separation. The reconstructed values are calculated from a rectangular region of interest centered in each material. The standard deviation reflects noise in the projection data as well as discretization errors.

There are a few results worth noting:

- The explosive material and the wearing apparel are reconstructed in the proper basis-function "bins" and with the correct density.
- Other low-Z materials (water, alcohol, printed material) end up in the explosives bin with approximately the correct densities. This is expected because the mass attenuation coefficients are so similar. Discrimination is still possible because the densities are sufficiently different from explosive.
- The high-Z materials (machine parts and electronics) are shared between the two bins. Machine parts has a negative contribution from the explosives basis function, while electronics is reconstructed with about the correct density in each bin.

Also note that if the imaging task is to detect explosives, then the only image of interest is the reconstruction of the coefficients of the explosives basis function. Additional basis functions could be used to move some of the other materials (e.g., electronics) out of the explosives image.

Using additional basis functions in the decomposition process should improve the separation between materials. Reconstructions using three basis functions are shown below. In the first example, the MAC for electronics was added to explosive and wearing apparel materials as the third basis function. Results of the decomposition and subsequent reconstruction are shown in Fig. 4 and Table 3.

There are two noteworthy changes for this case. The electronics is now partitioned into the electronics basis-function bin correctly. However, this occurs at the expense of more uncertainty in the estimated density in each of the other basis-function images. The electronics component has been removed from the explosives image, perhaps simplifying the explosives-detection task.

In another example, the MAC for alcohol was used as the additional basis function. These results are shown in Fig. 5 and Table 4. The material separation is dramatically worse for the case where alcohol replaces electronics as the third basis function. All of the low-Z materials share weight between the MACs for explosive and alcohol, and the mean density values are not even close to correct.

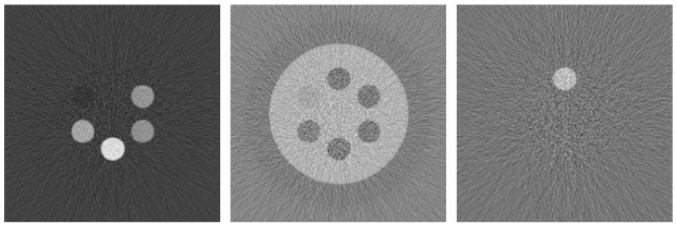


Figure 4 Reconstructions using explosives (left), wearing apparel (middle) and electronics as the basis functions. Images have been scaled independently.

	Density			
	Simulated	*Reconstructed*		
Material	*[g/cm³]*	*EXP Basis Ftn*	*WA Basis Ftn*	*EE Basis Ftn*
EXP	1.76	1.71 ± 0.17	0.00 ± 0.09	0.01 ±0.010
WA	0.16	-0.01 ± 0.16	0.16 ± 0.07	0.01 ± 0.09
H_2O	1.0	1.05 ± 0.14	0.01 ± 0.07	0.00 ± 0.08
MP	0.08	-0.08 ± 0.10	0.12 ± 0.06	0.01 ± 0.07
EE	0.24	-0.01 ± 0.15	0.00 ± 0.07	0.24 ± 0.09
CH	0.79	0.85 ± 0.12	-0.01 ± 0.07	0.00 ± 0.07
PM	0.8	0.82 ± 0.12	0.00 ± 0.07	0.00 ± 0.08

Table 3 Reconstructed densities for the 3-basis-function case using explosive, wearing apparel, and electronics as the basis functions.

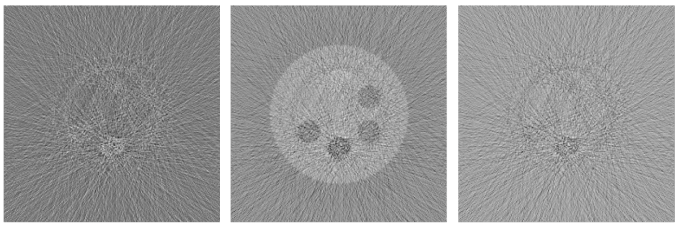


Figure 5 Reconstructions using explosives (left), wearing apparel (middle) and alcohol as the basis functions. Images have been scaled independently.

	Density			
	Simulated	*Reconstructed*		
Material	*[g/cm³]*	*EXP Basis Ftn*	*WA Basis Ftn*	*CH Basis Ftn*
EXP	1.76	10.23 ± 30.04	-0.07 ± 0.27	-7.79 ± 27.54
WA	0.16	-2.97 ± 19.72	0.19 ± 0.18	2.72 ± 18.08
H_2O	1.0	5.34 ± 24.59	-0.03 ± 0.22	-3.93 ± 22.55
MP	0.08	-2.86 ± 12.94	0.16 ± 0.12	2.56 ± 11.86
EE	0.24	4.54 ± 22.78	0.20 ± 0.21	-3.95 ± 20.89
CH	0.79	2.96 ± 20.15	-0.02 ± 0.18	-1.94 ± 18.48
PM	0.8	1.93 ± 21.21	-0.01 ± 0.19	-1.02 ± 19.45

Table 4 Simulated and reconstructed densities for the explosives, wearing apparel and alcohol basis functions

The mean coefficient values (Table 5) can be used to "reconstruct" the linear attenuation coefficients of the seven materials as shown in Fig. 6. These fits are fairly good

qualitatively, an indication that the basis-function decomposition represents the linear attenuation coefficient (even in the presence of noise), but there was some confusion about which of the two similar basis functions should contribute to a material. Plots of the reconstructed linear attenuation coefficients for the other two cases (EXP/WA and EXP/WA/EE basis functions) show qualitatively as good or better fits to the attenuation coefficients.

The challenge of using a third (and a fourth or fifth) basis function in the decomposition process is to find features of the attenuation coefficient that may be unique in a multi-dimensional space. The MAC for alcohol was chosen in the third example because it crosses the MAC of the other low-Z materials, but that feature was apparently not unique enough. On the other hand, the distinguishing feature of the EE MAC is the lead absorption edge; this feature cannot be fit by a linear combination of the other basis functions and allows the EE material to be separated from the other materials.

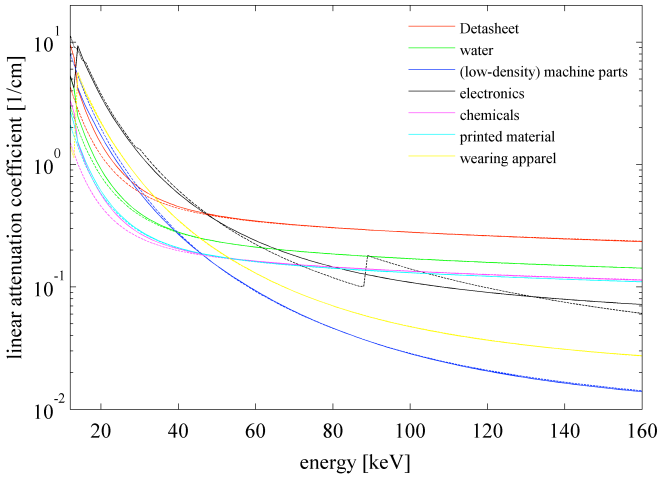


Figure 6 Reconstructed basis functions using coefficients from the mean values of the ROIs in Table 4 for the experiment using the MACs of explosive, wearing apparel, and alcohol. The solid lines are the simulated linear attenuation coefficients and the dashed lines are the fits from the reconstructed data. The fits are fairly good, although worse at lower energies. Note the inability to adequately fit the lead k-edge absorption line in the electronics material type.

IV. Discussion

The challenge of using multiple energy windows for material decomposition of x-ray CT data is the non-linearity in the decomposition process. It is likely that there are multiple local minima in the inversion process, and this results in fits that, while they correctly fit the attenuation coefficient, do not provide material discrimination in the image. For the linear case, for example, if the total attenuation is very small, we have observed that the decomposition performs much better.

We have shown that a 2-basis-function decomposition can be used to discriminate threats (explosives) from benign material, although higher Z_{eff} material may corrupt the imaging task when that material falls into the threat bin. We showed that discriminability may improve with three basis functions provided that the additional function has unique features (and that the increased uncertainty in pixel values did not add to the confusion). Using three or more basis functions to improve discriminability will depend upon finding unique features and defining a robust decomposition process.

References

[1] From the Transportation Security Administration website, http://www.tsa.gov/research/screening_statistics.shtm, accessed 5/24/2010.

[2] R.E. Alvarez and A. Macovski, "Energy selective reconstructions in x-ray computed tomography," *Phys Med Biol* 21(5), pp. 733-744, 1976.

[3] Z. Ying, R. Naidu, and C. Crawford, "Dual energy computed tomography for explosive detection," *Journal of X-Ray Science and Technology* 14, pp. 235-256, 2006.

[4] K. Taguchi, S. Srivastava, H. Kudo, and W.C. Barber, "Enabling photon counting clinical x-ray CT", *2009 IEEE Nuclear Science Symposium Conference Record* M13-45, 2009.

[5] E. Roessl and R. Proksa, "K-edge imaging in x-ray computed tomography using multi-bin photon counting detectors," *Phys. Med. Biol.* 52, pp. 4679-4696, 2007.

[6] K. Taguchi, M. Zhang, E.C. Frey, J. Xu, W.P. Segars, and B.M.W Tsui, "Image-domain material decomposition using photon-counting CT," *Medical Imaging 2007: Physics of Medical Imaging*, Proc. Of SPIE Vol. 6510, 2007.

[7] L. Lehmann, R. Alvarez, A. Macovski, W. Brody, N. Pelc, S. Riederer and A. Hall, "Generalized image comninations in dual-KVP digital radiography," *Medical Physics*, *, p. 659, 1981.

[8] J.P. Schlomka, E. Roessl, R. Dorscheid, S. Dill, G. Martens, T. Istel, C. Baumer, C. Herrmann, R. Steadman, G. Zeitler, A. Livne, and R. Proska, "Experimental feasibility of multi-energy photon-counting K-edge imaging in pre-clinical computed tomography," *Phys Med Biol* 53, pp. 4031-4047, 2008.

[9] R.C. Runkle, T.A. White, E.A. Miller, J.A. Caggiano, and B.A. Collins, "Photon and neutron interrogation techniques for chemical explosives detection in air cargo: A critical review," *Nucl Inst and Meth A*, 2009.

Combining the Virtual Fanbeam and Differentiated Backprojection Methods of Region-of-Interest Reconstruction

C. Mennessier, R. Clackdoyle, and M. Defrise

I. Introduction

INTEREST in 2D Region-of-Interest (ROI) tomographic reconstruction has increased since 2002, when it appeared possible to handle some incomplete data problems. By "tomographic reconstruction" we mean here only exact reconstruction methods. There exist two main approaches : the Virtual FanBeam (VFB) [1], [2] and the Differentiated BackProjection (DBP) methods [3]. The methods are exact because they produce the (same) correct answer when presented with ideal noise-free data. However, the approaches are not equivalent in the sense that, from a noisy sinogram, they do not lead to the same reconstruction. In addition, some incomplete data problems can be solved by one method but not the other.

The fact that there are two non-equivalent methods for ROI reconstruction means that data are redundant. The question is, can we exploit the redudancy to improve the quality of the reconstructed image? The goal of the present work is to combine the DBP and the VFB methods to improve the variance of the combined image. We consider the particular incomplete data problem that occurs in the bi-lateral truncation case (see section II), which can be handled by both methods.

II. Theory

Let $f(x)$ be an image function and $p(\phi, s)$ the corresponding sinogram i.e. the Radon transform of $f(x)$ for the parallel geometry. As both VFB and DBP methods correspond to linear operators, the value of the reconstructed functions $f^{VFB}(x)$ and $f^{DBP}(x)$ at each point x can be obtained as the inner products of $p(\phi, s)$ with specific weight functions $w_x^{VFB}(\phi, s)$ and $w_x^{DBP}(\phi, s)$, for the VFB and the DBP methods respectively [4]. More precisely we have :

$$f^{VFB}(x) = \int d\phi \int ds\ p(\phi, s) w_x^{VFB}(\phi, s) = < p, w_x^{VFB} >$$
$$f^{DBP}(x) = \int d\phi \int ds\ p(\phi, s) w_x^{DBP}(\phi, s) = < p, w_x^{DBP} > \quad (1)$$

For each x in the ROI where both VFB and DBP can be applied, we look for the scalar λ_x which is such that the combined function

$$f^{\lambda}(x) = \lambda_x f^{VFB}(x) + (1 - \lambda_x) f^{DBP}(x) \quad (2)$$

optimizes the variance of $f^{\lambda}(x)$.

Let us denote v_x^{VFB}, v_x^{DBP} and v_x^{λ} respectively the variances of $f^{VFB}(x)$, $f^{DBP}(x)$ and $f^{\lambda}(x)$. We put v^p for the variance of the sinogram and we assume that the noise covariance matrix is diagonal. Using (1) one has :

$$v_x^{VFB} = < v^p, \left(w_x^{VFB}\right)^2 >,\ v_x^{DBP} = < v^p, \left(w_x^{DBP}\right)^2 > \quad (3)$$
$$v_x^{\lambda} = < v^p, \left(w_x^{\lambda}\right)^2 > \quad (4)$$

From equations (1) and (2) the weight function for the combined reconstruction is $w_x^{\lambda} = \lambda_x w_x^{VFB} + (1 - \lambda_x) w_x^{DBP}$. Thus, developing $\left(w_x^{\lambda}\right)^2$ in (4) we obtain:

$$\left(w_x^{\lambda}\right)^2 = \lambda_x^2 (w_x^{VFB} - w_x^{DBP})^2 + 2\lambda_x \left(w_x^{VFB} w_x^{DBP} - \left(w_x^{DBP}\right)^2\right) + \left(w_x^{DBP}\right)^2 \quad (5)$$

For each point x, we determine the value λ_x^* that minimizes the variance v_x^{λ}. Equations (4) and (5), give a quadratic expression in λ_x which is minimized at

$$\lambda_x^* = \frac{< v^p, \left(w_x^{DBP}\right)^2 - w_x^{VFB} w_x^{DBP} >}{< v^p, (w_x^{VFB} - w_x^{DBP})^2 >} \quad (6)$$

The optimal variance for f^{λ} is deduced by putting equation (6) in equation (4) :

$$v_x^{\lambda^*} = v_x^{DBP} - \frac{\left(< v^p, w_x^{VFB} w_x^{DBP} - \left(w_x^{DBP}\right)^2 >\right)^2}{< v^p, (w_x^{VFB} - w_x^{DBP})^2 >} \quad (7)$$

Similarly

$$v_x^{\lambda^*} = v_x^{VFB} - \frac{\left(< v^p, w_x^{VFB} w_x^{DBP} - \left(w_x^{VFB}\right)^2 >\right)^2}{< v^p, (w_x^{VFB} - w_x^{DBP})^2 >} \quad (8)$$

From equations (7) and (8), we check that the variance of f^{λ^*} is smaller than or equal to the variances of f^{VFB} and f^{DBP}, with λ_x^* given by equation (6). Since both the VFB and DBP algorithms give exact reconstruction within the ROI, so does their linear combination. Note that even though the combination in equation (2) is linear, the optimal coefficient (6) does not depend linearly on the data when the noise variance is object dependent, and in this situation the proposed method is not a linear reconstruction algorithm.

The remaining question is to evaluate the magnitude of the gain in variance, and this has been done in preliminary numerical simulations. Evaluating the variances requires computation of the weight functions.

C. Mennessier is with the Chimie Physique Electronique and Laboratoire Hubert Curien, Lyon, France (e-mail: mennessier@cpe.fr).

R. Clackdoyle is with the Laboratoire Hubert Curien, Saint Etienne, France (e-mail: rolf.clackdoyle@univ-st-etienne.fr).

M. Defrise is with the University Hospital UZ-VUB, Brussels, Belgium (e-mail: mdefrise@vub.ac.be).

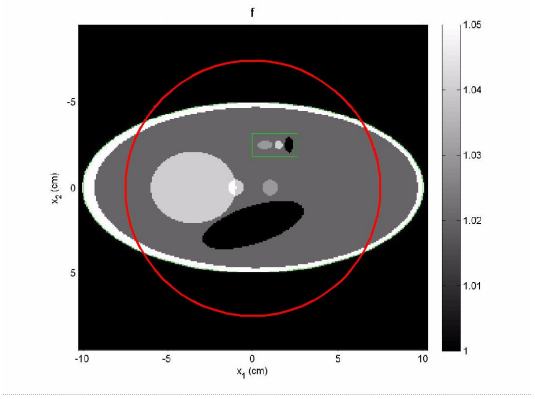


Fig. 1. The bilateral truncation (red circle: the field of view) and the ROI (green rectangle).

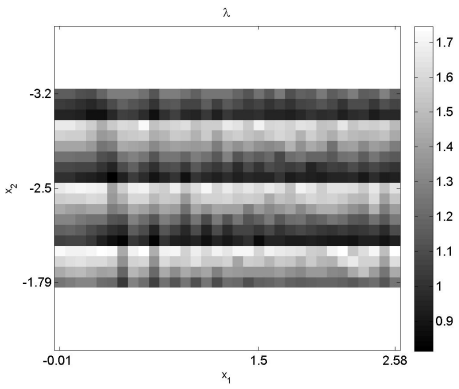


Fig. 2. The estimated λ_x^* coefficients, in the ROI.

III. Numerical simulations

We use a modified Shepp-Logan phantom for $f(x)$, with main ellipse of size $20cm \times 10cm$ (see figure 1). Its Radon transform is analytically calculated for each ray crossing a field-of-view defined by a disk, centered at the ellipse center, with radius equal to $7.5cm$. Rays not crossing that disk are not measured; this problem is known as the bi-lateral truncation problem. Even though almost the entire field-of-view can be reconstructed with both VFB and DBP from these bi-lateral truncated data, we focus on a small rectangular ROI, around the 3 small ellipses. The corners of this rectange are at $(-0.01, -1.79)$, $(-0.01, -3.2)$, $(2.58, -1.79)$, $(2.58, -3.2)$ (see figure 1). The reconstructed ROI was discretized into 33 x 19 pixels of size $0.08cm \times 0.08cm$.

In a first step we compute the weight functions w_x^{VFB} and w_x^{DBP}. For the VFB method we use a segment of the upper half-circle for the virtual trajectory to process the truncated parallel projections. The DBP method is based on the fact that the 1D finite Hilbert transform can be computed, and then inverted, along well chosen lines. In our case we choose vertical lines. The weight functions are actually distributions, and the analytical expressions in the previous section must be discretized and regularized in such a way that the spatial resolutions of the two methods are as close as possible. This is difficult because the two methods are very different and because the singularities of w_x^{VFB} and w_x^{DBP} require sophisticated numerical tools.

In this exploratory work, we discretize the data on a sinogram of size $n_\phi = 245$ and $n_s = 200$. For each sinogram bin (ϕ_0, s_0) we compute the weights $w_x^{VFB}(\phi_0, s_0)$ and $w_x^{DBP}(\phi_0, s_0)$ as follows : we reconstruct, using the VFB and the DBP algorithms, a sinogram that is equal to zero for all bins except for bin (ϕ_0, s_0) which is set equal to one. In the continuous description of section II, this is equivalent to inserting $p(\phi, s) = \delta(\phi - \phi_0, s - s_0)$ into equation (1) and then setting $w_x^{VFB}(\phi_0, s_0) = f^{VFB}_{(\phi_0, s_0)}(x)$ (and similarly for the DBP method). This procedure must be repeated for each of the sinogram bins (ϕ_0, s_0) so it requires 245×200 reconstructions for each method. The VFB method is much more time-consuming than the DBP method. The choice of sinogram size $n_\phi \times n_s$ was based on a study to balance reconstructed image quality against total time to generate the weight functions. For the simulation described above, 3 days were needed for the VFB weight function reconstruction using matlab codes on a 2.39 GHz and 1.98 Gb RAM, laptop.

In a second step, we used the weight functions to evaluate the optimized λ_x^* coefficients using equation (6). The sinogram variance was taken to be $v^p(\phi, s) = ((0.001/M)p(\phi, s))^2$ where $M = \max\{p(\phi, s)\}$. The λ_x^* values are displayed in figure 2. The minimum and maximum values of λ_x^* for the 33×19 values of the ROI were 0.8140 and 1.7459. In each row shown in figure 2, the values varied by less than 0.339. The sawtooth-like oscillations of the lambda values in the ROI are not currently understood.

In a third step, the variances of $f^{VFB}(x), f^{DBP}(x)$ and $f^{\lambda_x^*}(x)$ in the ROI were evaluated by performing $N = 50000$ reconstructions from noisy sinograms $p^n(\phi, s)$ defined by :

$$p^n(\phi, s) = p(\phi, s)(1 + k\ X), n = 1, ..., N \qquad (9)$$

where X is the standard centered gaussian distribution and $k = 0.001/M$ for $M = \max\{p(\phi, s)\}$. The f^λ function is calculated as given in equation (2). The reconstructions are shown in figures 3,4,5. The horizontal profiles indicate significant distortions that our simulations (not shown) suggest are due to the coarse sinogram-sampling used for time-computation reasons. The variances are shown in figure 6. For this particular simulation, the VFB variance is nearly 30% smaller than the DBP variance. The f^λ variance is very close to that of VFB, 2% smaller on average, and the improvement varies between 0% and 9% in the ROI, as shown in figure 6. Using the variances obtained from the N reconstructions, the two inequalities $v^\lambda < v^{VFB}$ and $v^\lambda < v^{DBP}$ hold for 98% of the 33×19 pixels in the ROI. The 2% of pixels for which the variance is not improved is presumed to be due to statistical uncertainty and would be removed by increasing N. The inequalities are verified everywhere in the ROI if the theoretical variances are used i.e. the variances obtained from equations (3) and (7). The results show that theoretically as well as experimentally, the combined method variance is smaller than the f^{VFB} and f^{DBP} variances. The values of both theoretical and experimental variances inside a 4×5 rectangle centered on the middle ellipse (see red rectangle in figure 5) are given in the tables I, II, III.

TABLE I
THEORETICAL AND EXPERIMENTAL VARIANCES : $10^6 v^{VFB}$

6.273 5.783 6.243 5.656	6.275 5.727 6.284 5.637
6.248 5.780 6.171 5.822	6.223 5.755 6.188 5.857
6.228 5.884 6.138 5.954	6.158 5.898 6.173 6.007
6.192 6.077 6.057 6.345	6.183 6.028 6.005 6.377
6.108 6.400 6.014 6.409	6.161 6.457 6.043 6.483

TABLE II
THEORETICAL AND EXPERIMENTAL VARIANCES : $10^6 v^{DBP}$

7.831 7.813 7.885 7.734	7.843 7.736 7.982 7.724
7.988 7.797 7.945 7.755	7.947 7.733 7.965 7.825
8.080 7.839 8.040 7.997	7.958 7.835 8.044 8.020
8.101 7.946 8.028 8.065	8.075 7.871 7.963 8.122
8.153 8.100 7.994 8.174	8.216 8.124 8.042 8.274

IV. CONCLUSION

The goal of this work was to exploit the non-equivalence of the DBP and VFB methods, to improve the image quality. We proposed to combine linearly the VFB and the DBP methods to improve the reconstructed image variance and we gave the expression of optimal coefficient λ_x^* which minimizes the variance of the combined function f^{λ^*} in each point x. Since the VFB and the DBP methods are not equivalent (they do not use the same data subset [4]), it is very unlikely that the optimal variance would be obtained by either of the two, and therefore our theoretical analysis shows that combining the two methods allows improved variance. The improvement in the variance might be quite small, as these simulations show.

One limitation of these numerical results is the lack of control of the resolution. It is important to consider issues of noise/resolution trade-off. It seems reasonable that the linear combination of two methods will produce a resolution function which is an average of the individual resolutions, and thus not worse than the poorer of the two resolutions. In principle the variance is improved, but future work will need to more carefully understand and regulate the noise/resolution trade-off, especially for two methods with different resolutions.

Another limitation is the high computation time of the procedure to calculate the optimal coefficients λ_x^*. However, it is only the calculation of the weight functions that is time-consuming, and this can be done once for a fixed set of reconstruction parameters (sinogram size, regularization parameters). These fixed weight functions could be used to calculate the optimal λ^* coefficients for different scans (provided the variance in the sinogram can be reliably estimated).

There are several areas of future study. An important one is to try to find a way to balance the resolution of the two methods before optimizing the variance. Another is to understand the sawtooth pattern of lambda values in the ROI. Another study would be to investigate the robustness of the optimal lambda values for different objects (with different sinogram noise properties) for fixed reconstruction parameters. Another study would be to investigate the sensitivity of the variance of $f^{\lambda^*}(x)$ to the coefficients λ_x^* for fixed x. We anticipate the variance of $f^{\lambda^*}(x)$ to be fairly insensitive to λ_x^*.

TABLE III
THEORETICAL AND EXPERIMENTAL VARIANCES : $10^6 v^{\lambda^*}$

6.272 5.772 6.243 5.638	6.275 5.718 6.284 5.619
6.241 5.777 6.166 5.816	6.216 5.752 6.182 5.853
5.921 5.519 5.817 5.565	5.868 5.529 5.862 5.631
5.913 5.802 5.777 6.168	5.910 5.771 5.734 6.193
5.923 6.315 5.857 6.298	5.975 6.380 5.881 6.363

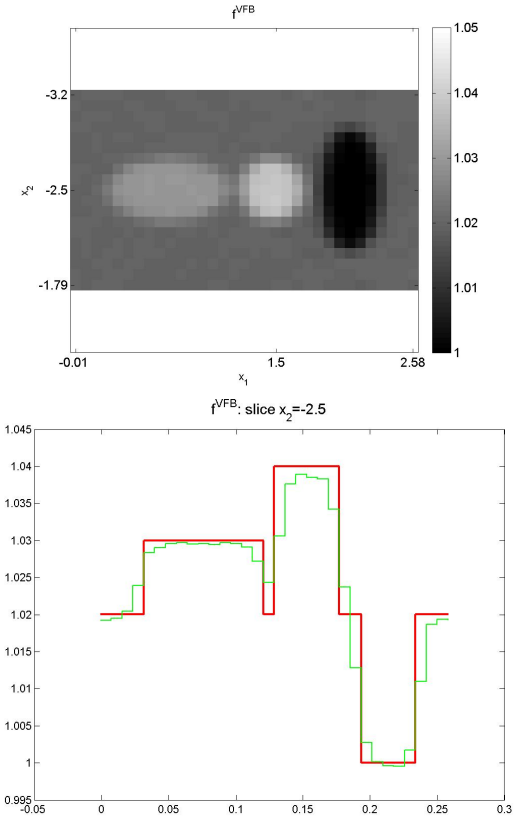


Fig. 3. Top line: VFB reconstruction of the ROI. Bottom line: the horizonal profile. Red thick line: the true function, green thin line: the reconstruction.

REFERENCES

[1] F. Noo, M. Defrise, H. Kudo, R. Clackdoyle. Image reconstruction from fan-beam projections on less than a short-scan. *IEEE Phys. Med. Biol.*, 47:2525–2546, 2002.

[2] R. Clackdoyle, F. Noo, J. Guo, J.A. Roberts. Quantitative reconstruction from truncated projections in classical tomography. *IEEE Trans. Nucl. Sci.*, 51:2570–2578, 2004.

[3] F. Noo, R. Clackdoyle, J.D. Pack. A two-step Hilbert transform method for 2D image reconstruction. *IEEE Phys. Med. Biol.*, 49:3903–3923, 2004.

[4] R. Clackdoyle, D.N. Ghosh Roy, M. Defrise, C. Mennessier, M.S. Ould Mohamed. Two-dimensional region-of-interest reconstruction: analysing the difference between Virtual Fan Beam and DBP-Hilbert reconstruction. *Conference Record of the 2009 IEEE Nuclear Science Symposium and Medical Imaging Conference. Orlando, FL.*, 2009.

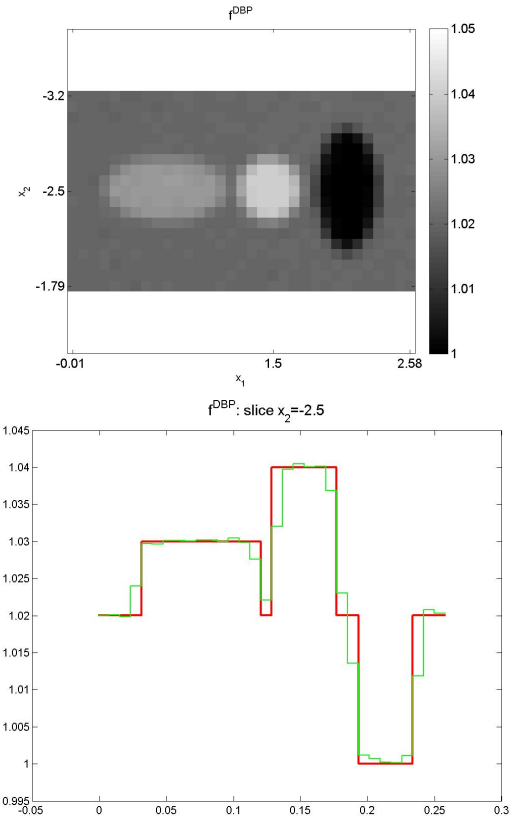


Fig. 4. DBP reconstruction of the ROI, and the central horizontal profile.

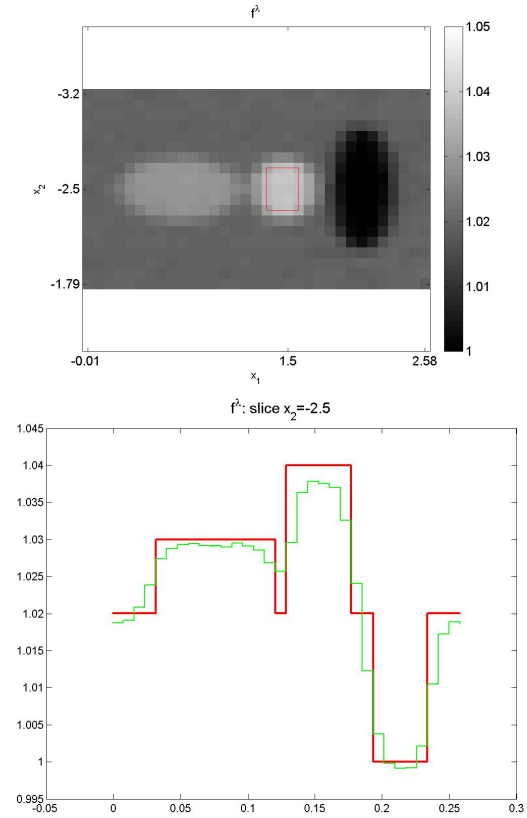


Fig. 5. Combined reconstruction $f^{\lambda *}$ of the ROI, with horizontal profile.

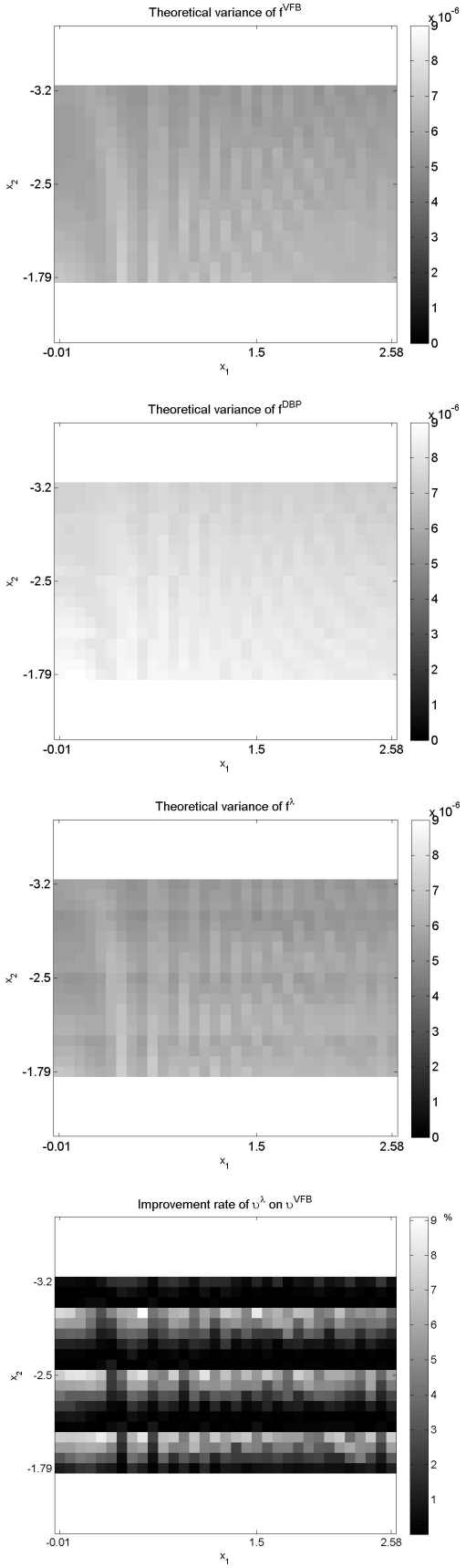


Fig. 6. The VFB (line 1), and DBP (line 2) variances computed using equation (3). The combined reconstruction variance (line 3) computed using equation (7) and the percentage improvement of υ^{λ} compared to υ^{VFB} (line 4).

X-ray source trajectories and their R-line coverage for long-object CB imaging with a C-arm system

Zhicong Yu, Frédéric Noo, Frank Dennerlein, Günter Lauritsch, Joachim Hornegger

***Abstract*—The geometry of the helix has been successfully applied in diagnostic CT for extended volume imaging without cone-beam artifacts. However, it cannot be used for C-arm systems due to the absence of slip-ring technology. For this reason, the reverse helix was proposed recently for C-arm systems, but efficient reconstruction from axially-truncated data collected on such a helix appears to be challenging. The main difficulty comes from the the missing R-line coverage in the central region of the scanned object. More specifically, the reverse helix is such that the theories that have been found for efficient handling of axial truncation cannot be applied, because large portions of the object are not intersected by R-lines**

In this work, we revisit the option of performing extended field-of-view imaging using a sequence of circular short-scans connected by line segments. We find that the R-line coverage is insufficient for a central region of interest when a line-segment is tightly fit between parallel circular arcs. On the other hand, extension of the line beyond the circular arc helps to increase R-line coverage in the central region of interest. Therefore, we propose a trajectory composed of two parallel circular arcs connected by an extended line. This trajectory does have a nice R-line coverage inside the ROI, but it has a discontinuity at the endpoints of the line. To overcome this problem, we suggest replacing the two parallel circular arcs by two helices, which can be duplicated along the axial axis conveniently and which, moreover, keeps the trajectory continuous and thus is more practical.

I. Introduction

Over the decade, cone-beam (CB) computed tomography has become a valuable tool in interventional radiology. Its success stems from its ability to provide the medical doctor with immediate feedback during a clinical procedure, thereby allowing on-the-fly adjustments. So far, circular data acquisition has been used, but more sophisticated geometries are being considered due to the problem of cone-beam artifacts and also due to limited volume coverage.

An attractive geometry for extended volume imaging with no cone-beam artifacts is the helix. This geometry has been very successful in diagnostic CT, but is unfortunately not practical for interventional radiology. The problem is the need for slip-ring technology, which is difficult to use (if possible at reasonable cost) with C-arm systems. To perform extended volume imaging with a C-arm system, another geometry must be found. Many options are possible, from using a combination of circles and lines, to using a reverse helix, as suggested by the group of X. Pan at the University of Chicago [1]. The reverse helix has many merits, but efficient reconstruction from axially-truncated data collected on such a helix appears to be challenging [1], [2]. The main difficult comes from the R-lines not covering the whole scanned object (an R-line is any line segment that connects two source positions together). More specifically, the reverse helix is such that the theories that have been found for efficient handling of axial truncation [3]–[6] cannot be applied, because large portions of the object are not intersected by R-lines.

In this work, we revisit the option of performing extended field-of-view imaging using a sequence of circular short-scans connected by line segments. In particular, we investigate R-line coverage with the goal of finding source-trajectory parameters such that a central region-of-interest within the object is fully covered by R-lines.

The paper is organized as follows. First, we describe the data acquisition geometry of interest. Next, we discuss R-line coverage resulting from two parallel circular arcs, and also R-line coverage resulting from connecting a line orthogonally to the endpoint of a circular arc. From there, we are then able to present the R-line coverage for the whole data acquisition geometry. Our results show that some parameters allow full coverage of a region-of-interest, but these parameters unfortunately come with practical implementation concerns. We discuss these concerns and potential remedies in the last section.

II. Data acquisition geometry

We consider extended volume imaging using periodic duplicates of a source trajectory consisting of two circular arcs connected by a segment of line. The patient is assumed to lie along the z-axis, the arcs are in parallel planes that are orthogonal to this axis, and the line is orthogonal to each arc through one of its endpoints. Figure 1 depicts this trajectory. Two options are considered: (a) the line is spatially limited by the arcs, (b) the line extends beyond the arcs. Figure 1 also shows how each circular arc is oriented relative to x and y-axes that form together with the z-axis a Cartesian system of coordinates. The distance in z between the arcs is $2\,H$, the radius of the arcs is R, and the line extension in the second path option is Δh on each side. Also, the plane $z = 0$ is chosen

Z. Yu and F. Noo are with the Department of Radiology, University of Utah, 729 Arapeen Drive, Salt Lake City, UT 84102 USA (e-mail: zyu@ucair.med.utah.edu, noo@ucair.med.utah.edu).

F. Dennerlein and G. Lauritsch are with Siemens AG, Healthcare Sector, 91301 Forchheim, Germany.

J. Hornegger is with the Chair of Computer Science 5 (Pattern Recognition), University of Erlangen-Nuremberg, 91058 Erlangen, Germany.

This work was partially supported by a grant of Siemens AG, Healthcare Sector and by the U.S. National Institutes of Health (NIH) under grant R21 EB009168. The concepts presented in this paper are based on research and are not commercially available. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

to be at mid-distance between the two arcs. To avoid confusion with other arcs that will appear later in the discussion, from now on we will use the term S-arcs to refer to the arcs forming the source trajectories.

III. Elemental R-line Coverage

In this section, we first discuss the R-line coverage resulting from connecting points from one arc to the other arc. Next, we discuss the additional coverage resulting from connecting points on the line to points on the arcs. Note that our data acquisition is symmetric relative to the $z = 0$ plane. Therefore, the R-line coverage at position $z = z_0$ is the same as the coverage at $z = -z_0$ for any $0 \leq z_0 \leq H$. Hence, we only discuss R-line coverage at positions $z \geq 0$.

A. Arc-to-arc coverage

To understand the R-line coverage in this case, we start by considering the simpler case where each S-arc has a length of 360 degrees. Figure 2 shows how the R-line coverage can be found in this case. Location $z = 0.2\,H$ is used for the illustration but a similar result would be obtained at any other z-location. Basically, we take a point on the upper S-arc, called A^i_+, and connect it to all points on the lower S-arc. Doing so, we create the surface of a cone that intersects the plane $z = 0.2\,H$ along a circle. This circle defines the R-line coverage coming from A^i_+ in the plane $z = 0.2\,H$. By moving A^i_+ along the upper S-arc, we obtain additional circles, as shown on the right side of Figure 2. The union of these circles is the full R-line coverage in the plane $z = 0.2\,H$; this union is an annular region with external boundary, $\mathcal{C}$, corresponding to the intersection between plane $z = 0.2\,H$ and the cylinder on which the source trajectory is drawn.Now, we consider the case where the S-arcs are shorter. A length of 230 degrees is used for the illustrations; similar results would be observed with other short-scan lengths. Figure 3 shows the coverage in plane $z = 0.2\,H$, whereas figure 4 illustrates the coverage at $z = 0$.

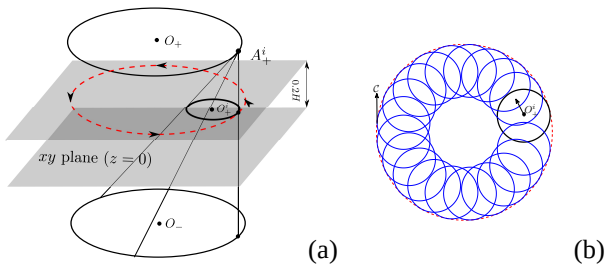


Fig. 2. Arc-to-arc R-line coverage when the length of the arcs is 360 degrees. (a) Coverage in plane $z = 0.2\,H$ due to one point from the upper S-arc; this coverage is a circle centered on a point denoted as O^i_+ in the figure. (b) Coverage in the same plane as on the left, but due to several points on the upper S-arc.

First, comparing figure 2a with figure 3a, we see that the coverage from A^i_+ is not a full circle anymore, it is a circular arc, which we refer to as an R-arc to avoid ambiguity between such an arc and the S-arcs that form the source trajectory. By moving A^i_+ along the upper S-arc, we obtain a union of R-arcs that defines the R-line coverage at $z = 0.2H$. Identifying the region covered by this union is not too complex. First, we observe that the endpoints of the R-arc, A_h and B_h, move along two arcs. The arc along which A_h moves is shown in figure 3c; it is the intersection between the plane $z = 0.2\,H$ and the (open) cone surface that results from connecting A^i_- to the upper S-arc, where A^i_- is the point on the lower arc that is at the same (x, y) location as A^i_+. The arc along which B_h moves is shown in figure 3d; this arc is created from the cone based on the upper S-arc and the endpoint A^e_- on the lower S-arc. Secondly, compared to figure 3, we illustrate the variation of the R-line coverage along the z-axis by looking at the plane $z = 0$, as shown in figure 4. Using the same process as in figure 3, we take a point A^i_+ on the upper S-arc and connect it to the lower S-arc; with this procedure, we get a surface consisting of a partial cone, which intersects with the plane $z = 0$ along a partial circular arc whose two endpoints are denoted as A_h and B_h. By moving A^i_+ along the upper S-arc, we acquire the combination of R-arcs that forms the R-line coverage in the plane $z = 0$, as illustrated in figure 4b. The identification of the R-line coverage in the plane $z = 0$ could be performed similarly to the case depicted in figure 3. The difference is that the orbit of the endpoint A_h is the R-arc itself corresponding to the point A^i_+ on the upper S-arc, as shown in figure 4a; while the point B_h follows the dashed partial circular arc as illustrated in figure 4c. Both tracks of points A_h and B_h go through the center of the cylinder, on whose surface the two S-arcs are located. Figure 4 also contains the results of a numerical simulation. Combining the information in figure 3, we observe that in planes orthogonal to the z-axis, as $|z|$ increases, the R-line coverage corresponding to one point on the upper S-arc becomes smaller and the orbits of the endpoints of the R-arc become larger, and thus the R-line coverage around z axis becomes worse.

B. Arc-to-line coverage

In this subsection, we investigate another elemental data acquisition geometry: a line orthogonally attached to a partial circular arc at one of its endpoints. Recall from figures 1a and 1b that two S-arcs are symmetric relative to the plane $z = 0$. In this case, the R-line coverage resulting from connecting the points on the line to points on the upper S-arc is symmetric to that obtained from connecting the points on the line to the points on the lower S-arc. Therefore, we consider here only the R-line coverage that results from the line located above the S-arc. Figure 5 depicts this R-line coverage. Here, we used location $z = 0$ and angular coverage $230°$ as an example, but a similar result could be obtain at any location $-H \leq z \leq H$. Figure 5b shows the R-line coverage resulting from the numerical simulation.

As observed, the R-line coverage of the arc-to-line in the plane $z = 0$ is a partial disk. Compared to Figure 4, we can see the outline of the partial disk is actually the union of the R-lines connecting the point A^i_+ to the lower S-arc in Figure 4. The union of all the disks along the line is a cone whose base is bounded by the lower S-arc and with peak at the

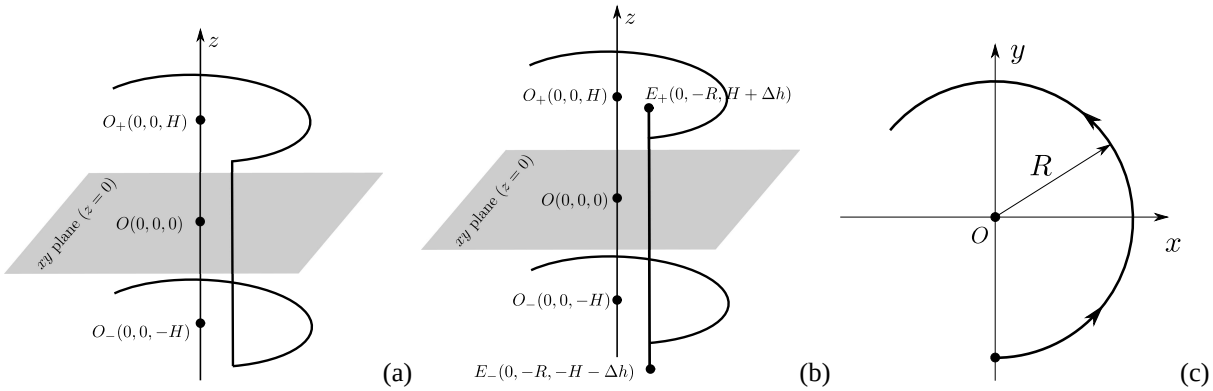


Fig. 1. Data acquisition geometry. Extended volume imaging is performed using duplicates of a path consisting of two arcs plus a line. Two options are considered for this path: (a) the line is tightly fit between the arcs, so that each endpoint of the line corresponds to one endpoint of an arc; (b) the line extends beyond the arcs by a distance Δh on each side. (c) Orthogonal projection of the source trajectory onto the $x-y$ plane.

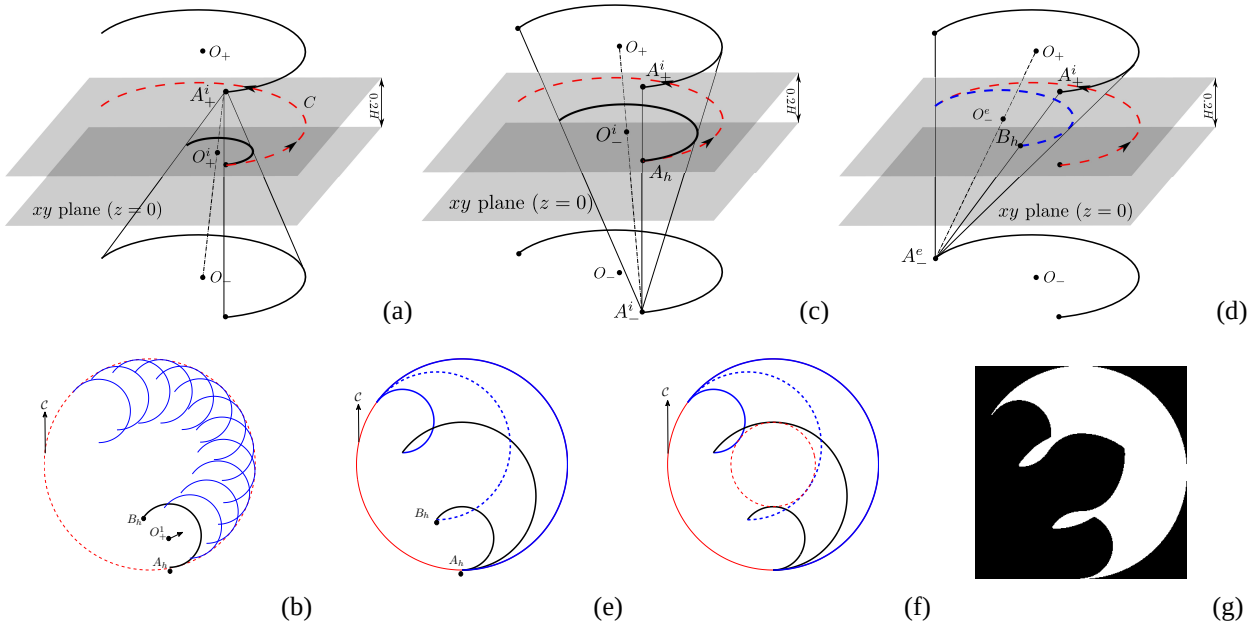


Fig. 3. Arc-to-arc R-line coverage when the length of the arcs is 230 degrees. (a) Coverage in plane $z=0.2H$ due to one point from the upper S-arc; this coverage is a partial circular arc centered on a point denoted as O^i_+ in the figure. (b) Coverage in the same plane as in (a), but due to several points on the upper S-arc. (c) The track (the solid arc centered on the point O^i_-) which A_h follows when moving point A^i_+ along the upper S-arc. It is the intersection between the cone surface due to A^i_- and the plane $z=0.2H$. (c) The track (the dashed arc centered on the point O^e_-) of B_h when moving the point A^i_+ along the upper S-arc. (e) Combination of arcs from (b), (c) and (d). (f) Additional outlines are needed for R-line coverage of the two S-arcs. The reason is that when the R-arc moves along the C, due to the large angular coverage of the S-arcs, A_h and B_h are not always the furthest point to the z axis. (g) the numerical simulation result.

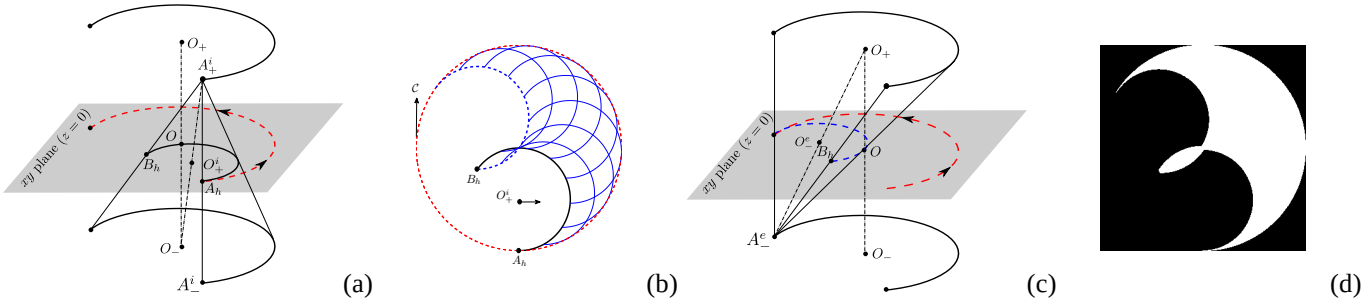


Fig. 4. Arc-to-arc R-line coverage when the length of the arcs is 230 degrees. (a) Coverage in plane $z=0$ due to one point from the upper S-arc; this coverage is a partial circular arc centered on a point O^i_+. (b) Coverage in the same plane as in (a), but due to several points on the upper S-arc. (c) The track (the dashed partial circular arc) of B_h when A^i_+ moves along the upper S-arc. (d) the numerical simulation result.

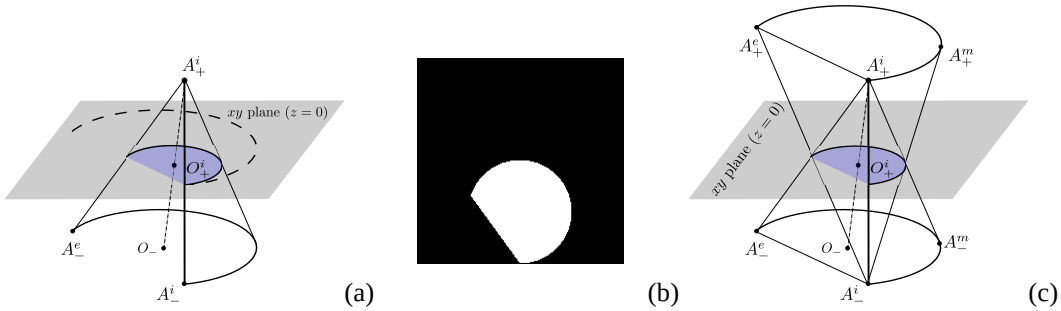


Fig. 5. Arc-to-line R-line coverage when the length of the arcs is 230 degrees. The line is attached to one of the endpoints of the S-arc. (a) Coverage in plane $z = 0$ due to one point A^i_+. This coverage is a partial disk centered on a point denoted as O^i_+. (b) the numerical simulation result. (c) the union of two symmetric cones of R-line coverage, which is resulting from connecting points on the line to the points on the two S-arcs.

point A^i_+. Due to the symmetry of the trajectory as illustrated in Figure 1b, the R-line coverage between the line and the upper S-arc would be an upside-down cone, congruent to that between the line and the lower S-arc. The union of these two cones intersected with the $z = 0$ plane is shown in Figure 5c. As we can see, the R-line coverage resulting from connecting points on the line to the two R-arcs has its minimum in the plane $z = 0$ and increases when $|z|$ becomes larger.

To give a better understanding of the relation between the R-line coverage of the two R-arcs and that of a line and an S-arc, we offer an illustration in Figure 6. Here we choose the plane at height $z = 0.2H$ as an example, but a similar result could be obtained for any other locations with $-H \leq z \leq H$. Figure 6 depicts the combination of R-lines from the elemental trajectories: arc-to-arc and arcs-to-line . We observe that the R-line coverage from the line and the S-arc and the R-line coverage from the two S-arcs compensate each other quite well.

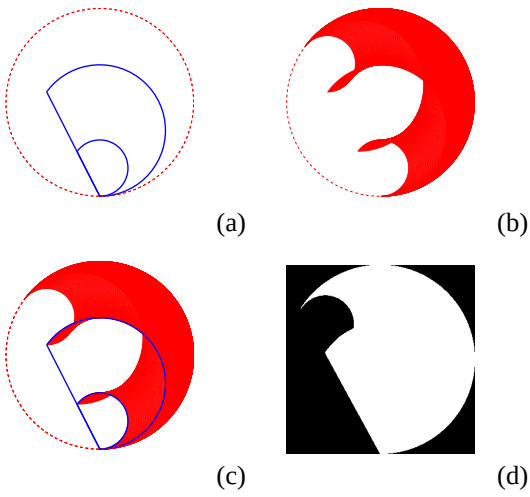


Fig. 6. Combination of R-line coverage from arc-to-arc and arc-to-line. (a) the outline of the R-line coverage from arc-to-line in the plane $z = 0.2H$. The small partial circular arc is due to the top endpoint of the line and the big one due to the bottom endpoint of the line. (b) Coverage between the two S-arcs. (c) Combination of R-line coverage of (a) and (b). (d) the numerical simulation result.

IV. Tight Line Plus Arcs

Using the R-line coverage of the elemental trajectories, we obtain the R-line coverage for the trajectory composed of two parallel S-arcs and a tight line, as illustrated in Figure 1. Recall from section III that the R-line coverage of the arc-to-arc and the arc-to-line compensate each other quite well. However, the R-line coverage of the arc-to-line has a minimum in the plane $z = 0$ while the R-arc obtained by connecting one point on the upper S-arc to the lower S-arc reaches a maximum length; the R-line coverage of the arc-line-arc is worst in the plane $z = 0$. Figure 7 illustrates the R-line coverage of the arc-line-arc in the plane $z = 0$ with source-angular coverage of 230° and 310°. As we can see, the R-lines are not fully covering the Region-of-Interest (ROI) for both angular coverages. In fact, as long as the angular coverage is less than 360°, there is always an angular space touching the z axis that is not covered by R-lines.

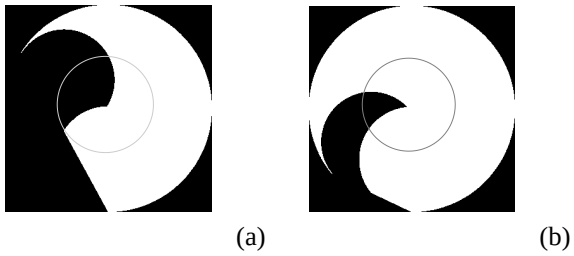
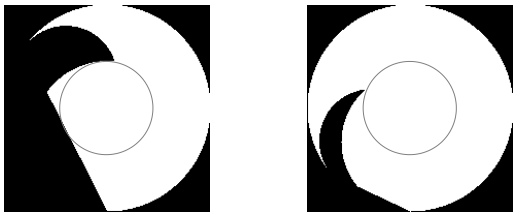


Fig. 7. Numerical simulation results of R-line coverage for two parallel circular arcs plus a tight line (ALA). (a) R-line coverage of ALA in the plane $z = 0$ with angular coverage of 230°. (b) R-line coverage of ALA in the plane $z = 0$ with angular coverage of 310°.

V. Extended Line Plus Arcs

To overcome the problem mentioned in the previous section, we here propose the trajectory composed of two parallel S-arcs and an extended line as shown in Figure 1b. In this case, the R-line coverage from an arc-to-line trajectory occupies a larger area that can offer more R-lines in the central region. How large the R-line coverage from an arc-to-line trajectory is depends on how long the line extends beyond the two S-arcs. Details will be given later in this section.

First, we would like to show the numerical simulation result of the R-line coverage for an arc-extended-line-arc trajectory with angular coverage of 230° and 310° in Figure 8. Here for the angular coverage of 230° we use $\Delta h = 0.832(2H)$ and for the angular coverage of 310° we use $\Delta h = 0.485(2H)$. The circle in the middle is the ROI corresponding to the angular coverage of 230°. Compared to the simulation results depicted in Figure 7, we can see the ROI is already fully covered.



(a) (b)

Fig. 8. Numerical simulation result of two parallel circular arcs plus an extended line (AELA). (a) R-line coverage of AELA in the plane $z = 0$ with angular coverage of 230° and $\Delta h = 0.832(2H)$. (b) R-line coverage of AELA in the plane $z = 0$ with angular coverage of 310° and $\Delta h = 0.485(2H)$.

We still have one open question: how far should we extend the line, i.e., how large Δh should be? From the analysis in seciton III, we can see that the R-line coverage of the trajectory of two parallel circular arcs plus a tight line has its minimum in the plane $z = 0$. Therefore, we only investigate here the R-line coverage for the arc- (extended)line-arc trajectory in the plane $z = 0$.

Suppose the desired ROI radius is given and denoted as $RROI$. To be practical, we calculate the fan angle using $RROI$ and denote it as α. We first start from the short scan and calculate the minimum Δh that is needed to cover the ROI; then we increase the angular coverage of the trajectory and see whether the value of Δh depends on the angular coverage or not. Figure 9a illustrates the R-line coverage of the arc-line that needed to cover the ROI. In this case, the R-line coverage from an arc-to-line trajectory has to cover the whole ROI, and thus we can calculate the minimum line extension using the following equation:

$$\frac{\Delta h}{2H} = \frac{RROI}{R - RROI} \tag{1}$$

If Δh is bigger than the value obtained from equation 1, we get additional gain of R-line coverage in the central region, as illustrated in Figure 9b.

With the increase of the angular coverage, the R-line coverage from the arc-to-arc becomes larger in the plane $z = 0$. Therefore, at some critical angular coverage, the requirement of Δh is reduced. Comparing Figure 9a and 9c, point K is the intersection between the ROI and the circle centered on the point O^i. In Figure 9a, point K is on the boundaries of the ROI and the missing R-line coverage region, but it is not the furthest such point from A^e. However, when the angular coverage increases, for example, in Figure 9b, point K is the furthest point from A^e such that it is still on the boundaries of the ROI and the missing R-line coverage region. Connecting point K and O and extending it along the direction from K to O, we get an intersection A^c, as shown in Figure 9c; we denote the corresponding angular coverage as θ_c. The requirement for Δh changes when the angular coverage pass through θ_c. Suppose now the angular coverage is greater than θ_c, as illustrated in Figure 9c, and draw a dashed circle through points K and A^e such that it is tangent to the big circle where A^i and A^e are located. Draw a line from point A^e to O and intersect this line with the dashed circle at point W. Letting the distance between O and W be R_x, we can obtain the minimum requirement of Δh using the equation below:

$$\frac{\Delta h}{2H} = \frac{r_x}{R - r_x}. \tag{2}$$

Similar to the situation in Figure 9b, if we adopt Δh larger than the requirement shown in equation 2, we get an additional gain of R-line coverage in the central region in the plane $z = 0$, as depicted in Figure 9d.

Let θ be the angular coverage, combining equations 2 and 1, we get a complete formula expressing the minimum requirement for Δh:

$$\frac{\Delta h}{2H} = \begin{cases} \frac{RROI}{R-RROI} & \pi + 2\alpha \le \theta \le \theta_c \\ \frac{r_x}{R-r_x} & \theta_c < \theta \le 2\pi \end{cases} \tag{3}$$

Figure 10 depicts the minimum requirement for Δh for different source-angular coverage according to equation 3. In the figure, three different $RROI$ are used, i.e., 0.454R, 0.300R and 0.156R. Figure 10 shows that when the angular coverage

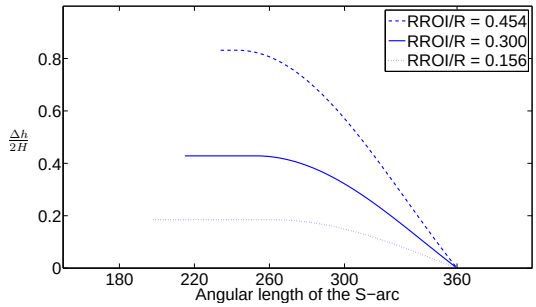


Fig. 10. Minimum Δh requirement for different angular coverage. Three $RROI$ are used, i.e., $0.454R$, $0.300R$ and $0.156R$, where R is the radius of the S-arc.

approaches 360°, the minimum requirement of Δh reduces to zero. This is consistent with the conclusion in section III that the two parallel circles plus an orthogonally attached tight line has full R-line coverage. Although there is always a Δh available to make sure that the ROI is fully covered by R-lines, it is too expensive to do so when the angular coverage is too small and the required ROI is too large, which could lead to an impractical trajectory. However, for a trajectory with large angular coverage and an ROI requirement that is not too great, the two parallel circular arcs plus an extended line does offer nice R-line coverage in the central region, while keeping the cost of line extension low.

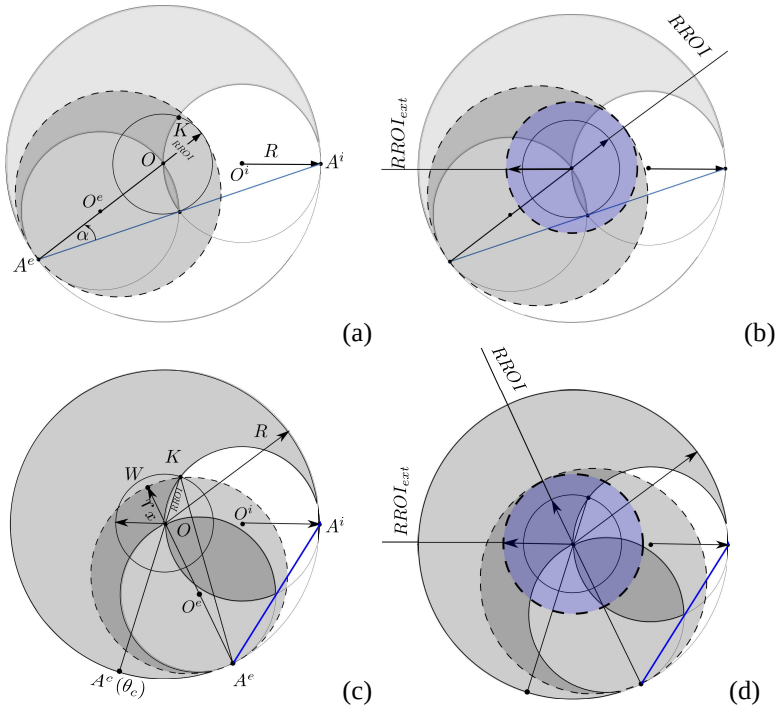


Fig. 9. Relation between Δh and $RROI$ (the desired radius of ROI), such that the R-line coverage of the arc-line covers the whole ROI in the plane $z = 0$. The minimum short scan here is design according to RROI. (a) The R-line coverage from the arc-line should at least cover the whole ROI for the minimum short scan, and we denote the corresponding extension of line as $\Delta h_R OI$. (b) the increase of Δh leads to the increase of central R-line coverage as described by the central dashed circle. (c) The R-line coverage from the arc-line is not necessary to cover the whole ROI thanks to the additional R-line coverage from the arc-to-arc, when the angular coverage is bigger than θ_c that corresponds to A_c in the figure. (d) As in Figure (b), the increase of Δh will give us some gain of the R-line coverage in the central region.

VI. Conclusion

We have investigated R-line coverage for a trajectory consisting of two circles connected by a line. We have observed that the R-line coverage is insufficient for a central region of interest when the line is tightly fit between the circles. On the other hand, if the line extends beyond the circle, we find the nice result that there exists a central ROI that is fully covered by R-lines. How big this central region is depends on how big Δh is. Also, we have shown that some further gain is achieved when the length of the short-scan is increased beyond its minimal value. From a practical implementation viewpoint, the line extension is not very satisfactory because it requires a short pause in exposure. One way to circumvent this problem is to replace the two S-arcs by helical arcs that touch the extended line at its endpoints as shown in Figure 11. We have done preliminary numerical simulations with this trajectory and have observed that it retains essentially the same properties of the trajectory studied in this work, while being more practical.

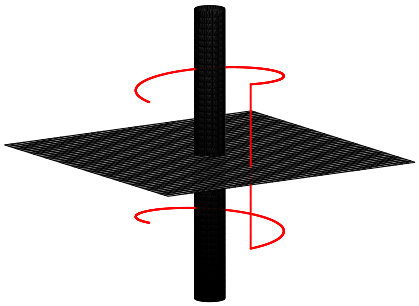


Fig. 11. Trajectory composed of two helices and a line.

References

[1] S. Cho, D. Xia, C. A. Pelizzari, and X. Pan, "Exact reconstruction of volumetric images in reverse helical cone-beam ct," *Med.Phys*, vol. 35, no. 7, pp. 3030–040, July 2008.

[2] F. Noo, A. Wunderlich, G. Lauritsch, and H. Kudo, "On the problem of axial data truncation in the reverse helix geometry," in *10th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, September 2009, pp. 90–93.

[3] A. I. Katsevich, "An improved exact filtered backprojection algorithm for spiral computed tomography," *Adv. Appl. Math*, vol. 32, no. 4, pp. 681–697, May 2004.

[4] J. D. Pack, F. Noo, and R. Clackdoyle, "Cone-beam reconstruction using the backprojection of locally filtered projections," *IEEE Trans. Med. Imaging*, vol. 24, no. 1, pp. 70–85, Jan 2005.

[5] Y. Zou and X. Pan, "Exact image reconstruction on pi-lines from minimum data in helical cone-beam ct," *Phys. Med. Biol.*, vol. 49, no. 6, pp. 941–959, Feb 2004.

[6] Y. Ye, S. Zhao, H. Yu, and G. Wang, "A general exact reconstruction for cone-beam ct via backprojection-filtration," *IEEE Trans. Med. Imaging*, vol. 24, no. 9, pp. 1190–1198, Sep 2005.

Noise Properties of Grating Based X-Ray Phase Contrast Computed Tomography

Thomas Köhler[1], Klaus Jürgen Engel[1], and Ewald Roessl[1]

Abstract— **We investigated the noise properties of a grating based phase contrast imaging setup in simulations with respect to its spectrum and the energy dependence of the achievable contrast to noise ratio (CNR). The noise in reconstructed phase contrast images shows a maximum at low frequencies. The energy dependence of the CNR was evaluated by simulating measurements at constant dose for different energies. For a 11 cm size object, we found that the CNR reaches a maximum around 50 keV.**

I. Introduction

In conventional x-ray imaging, the contrast is generated by the attenuation of the x-ray beam by the photoelectric effect and the Compton effect. An alternative way to generate contrast is to measure the phase shift that an object imposes on the "wave front" of an x-ray beam.

Although a number of phase contrast imaging (PCI) techniques exist [1], PCI plays only a minor role in the domain of medical imaging, mainly because of the stringent requirements of known techniques on coherence and monochromaticity of the x-ray beams. These are met by synchrotron x-ray sources but unmet by conventional x-ray tubes as almost exclusively used in the medical domain. Recently, a technique well known in the visible range of wavelengths has been adopted to the x-ray domain, the so-called Talbot-Lau interferometry [2–4]. This technique permits the use of conventional tubes and functions with partially coherent beams generated by gold gratings placed close to the x-ray tube's exit window. Among all known techniques of phase-contrast imaging in the hard (medical) x-ray domain, Talbot-Lau interferometry requires the smallest coherence volume for phase-sensitive imaging [3].[1]

The grating based setup allows tomographic reconstruction of the three-dimensional distribution of the refractive index [6–8]. The resulting images of biological samples like a rat brain [6] or a rat heart [9] show excellent soft tissue contrast that would likely add clinical value in diagnostic imaging if it is applicable to human imaging.

Currently, experimental setups for grating based PCI use x-ray energies in the range from 20 to 30 keV. For diagnostic imaging, this energy range is useful mainly for mammography. However, for other applications in radiology, higher x-ray energies are required. Concerning the need for higher energies, several comments are in order: Firstly, gratings with higher aspect ratios are required, which are more demanding to manufacture. Secondly, the phase contrast mechanism does not require absorption of photons. Therefore, it is commonly argued that phase contrast imaging potentially benefits from higher energies since fewer photons are absorbed – at least if the photoelectric effect dominates the total attenuation – and vice versa, more photons can be used for phase contrast imaging [1].

Noise in phase contrast x-ray imaging can be very different to that of absorption contrast imaging as shown, e. g., in [10] for the case of propagation based PCI. In fact the noise shows a broad spatial autocovariance profile, or, in other words, there is a significant amount of low-frequency noise.

The goal of the present work is to investigate the noise features of tomographic grating based PCI, viz., the noise spectrum and the energy dependence of the contrast to noise ratio (CNR) that can be achieved.

II. Method

The setup for grating based PCI is show in Fig. 1, where only the most important optical elements required for phase contrast imaging are shown [2, 4]. A plane wave of coherent x-rays hits the sample. Due to refraction, the phase front behind the object is distorted. The distorted wavefront passes a beam-splitter grating G_1, which creates a characteristic interference pattern at the location of the analyzer grating G_2. The interference pattern is imaged by measuring the x-ray intensity using a detector D immediately behind G_2 at several relative transverse positions of the gratings G_1 and G_2.

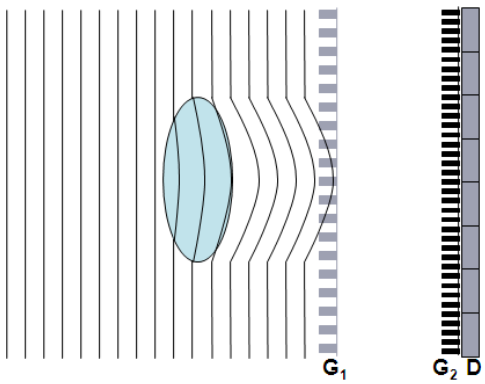


Fig. 1. Grating based phase contrast CT setup.

A. Simulation

Fig. 2 shows the mathematical phantom used in the study. The material compositions were taken from [11], and the optical properties of these compositions were calculated

[1] Philips Technologie GmbH Forschungslaboratorien

[1]We do not consider here the alternative grating based PCI setup proposed in [5].

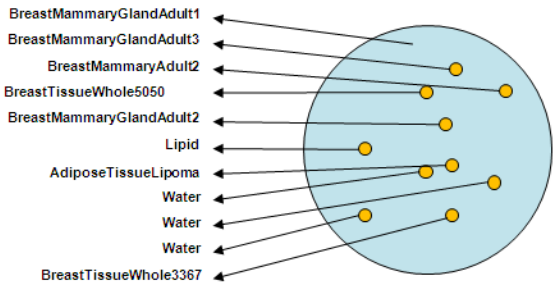


Fig. 2. Phantom used for simulations. The diameter of the object is 110 mm.

based on the formulas given in [12] using atomic form factors tabled in [13]. The projection approximation [14] was used to generate 721 monochromatic parallel beam projections over 180° with 1024 detector pixels covering 13 cm field of view. For each projection angle, eight different relative positions of G_1 with respect to G_2 were simulated in order to retrieve the absorption contrast and phase contrast images. The projections were simulated at four different energies from 25 to 85 keV assuming monochromatic x-rays in order evaluate the energy dependence of the CNR. The grating pitches were kept constant, whereas the distance of G_2 to G_1 was adjusted for each energy to the first fractional Talbot distance (i. e., the distance was 40, 76, 108, and 140 mm, respectively). Noise was added to the data sets such that for each energy the total dose deposited into the object was constant, thus allowing for a meaningful comparison of the CNR in the different reconstructed images. The dose calculation was performed using a Monte Carlo simulation considering photoelectric absorption and Compton scattering.

B. Reconstruction

Reconstruction was performed using filtered back-projection using a conventional ramp-filter for the absorption contrast image [15] and a Hilbert filter for the phase contrast image [16]. In order to suppress aliasing, four point interpolation was used during back-projection.

III. Results

A. Noise Power Spectrum

Since linear reconstruction methods are used here, the noise power spectrum can be determined from projections without an object. Projections with white Gaussian distributed noise were simulated and reconstructed. White noise is appropriate since the processing in grating based PCI to obtain the absorption projection and the differential phase contrast projection is performed on each detector pixel individually. Thus, no spatial correlation is introduced by the processing. The reconstructed images are shown in Fig. 3. The noise power spectrum of these images is determined by 2D FFT followed by a binning into several frequency bands of 3 lp/cm width and a calculation of the total signal intensity in each frequency bin. The resulting spectra shown in Fig. 4 are further normalized to the total noise power in the image.

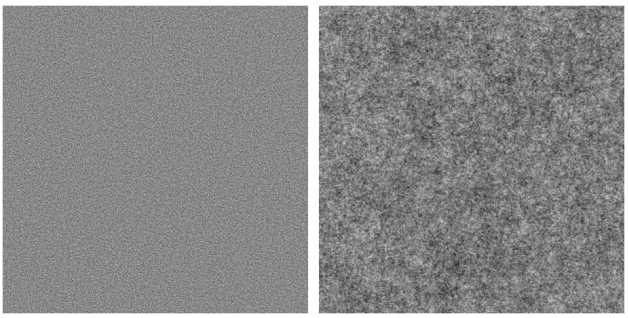


Fig. 3. Reconstruction of a noisy air scan. Left: Absorption contrast image, right: Phase contrast image.

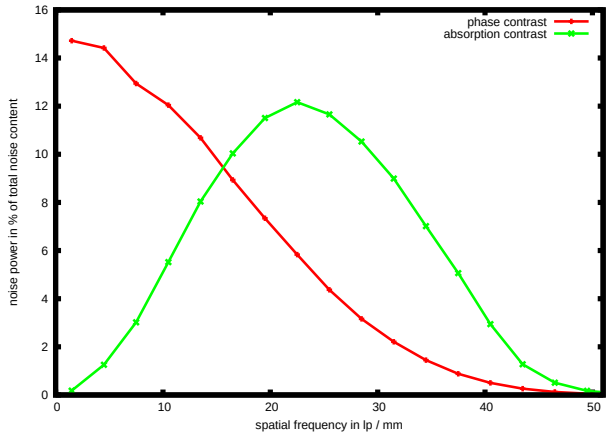


Fig. 4. Noise power spectrum of the reconstructed absorption and phase contrast images.

In order to illustrate one important aspect of the different noise power spectra, a region of interest (ROI) of the simulated phantom scan at 45 keV was reconstructed once at system resolution (≈ 0.17 mm) and at lower resolution (≈ 0.5 mm). The images are shown in Fig. 5. The noise level in the absorption contrast images decreases by the smoothing by a factor of 6.1, whereas the noise reduction for the phase contrast image is only a factor of 2.0.

B. Energy Dependence of CNR

Fig. 7 shows an overview of the reconstruction results for the absorption and phase contrast images. Note that there are low-frequency artifacts in the phase contrast images which are due to the very high gradient at the phase front created by the object boundary, which cannot be recovered properly. Fig. 6 shows the central part of the phantom containing four objects which were selected for a detailed

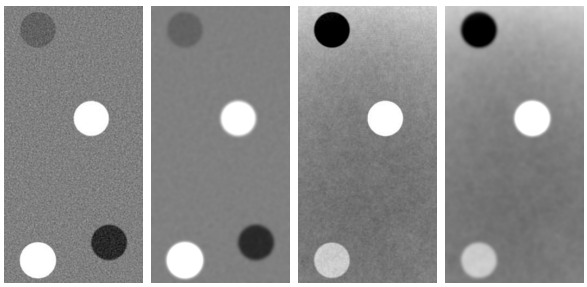


Fig. 5. ROI reconstruction at 45 keV at system resolution (first and third image) and at 0.5 mm resolution (second and fourth image) for absorption contrast (left) and phase contrast (right).

analysis of the CNR as a function of x-ray energy. This analysis is shown in Fig. 8.

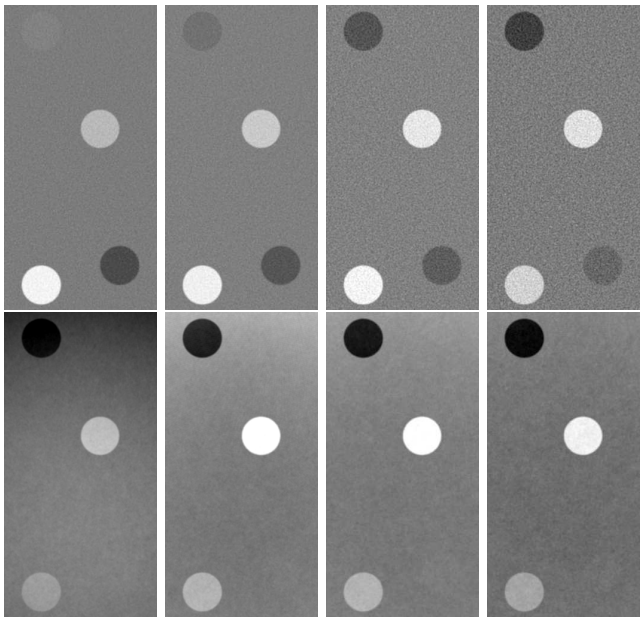


Fig. 6. Central region of interest of the images shown in Fig. 7. A wider window than in Fig. 5 was selected to show the full dynamic range of the region of interest. The display level was set to the value of the background material.

IV. Discussion

The noise power spectrum of the phase contrast image shows a maximum at low frequencies, see Fig. 4. This can be easily understood from the reconstruction kernel used in the reconstruction: The ramp filter used in absorption contrast imaging damps low frequencies and boosts the higher frequencies leading to the well-known peak in the noise power spectrum near the Nyquist frequency of the detection system (actually, due to the smoothing effect of interpolation during back-projection, the peak is well below the Nyquist frequency of 39 lp/cm). On the other hand, the Hilbert filter used in phase contrast imaging treats all frequencies equally, leading to a higher noise power at lower frequencies.[2] One of the important implications of this feature for tomographic medical imaging is that the superior CNR of phase contrast images compared with absorption contrast as demonstrated in [6, 9] will diminish for applications, where a lower spatial resolution is required: In absorption contrast imaging, noise and resolution can be traded off very efficiently since smoothing also suppresses the peak in the noise power spectrum. However, for phase contrast imaging, smoothing only suppresses the tail of the noise power spectrum and does not lead to a substantial increase in CNR. For the example shown in Fig. 5, the CNR of the absorption contrast image increases three times more than in the phase contrast image. This effect should always be considered if the two modalities are compared with respect to the achieved CNR. Regarding CNR, it should also be noted that the noise in the phase contrast image can be decreased by increasing the distance between G_1 and G_2 to a higher fractional Talbot order because the sensitivity to the gradient of the phase front is increased [2]. The price to pay for this increase in CNR is that the phase wrapping (that leads to the cupping and capping artifacts in Fig. 7) will also appear at even smaller phase gradients and the problem of unwrapping the phase gradient will become a demanding task.[3]

The CNR of phase contrast images measured at constant dose for the sample phantom are qualitatively very similar: For all objects, the magnitude of the CNR shows a maximum around 50 keV. The existence of a maximum at a distinct energy independent of the object is caused by several effects: First, the phase shift introduced by the objects is exclusively created by the electron density with an energy dependence of E^{-1}. Thus, it is expected that the energy dependence of the CNR of all objects in the phantom is qualitatively the same. The second cause is the different contributions of the photoelectric effect and Compton scatter to the total attenuation. For lower energies, the photoelectric effect dominates. However, it decreases approximately as E^{-3}, whereas the phase contrast decreases only as E^{-1}. Thus, going to higher energies decreases the contrast in the phase image, but this is more than compensated by the fact that at constant dose more photons can be used for imaging since fewer of them are absorbed. However, as soon as the Compton effect contributes significantly to the attenuation at higher energies, the total transmission decreases again and the decreasing phase contrast is – in the constant dose scenario – no longer compensated by an even faster decreasing absorption.

Herzen et al. [17] showed in experiments using several different liquids that phase and absorption contrast images show complementary information in the sense that the CNR is sometimes larger in the phase contrast image and sometimes in the absorption contrast image. Our simulation using material compositions of biological tissues confirms that this feature is relevant for medical imaging.

[2]Due to the inhomogeneous sampling of the Radon space and thus also the Fourier space with a series of parallel projections, this equal treatment of all frequencies does not lead to flat noise power spectrum.

References

[1] R. A. Lewis. *Phys Med Biol*, 49(16) (2004) 3573.
[2] T. Weitkamp et al. *Opt Express*, 13(16) (2005) 6296.
[3] F. Pfeiffer et al. *Phys. Rev. Lett.*, 98(10) (2007) 108105.
[4] A. Momose et al. *Opt Express*, 17(15) (2009) 12540.
[5] Z.-F. Huang et al. *Physical Review A*, 79 (2009) 013815.
[6] F. Pfeiffer et al. *Phys. Med. Biol.*, 52(23) (2007) 6923.
[7] F. Pfeiffer et al. *Nucl. Instrum. Meth. Phys. Res. A*, 580 (2007) 925.
[8] M. Bech et al. *Phys. Med. Biol.*, 54(9) (2009) 2747.
[9] T. Weitkamp et al. *Eur J Radiol*, 68 (2008) S13.
[10] C.-Y. Chou and M. A. Anastasio. *Med. Phys.*, 37(1) (2010) 270.

[3]The dynamic range of the phase gradient measurement for a setup using the first fractional Talbot distance is $\pm\pi$ phase difference per one half of the grating pitch of G_2 [2]. Larger gradients are wrapped into this interval. Since the phase shift decreases with increasing energy, the low-frequency cupping or capping artifact are more severe for low energies. Note that these artifacts are not visible in the experiments reported in [6] and [9] since the objects were embedded in a fluid rather than in air as in our simulation.

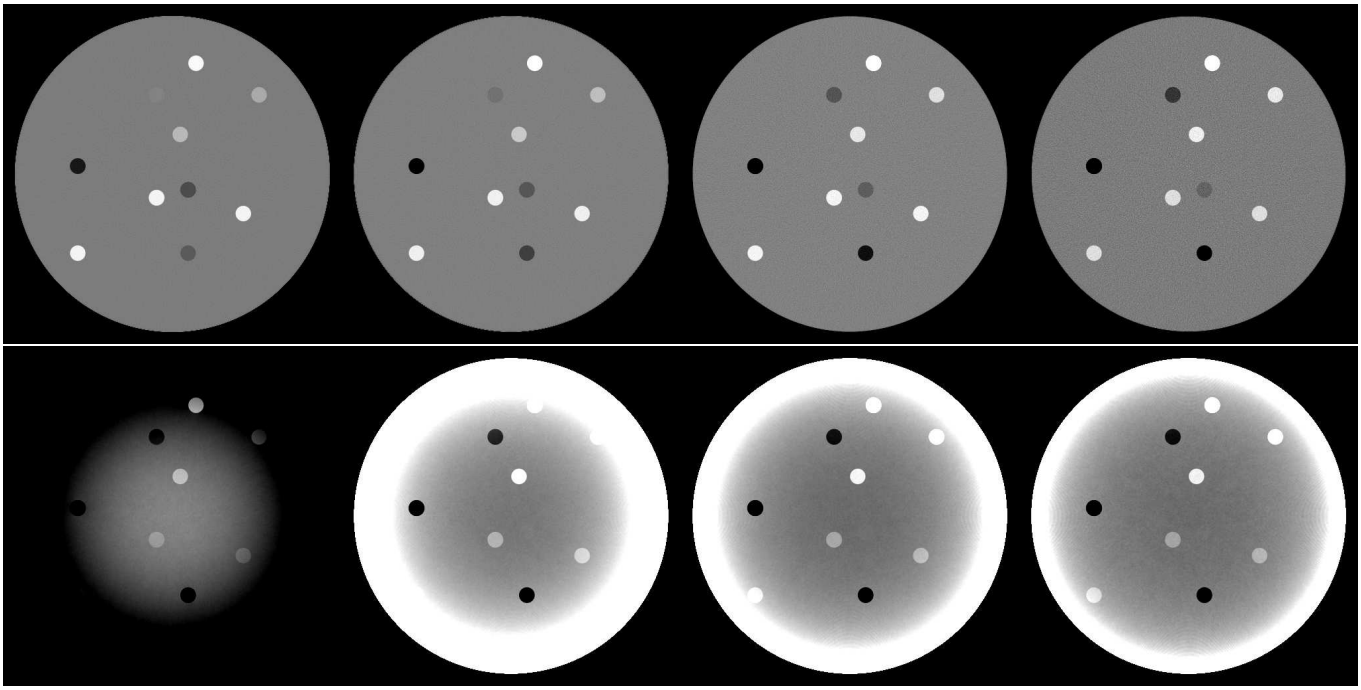


Fig. 7. Reconstruction results of the phantom simulated. From left to right 25, 45, 65, and 85 keV. The top row shows the absorption contrast images, the bottom row the reconstructed phase contrast images. Level and window are set as in Fig. 6.

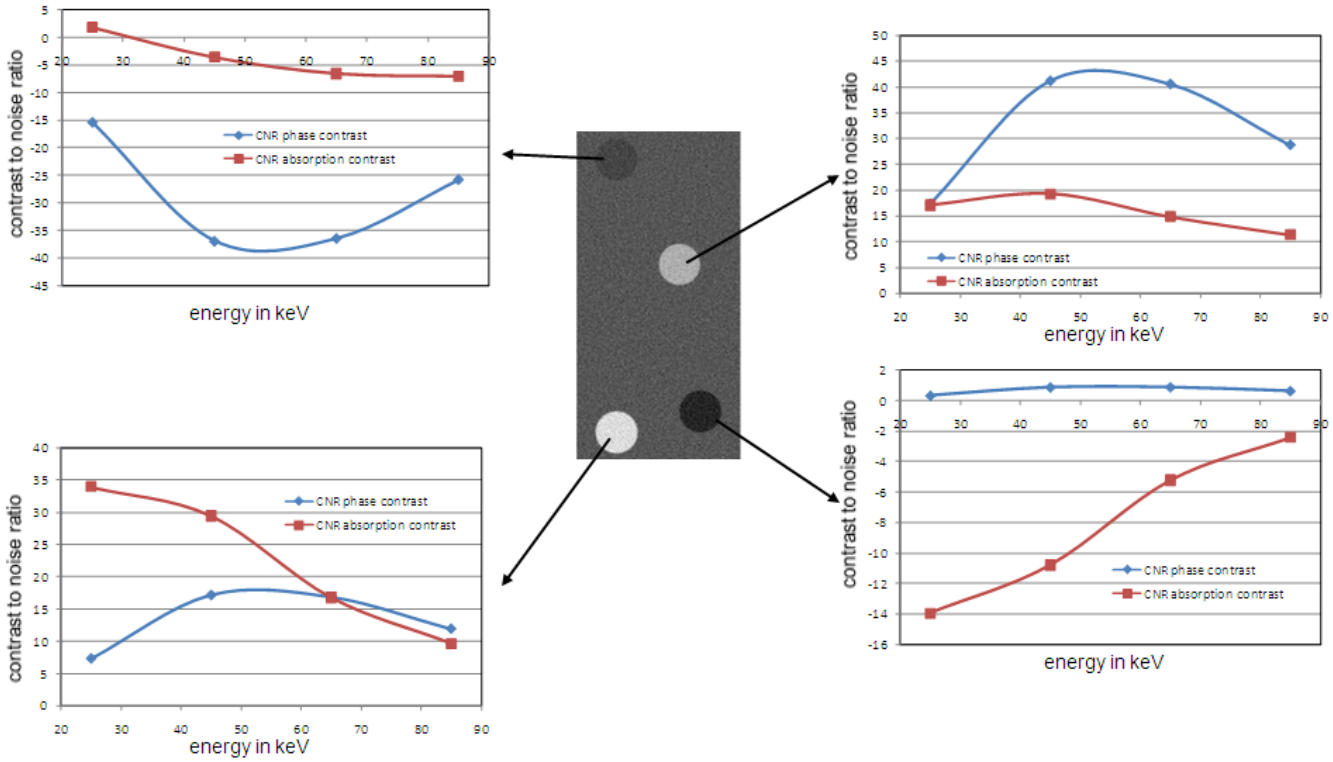


Fig. 8. CNR measurements for the objects within the ROI as a function of energy.

[11] Photon, electron, proton, and neutron interaction data for body tissues. ICRU report 46. International Commission on Radiation Units and Measurements, 1992.

[12] B. Henke et al. *Atomic Data and Nuclear Data Tables*, 54(2) (1993) 181.

[13] C. T. Chantler et al. X-ray form factor, attenuation, and scattering tables (version 2.1). National Institute of Standards and Technology, 2005.

[14] D. M. Paganin. *Coherent X-Ray Optics*. Oxford University Press, 2006.

[15] A. C. Kak and M. Slaney. *Principles of Computerized Tomographic Imaging*. IEEE Press, New York, 1987.

[16] G. W. Faris and R. L. Byer. *Applied Optics*, 27(24) (1988) 5202.

[17] J. Herzen et al. *Opt Express*, 17(12) (2009) 10010.

Local tomography and the motion estimation problem

Alexander Katsevich, Michael D. Silver, and Alexander A. Zamyatin

***Abstract*—We propose a novel motion estimation algorithm. It works with a sufficiently short time window, which is typically equivalent to that of a short scan. The center of the window is taken as reference time. The problem is to estimate motion within the window relative to the reference time. The proposed algorithm is based on local tomography (LT). If there is some uncompensated motion in f, the edges of f and of the LT function $\mathcal{B}f$ in general do not coincide. This means that if motion is not known (or, is known incorrectly), edges in the reconstructed image spread out. A single edge in f produces multiple edges in $\mathcal{B}f$ at random locations. Consequently, the reconstructed image will look chaotic. We can use a measure of chaos in the reconstructed image $\mathcal{B}f$ to tell whether our motion model is accurate or not. The main idea is to iteratively refine the motion model, so that chaos in $\mathcal{B}f$ is as small as possible. The algorithm is quite flexible and is used also for solving a misalignment correction problem. The results of numerical experiments on motion estimation and misalignment correction are very encouraging.**

***Index Terms*—motion estimation, local tomography, iterative reconstruction, cardiac CT.**

I. Introduction

Cardiac and, more generally, dynamic imaging is one of the top challenges facing modern computed tomography (CT). In cardiac CT there are two major groups of approaches for dealing with this issue. One is based on gating. The second approach, known as motion compensation, is based on incorporating a motion model into a reconstruction algorithm (see e.g. [1], [2], [3], [4]). Motion compensation algorithms are preferable, because they use all data and have the potential to provide good image quality with reduced x-ray dose. The main difficulty of using such algorithms is that motion model needs to be known. There are motion estimation algorithms available, see e.g. [1], [5], [6], [7], [8] for the most recent advances, but still significant research needs to be done to improve efficiency, accuracy, stability with respect to noise, flexibility, etc.

In this paper we propose a novel approach to motion estimation, which is based on local tomography (LT). We show that if any given edge in the unknown function f is seen from the data from two or more source positions, then in the case of incorrectly known motion, the single edge "spreads" and becomes a double edge. As a result, the image looks more cluttered. Our main idea is to iteratively improve the motion model so that image clutter is minimized. We propose an empiric measure of clutter, which we call "edge entropy". The use of LT provides a number of benefits in terms of computational efficiency, flexibility, and ability to handle truncated data.

Let us mention some other attractive features of the proposed approach. First, it is local in time. Motion estimation is done inside a reasonably short time window, e.g., not much longer than the length of a short scan. Second, the approach is fairly general and can be used for several types of motion, e.g., cardiac, breathing, etc. Finally, with simple modifications the approach can be applied to solving other practically important problems. As an example we show how to solve a misalignment correction problem for a distorted circular scan. A similar iterative algorithm, which is based on Feldkamp inversion formula, is proposed in [9]. Since our algorithm is based on LT, it is faster and it is not limited to a source trajectory for which accurate and efficient inversion formulas exist.

AK is with the Department of Mathematics, University of Central Florida, Orlando, FL 32816. MS and AZ are with Toshiba Medical Research Institute USA, Inc. Vernon Hills, IL 60061. Corresponding author: Alexander Katsevich, E-mail: akatsevi@mail.ucf.edu.

II. Local tomography (LT)

Let C be a smooth curve in $\mathbb{R}^3$

$$I \ni s \to y(s) \in \mathbb{R}^3, \ |y_s'(s)| \neq 0, \tag{1}$$

where $I \subset \mathbb{R}$ is an interval. Usually the source moves along C with constant speed, so we identify s with time variable.

Fix any $s_0 \in I$. We refer to $s = s_0$ as the reference time. Suppose $y = \psi(s, x)$ is the position of the particle at time s, which is located at x at the reference time $s = s_0$. We assume that for each $s \in I$ the function $\psi(s, x) : \mathbb{R}^3 \to \mathbb{R}^3$ is a diffeomorphism. Physically this means that two distinct points cannot move into the same position. This assumption is quite natural, since cardiac motion is not infinitely compressible (see e.g. [10]). The inverse of ψ is the function $x = \mu(s, y) : \mathbb{R}^3 \to \mathbb{R}^3$. It gives the original position x of the particle at the reference time, which is located at y at time s. We assume that both ψ and μ are identity maps outside of some open set U, which contains the support of the object, and $\psi, \mu \in C^\infty(I \times \mathbb{R}^3)$. As usual, we assume that C is at a positive distance from U. Obviously,

$$\mu(s, \psi(s, x)) \equiv x, \ \psi(s, \mu(s, x)) \equiv x. \tag{2}$$

In what follows we use the convention that the operator ∇ acts with respect to space variables. Thus $\nabla\mu(s, y) = \nabla_y \mu(s, y)$ and $\nabla\psi(s, x) = \nabla_x \psi(s, x)$.

Since matter is conserved, the x-ray density at time s and point y is given by $|\nabla\mu(s,y)|f(\mu(s,y))$. Hence the data are

$$D_f(s,\beta) := \int_0^\infty |\nabla\mu(s,y(s)+t\beta)|f(\mu(s,y(s)+t\beta))dt,\ s\in I, \tag{3}$$

where β runs through a subset of the unit sphere determined by the detector.

Introduce the following LT function

$$(\mathcal{B}f)(x) = \int_I \varphi(s,x)\frac{\partial^2}{\partial q^2}\ D_f(s,\beta(s,x+q\Theta(s,x)))\big|_{q=0}\,ds, \tag{4}$$

where

$$\beta(s,x) = \frac{\psi(s,x)-y(s)}{|\psi(s,x)-y(s)|}, \tag{5}$$

$\Theta(s,x) : I\times U \to \mathbb{R}^3\setminus 0$ is a smooth function, and $\varphi\in C_0^\infty(I\times U)$. The function $\Theta(s,x)$ specifies the direction along which the data are differentiated. We establish several properties of the LT function $\mathcal{B}f$. First, if motion is known, useful singularities of f and $\mathcal{B}f$ coincide (if they are visible from the data). Suppose that instead of the motion function ψ we know its approximation $\psi_\epsilon(s,x)=\psi(s,x)+\epsilon\psi_1(s,x)$. We show that in this case useful singularities of $\mathcal{B}f$ are shifted by a distance of order $O(\epsilon)$ away from those of f and obtain a formula for the shift.

III. A motion estimation algorithm

As stated in the previous section, if there is some uncompensated motion in f, the edges of f and $\mathcal{B}f$ no longer coincide and edges in the reconstructed image spread out. A single edge produces multiple edges at several locations. Consequently, the reconstructed image looks clattered or random. We can use a measure of randomness in the reconstructed image $\mathcal{B}f$ to gauge whether our motion model is accurate or not. In what follows we call this measure "edge entropy". Using this idea, we summarize the proposed motion estimation algorithm as follows.

1) Assume some motion model;
2) Perform motion-compensated LT image reconstruction using current motion model;
3) Compute edge entropy of the LT image;
4) If edge entropy is low (i.e., the edges have not spread too much), stop. If edge entropy is high, change the motion model and go to step (2).

A similar idea was used in [9] for misalignment correction in circular cone beam CT. The main novelty of our approach is that we use LT instead of global Feldkamp type (FDK) reconstruction. On one hand, the use of LT allows us to significantly speed up the iterations. On the other hand, many tools that work with conventional images (most notably, image entropy) do not work with LT images, so we had to develop alternative tools from scratch. We now describe the key steps of the algorithm in more detail.

A. Motion model

Let $[s_l,s_r]\subset I$ be a parameter/time window, which is used for motion estimation. The center point $s_0=(s_l+s_r)/2$ is taken as reference time. The primary purpose of the algorithm is to perform local (in time) motion estimation, thus the width of the window $S:=s_r-s_l$ is usually rather short. In our experiments S is typically less than one gantry rotation. Let $D\subset U$ denote the region where motion takes place. We assume that D is a rectangle, i.e. $D:=\{(x_1,x_2,x_3)\in\mathbb{R}^3 : L_k\le x_k\le R_k, k=1,2,3\}$. To represent motion, we consider a regular grid over D. The grid planes are

$$x_k=\zeta_{ik}:=L_k+i\Delta x_k,\ 0\le i\le N_k+1,\ k=1,2,3, \tag{6}$$

where $\Delta x_k=(R_k-L_k)/(N_k+1)$ is the step-size along the k-th axis. Thus, grid (6) has $N_1N_2N_3$ interior nodes, and for each direction k there are N_k+2 planes $x_k=\zeta_{0k},\cdots,x_k=\zeta_{N_k+1,k}$. Because of motion, the grid planes deform over time:

$$\begin{aligned}
x_1 &= \zeta_{i1}+a_{i1}(s)\phi[(x_2-L_2)/(R_2-L_2)]\\
&\quad\times\phi[(x_3-L_3)/(R_3-L_3)],\ 1\le i\le N_1,\\
x_2 &= \zeta_{i2}+a_{i2}(s)\phi[(x_1-L_1)/(R_1-L_1)]\\
&\quad\times\phi[(x_3-L_3)/(R_3-L_3)],\ 1\le i\le N_2,\\
x_3 &= \zeta_{i3}+a_{i3}(s)\phi[(x_1-L_1)/(R_1-L_1)]\\
&\quad\times\phi[(x_2-L_2)/(R_2-L_2)],\ 1\le i\le N_3.
\end{aligned} \tag{7}$$

Each line in (7) defines a separate surface, which corresponds to a deformation of one of the original planes (6). We assume that motion equals zero at the boundary of D, so the boundary grid planes (i.e. those given by $x_k=\zeta_{ik}$, $i=0$ or N_k+1, $k=1,2,3$) do not deform. In (7), the function ϕ is smooth, defined on the interval $[0,1]$, and equals zero at both endpoints of the interval. Since the time window $[s_l,s_r]$ is sufficiently short, we assume that the functions $a_{ik}(s)$ are linear:

$$a_{ik}(s)=a_{ik}(s-s_0)/(0.5S), k=1,2,3, \tag{8}$$

where $a_{ik}, 1\le i\le N_k$, $k=1,2,3$, are constants to be determined. Equations (7) and (8) allow us to describe motion of every point in D. To determine where a node from the original grid (6) is located at time s, we identify the three planes where the node is located, deform them according to (7), and then find the point of intersection of the three resulting surfaces. Location of all other pixels is computed using trilinear interpolation.

B. Edge entropy

Suppose $\mathcal{B}f$ is computed on a regular grid $(x_{i_1},x_{i_2},x_{i_3}), 1\le i_k\le M_k$, $k=1,2,3$, which covers D. Suppose, for simplicity, that the step-size of the grid is the same along every axis and equals Δx. Nodes of the grid are denoted $x_I:=(x_{i_1},x_{i_2},x_{i_3})$, where $I=(i_1,i_2,i_3)$. Of course, this grid should be much more dense than the one in (6) ($M_k\gg N_k$). We also need a shifted grid with nodes $\bar{x}_I:=(\bar{x}_{i_1},\bar{x}_{i_2},\bar{x}_{i_3})$, where $\bar{x}_{i_k}=x_{i_k}+\Delta x/2$, $1\le i_k\le M_k-1$, $k=1,2,3$. Introduce the distance function:

$$\mathrm{dist}(\bar{x}_I,\bar{x}_J)=\max(|i_1-j_1|,|i_2-j_2|,|i_3-j_3|). \tag{9}$$

Calculation of edge entropy consists of several steps. Let parameter $\kappa, 0<\kappa<1$, be fixed.

1) Using finite differences, compute the norm of the gradient at the nodes of the shifted grid $|\nabla(\mathcal{B}f)(\bar{x}_I)|$;

2) Compute the empirical histogram of the norm of the gradient;
3) Using the histogram, estimate the value M such that $|\nabla(\mathcal{B}f)(\bar{x}_I)| > M$ for 100κ percent of the points (such points are called "bright");
4) By running a sliding window over the image compute the total number of points $\bar{x}_I$ whose distance (in the sense of (9)) to the closest bright point equals either 2, 3, or 4;
5) Divide this number by the total number of nodes in the grid and multiply by 100 (to get percents). The result is the edge entropy of the image $\mathcal{B}f$.

IV. Numerical experiments

The original phantom is a superposition of seven balls (to save on space, images of the phantom are not shown here). The motion of the medium is described by the function

$$\psi(s,x) = x + \begin{cases} 25\cos(0.35(s-s_0))\Theta, & |x| < 10, \\ 25\cos(0.35(s-s_0))\frac{75-|x|}{65}\Theta, & 10 \le |x| < 75, \\ 0, & |x| \ge 75; \end{cases}$$
$$\Theta = (\cos\theta_2\cos\theta_1, \cos\theta_2\sin\theta_1, \sin\theta_2). \tag{10}$$

Here s_0 is reference time, $\theta_1 = 70^\circ$, $\theta_2 = 30^\circ$. In this section the units of length are always mm. The detector array is curved and passes through the isocenter. Pixel size on the detector is 0.5 along columns, and 10^{-3} radians along rows. The source trajectory is circular: $x_1 = R\cos s, x_2 = R\sin s, x_3 = 0$, and the source to isocenter distance is $R = 600$. There are 1000 projections per one rotation, $0 \le s < 2\pi$. The time corresponding to projection 501 was chosen as reference time: $s_0 = \pi$. For motion estimation we used only the data corresponding to the range of projections [101,900]. Following the common practice in medical imaging, we did not track the changes in density due to motion (see e.g., [10], [11]). This is equivalent to setting $|\nabla\mu| = 1$ in equation (3).

The function $\mathcal{B}f$ is computed on the $112\times112\times112$ regular grid covering the cube $-75 \le x_k \le 75$, $k = 1,2,3$. To make the resulting algorithm as numerically efficient as possible, we use the simplest version of the motion compensated LT. To this end the derivative $\partial^2/\partial q^2$ in (4) is replaced by the second derivative of the cone beam data along detector rows.

Let D_l be a box-like region bounded by six neighboring planes (6). As is easily seen, the values of $\mathcal{B}f(x)$ for all $x \in D_l$ depend only on the six parameters describing the deformation of the six planes that form its boundary. Minimization of edge entropy uses this observation and is done using the following approach.

Step 1. Let $a_{ik}, i = 1,\dots N_k, k = 1,2,3$, be the current best estimate of the motion parameters. Let some $\Delta a \ne 0$ be given. Pick one of the D_l's. Let $a_{i_1k_1},\dots,a_{i_6k_6}$ be the six parameters affecting the chosen region. Compute 3^6 subimages $\mathcal{B}f(x)$, $x \in D_l$, corresponding to the sets $\tilde{a}_{i_1k_1},\dots,\tilde{a}_{i_6k_6}$, where each $\tilde{a}_{ik}$ equals either a_{ik}, or $a_{ik}-\Delta a$, or $a_{ik}+\Delta a$. Store all the subimages on the disk, and repeat for all D_l's.

Step 2. Run the loop over all $3^{N_1+N_2+N_3}$ sets $a_{ik}, i = 1,\dots N_k, k = 1,2,3$, where each $\tilde{a}_{ik}$ equals either a_{ik}, or $a_{ik}-\Delta a$, or $a_{ik}+\Delta a$. This is done by reading the appropriate subimages from the disk and combining them into a single image of $\mathcal{B}f(x)$, $x \in D$. Then compute edge entropy for the obtained image. From the $3^{N_1+N_2+N_3}$ sets of parameters find the one which produces the image with the smallest entropy.

Steps 1 and 2 constitute a single iteration. The initial values of a_{ik} are chosen to be zero (which is the no motion assumption). The value of Δa is chosen from some *a priori* considerations. After the end of each iteration, the optimal set of parameters identified at Step 2 is passed on to Step 1. Also, the value of Δa is decreased. In our experiments we used $\Delta a_{\text{initial}} = 10$, $\Delta a_{\text{new}} = 0.75\Delta a_{\text{old}}$, and three iterations were performed.

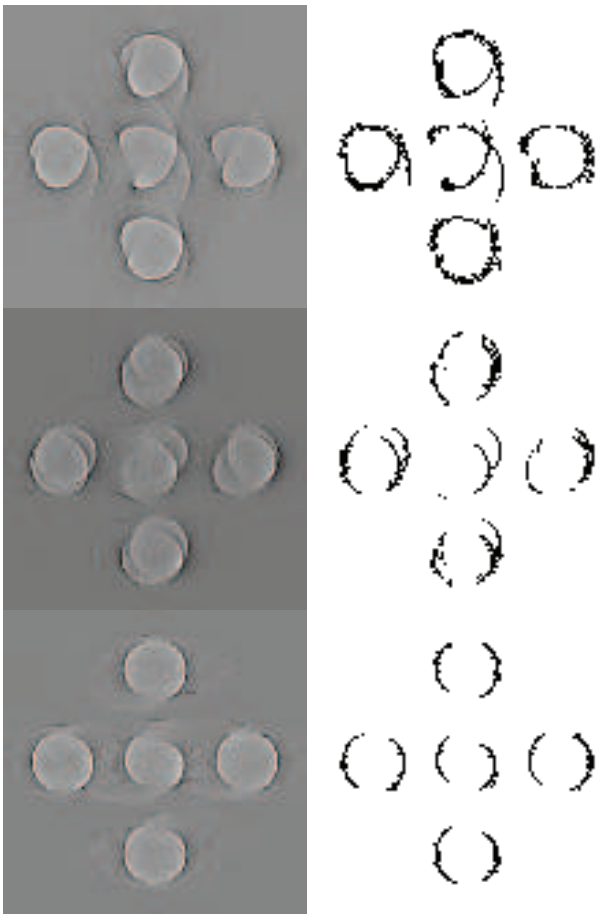


Fig. 1. Left panel: density plots of $\mathcal{B}f$ at the beginning of iterations, i.e. when zero motion is assumed. Right panel: corresponding images of bright points. Top to bottom: cross-sections x_1x_2, x_1x_3, and x_2x_3 through the center of the grid. Entropy 9.81%.

Results of experiments are shown in Figs 1 and 2. Fig. 1 shows the initial image of $\mathcal{B}f$ computed under the (incorrect) assumption of no motion. Fig. 2 shows the final image of $\mathcal{B}f$ computed for the motion model, that was determined by the algorithm. In these experiments we used $N_1 = N_2 = N_3 = 4$ (cf. (7)). We used $\kappa = 0.0125$ to compute edge entropy. At the beginning of iterations (Fig. 1) the value of entropy is 9.81%, and at the end - 8.40% (Fig. 2).

Application of the proposed approach for solving a misalignment correction problem will be presented at the conference. In the case of noise-free (or low noise) data, the

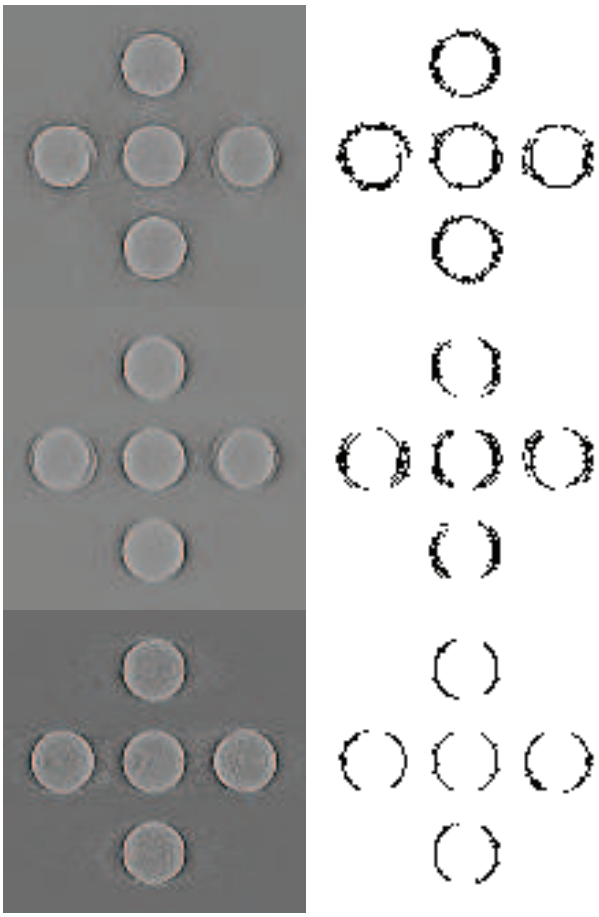


Fig. 2. Left panel: density plots of $\mathcal{B}f$ at the end of iterations. Right panel: corresponding images of bright points. Top to bottom: cross-sections x_1x_2, x_1x_3, and x_2x_3 through the center of the grid. Entropy 8.40%.

algorithm based on LT as outlined above works well. If data are sufficiently noisy, the proposed scheme is unstable. A more stable modification of the scheme and the results of testing on noisy data will also be presented at the conference.

ACKNOWLEDGMENT

The work of AK was supported by NSF grant DMS-0806304 and Toshiba Medical Research Institute USA, Inc

REFERENCES

[1] A. A. Isola, M. Grass, and W. J. Niessen, "Fully automatic nonrigid registration-based local motion estimation for motion-corrected iterative cardiac CT reconstruction," *Medical Physics*, vol. 37, pp. 1093–1109, 2010.

[2] U. van Stevendaal, C. Lorenz, J. von Berg, and M. Grass, "Motion-compensated reconstruction in helical cardiac CT," in *Proceedings of the IXth International Conference on Fully 3D Reconstruction in Radiology and Nuclear Medicine. Lindau, Germany. July 9 - 13, 2007*, 2007, pp. 437–440.

[3] K. Taguchi and H. Kudo, "Motion compensated fan-beam reconstruction for computed tomography using derivative backprojection filtering approach," in *Proceedings of the IXth International Conference on Fully 3D Reconstruction in Radiology and Nuclear Medicine. Lindau, Germany. July 9 - 13, 2007*, 2007, pp. 433–436.

[4] L. Desbat, S. Roux, and P. Grangeat, "Compensation of some time dependent deformations in tomography," *IEEE Transactions on Medical Imaging*, vol. 26, pp. 261–269, 2007.

[5] C. Blondel, G. Malandain, R. Vaillant, and N. Ayache, "Reconstruction of coronary arteries from a single rotational X-ray projection sequence," *IEEE Transactions on Medical Imaging*, vol. 25, pp. 653–663, 2006.

[6] R. Zeng, J. A. Fessler, and J. M. Balter, "Estimating 3-d respiratory motion from orbiting views by tomographic image registration," *IEEE Transactions on Medical Imagaging*, vol. 26, pp. 153–163, 2007.

[7] M. Reyes, G. Malandain, P. M. Koulibaly, M. A. Gonzalez-Ballester, and J. Darcourt, "Model-based respiratory motion compensation for emission tomography image reconstruction," *Physics in Medicine and Biology*, vol. 52, pp. 3579–3600, 2007.

[8] M. Wierzbicki, G. M. Guiraudon, D. L. Jones, and T. Peters, "Dose reduction for cardiac CT using a registration-based approach," *Medical Physics*, vol. 34, pp. 1884–1895, 2007.

[9] Y. Kyriakou, R. M. Lapp, L. Hillebrand, D. Ertel, and W. Kalender, "Simultaneous misalignment correction for approximate circular cone-beam computed tomography," *Physics in Medicine and Biology*, vol. 53, pp. 6267–6289, 2008.

[10] C. Blondel, R. Vaillant, G. Malandain, and N. Ayache, "3D tomographic reconstruction of coronary arteries using a precomputed 4D motion field," *Physics in Medicine and Biology*, vol. 49, pp. 2197–2208, 2004.

[11] S. Rit, D. Sarrut, and L. Desbat, "Comparison of analytic and algebraic methods for motion-compensated cone-beam CT reconstruction of the thorax," *IEEE Transactions on Medical Imaging*, vol. 28, pp. 1513–1525, 2009.

Using the Positivity Constraint to Enhance Temporal Resolution in CT

Harald Schöndube, Holger Kunze, Herbert Bruder and Karl Stierstorfer

***Abstract*—We present a new idea to enhance temporal resolution in CT beyond the short-scan limit by making use of a positivity constraint. If the moving objects are seen against an empty background, it can be demonstrated that an iterative reconstruction scheme which makes use of the positivity constraint in the image domain can image objects sharply even when they are stationary only for one third of the rotation time of the CT scanner. This idea can be propagated further by decomposing an object consisting of small moving parts against a large stable background into these components and making use of the positivity constraint in for the reconstruction of the moving part.**

***Index Terms*—Computed Tomography, Temporal Resolution, Positivity Constraint, Iterative Reconstruction**

I. Introduction

IMAGING the heart remains a challenge for Computed Tomography, even after the introduction of fast-rotating multislice scanners. Dual-source CT were introduced which consistently cuts temporal resolution by half [1]. Downside of these scanners, of course, is the substantial hardware effort needed which makes this technology prohibitive for the budget minded customer, which prefer single source multislice scanners. For these devices, different iterative reconstruction algorithms were proposed. Besides [2] motion compensated reconstruction, which uses all projections for the final reconstruction, but requires the time consuming estimation of motion vector fields, a compressed sensing technique [3] was recently proposed to increase temporal resolution. This technique uses a sparsifying transform to optimize iteratively a combination of a smoothness constraint, a constraint of a prior image and a constraint matching the forward projection of the image to data less than a short-scan.

Goal of this paper is to propose an alternative algorithm which is also based upon using as little data as possible for the final reconstruction but is potentially more robust and easier to implement than the other two algorithms.

II. Basic Idea of the Algorithm

The basic idea of the algorithm is simple: Suppose we were imaging a set of small moving objects without any background attenuation. In the images reconstructed with FBP, motion artifacts would show up as positive and negative patterns around the moving objects. Using an iterative reconstruction scheme which simply requires the image to match data from a subset of a short-scan interval (e.g. 120°) will not help to suppress these artifacts, because they are consistent with the data: positive and negative parts will approximately cancel in the forward projection. However, this cancellation will be thrown off balance if we impose a positivity constraint by simply setting the negative parts to zero after each iteration. This will have the effect that also the positive parts of the artifacts reduce in order to match the data with the forward projection of the image. As we will demonstrate in this paper, this will lead to a substantial improvement in image quality.

Using Iterative Filtered Backprojection [4] as the standard reconstruction algorithm, the update formula of the iteration would then be

$$\begin{aligned} f_0 &= FBP(D_{short-scan}), \\ \hat{f}_{k+1} &= f_k + \alpha\, FBP\left(D_{subset} - \mathrm{Proj}(f_k)\right), \\ f_{k+1} &= \max(\hat{f}_{k+1}, 0). \end{aligned} \tag{1}$$

Here, D stands for projection data of a short-scan interval or a subset of it, depending on the index. FBP is the Filtered Backprojection operator. Proj(f_k) denotes the forward projection operator, α is an iteration relaxation factor. The projection based update formula is alternated with imposing the non-negativity constraint.

Our idea can also be used with other iterative reconstruction algorithms, such as SART and SIRT [5]. In this case, starting with an empty estimate image f_0 (i.e., an image that contains only zeros) is also possible.

III. Extension of the Algorithm to More General Objects

In clinical practice, of course, the scanned object is not a collection of moving objects with no background attenuation. Usually, we face a set of small objects superimposed over relatively smooth background attenuation. In this section we sketch two approaches how the basic idea of the algorithm can be applied to more realistic data sets.

A. Sinogram Decomposition

Our first approach was to separate moving and non-moving objects in the sinogram. The first step is then to decompose the sinogram D into two sinograms:

All authors are with Siemens Healthcare, Siemensstr. 1, D-91301 Forchheim, Germany. E-mail: harald.schoendube@siemens.com

(2) $D = D_{low} + D_{high}$,

so that

(3) $D_{high} > 0$.

Thus we obtain the attenuation from the moving objects in the partial sinogram D_{high} while the nearly stationary background containing only low spatial frequencies is contained in D_{low}. Once this is achieved, we can reconstruct the two parts separately: D_{low} is reconstructed in a standard way, e.g. using an FBP short-scan reconstruction. With D_{high}, we can proceed as sketched in section II. Essential for this, of course, is the positivity constraint of equation (3). One additional requirement should be that the low frequency sinogram D_{low} is consistent in the sense that its reconstruction with FBP yields an object with finite support. The final image is then the sum of the two partial images.

B. Using a Histogram Constraint

A reliable decomposition of the sinogram into a moving and a non-moving part proves to be difficult. Therefore we propose to use an extension of the positivity constraint to a more general histogram constraint. In [5], [6] it was shown that a good reconstruction of objects that are located outside the boundary of the CT field-of-view (FOV) can be obtained by combining an iterative reconstruction with a histogram constraint. The algorithm presented in these publications partitions the image into regions inside and outside the FOV and uses the image gray-scale distribution inside of the FOV as a-priori information for the image values outside of the FOV for a statistical reconstruction algorithm.

Our approach, which we have dubbed "Temporal Resolution Improvement Method" (TRIM), employs this principle for the goal of a enhancing the temporal resolution of CT images. A flow chart of the algorithm is shown in figure 1: First, a cardiac short-scan reconstruction of the moving object is performed, resulting in a temporal resolution of 180° of rebinned data. This 180° image is partitioned into smaller regions, for each of those regions the corresponding histogram is computed. Afterwards, an iterative reconstruction using only a subset of the short-scan interval is performed. After each iterative step an additional step is performed where the gray scale values in each image region are manipulated such as to approximately correspond to the respective histogram of the 180° image. The underlying assumption here is that the shape of the histogram of the image does not change significantly in the presence of motion artifacts.

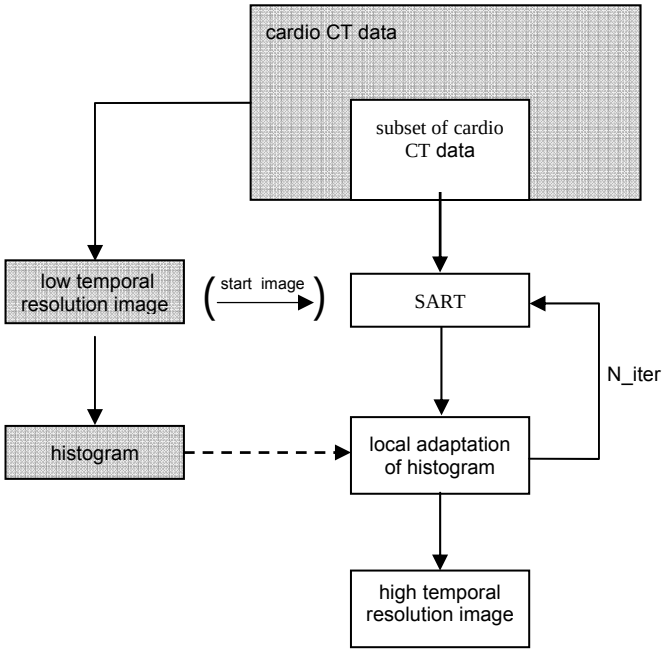


Figure 1: Flow chart of the TRIM algorithm.

IV. Results

We have performed some preliminary studies to evaluate the feasibility of our idea and have obtained encouraging results. Some of the experiments we have performed will be presented in this section.

A. Reconstruction with Positivity Constraint

For our experiments involving the positivity constraint discussed in section II we have used two different simulated moving objects, which we have reconstructed using a rebinned two-dimensional SART algorithm. The positivity was enforced after every iterative step. Our first experiment involved a phantom which consists of several high-contrast inserts representing vessels filled with contrast-agent. The high-contrast objects remain at fixed positions for a scan-range of 120° and then start a linear movement towards the image center. A depiction of an FBP reconstruction of the phantom from *static* data serving as a reference image is shown in figure 2. An FBP cardiac short-scan reconstruction of the phantom from data with simulated object movements is depicted in figure 3, while an SART reconstruction with positivity constraint from 90° of rebinned data after seven iterations is shown in figure 4. The same two images using a different, narrower gray-level window are again shown in figures 5 and 6.

In the SART reconstructed images, a significant reduction of motion artifacts can be observed. It is also discernable that the limited angle of projections used for the SART image leads to a slightly non-isotropic resolution. However, the advantage of having virtually no motion artifacts in the SART reconstruction should prevail over the disadvantage of getting a non-isotropic resolution.

The second phantom was employed to study how internal structures of moving objects are reproduced by the method. It consists of a cylinder of the same density as the high contrast objects of the first phantom located near the isocenter with a smaller cylindrical off-center insert (see figure 7 for a reference FBP reconstruction from non-moving data). In the data sets with simulated motion, the object was held static over a scan range of 135° and then started a linear motion similar to the small objects of the first phantom.

We again display the reconstructions of the second phantom from moving data in two different gray scales to demonstrate

the effect of using a positivity constraint. FBP cardiac short-scan reconstructions are shown in figures 8 and 10, while SART reconstructions using the positivity constraint from 120° of rebinned data after 15 iterations are depicted in figure 9 and figure 11.

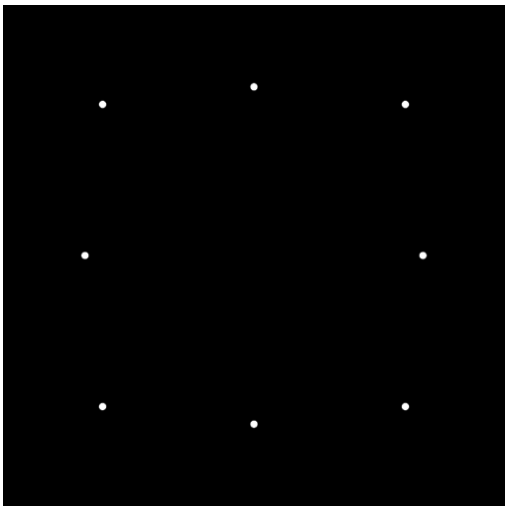


Figure 2: Reference FBP reconstruction of the first phantom from *static* data.

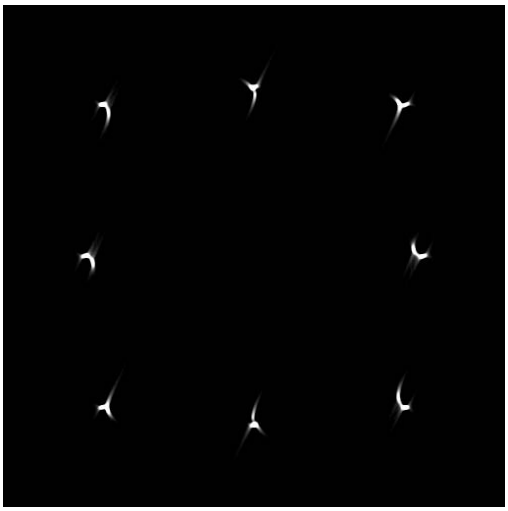


Figure 3: FBP cardiac short-scan reconstruction of the first phantom from moving data. Window: [-950 HU, -250 HU]

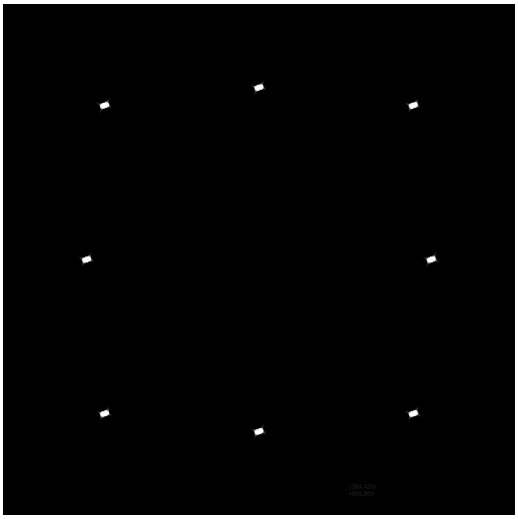


Figure 4: SART reconstruction with positivity constraint of the first phantom. Window: [-950 HU, -250 HU]

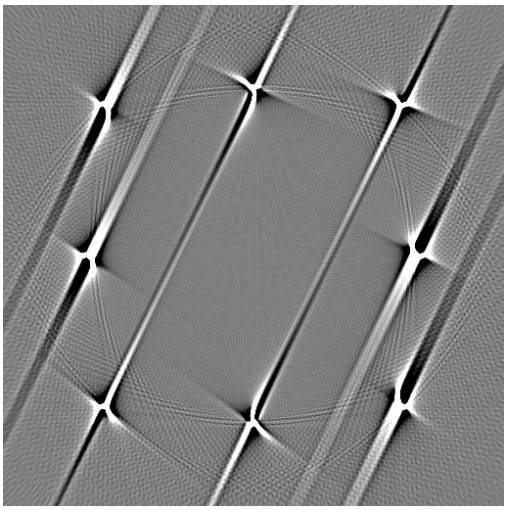


Figure 5: FBP cardiac short-scan reconstruction of the first phantom. Window: [-1050 HU, -950 HU]

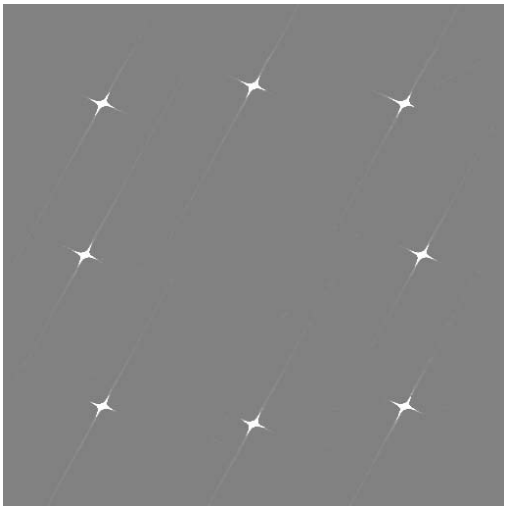


Figure 6: SART reconstruction with positivity constraint of the first phantom. Window: [-1050 HU, -950 HU]

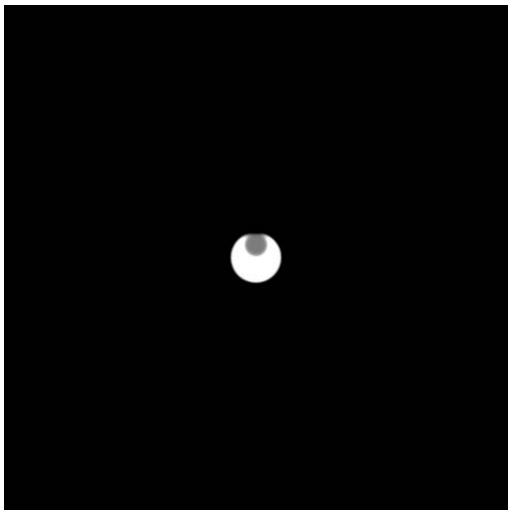


Figure 7: FBP reference reconstruction of the second phantom from static data, window: [0 HU, 1400 HU]

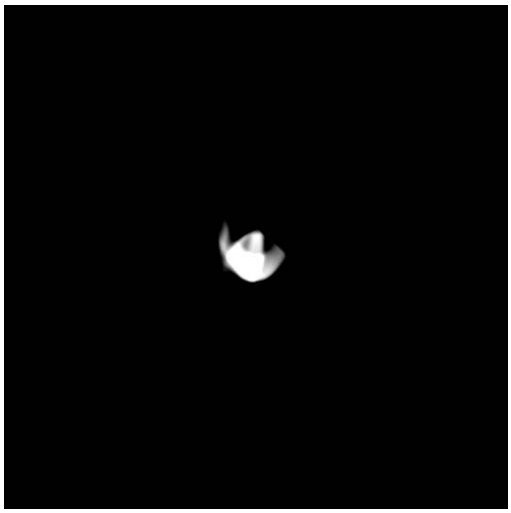


Figure 8: FBP cardiac short-scan reconstruction of the second phantom from moving data, window: [0 HU, 1400 HU]

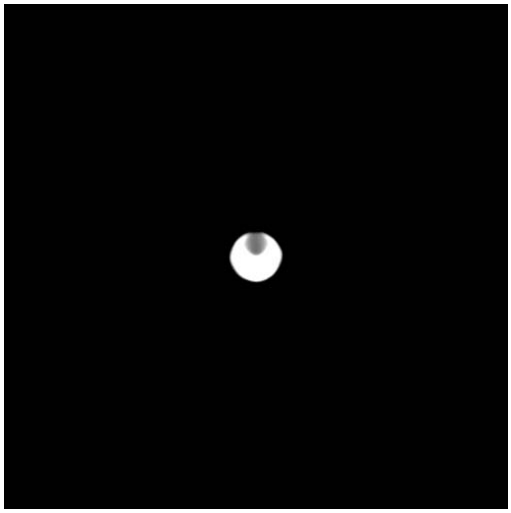


Figure 9: SART reconstruction with positivity constraint of the second phantom, window: [0 HU, 1400 HU]

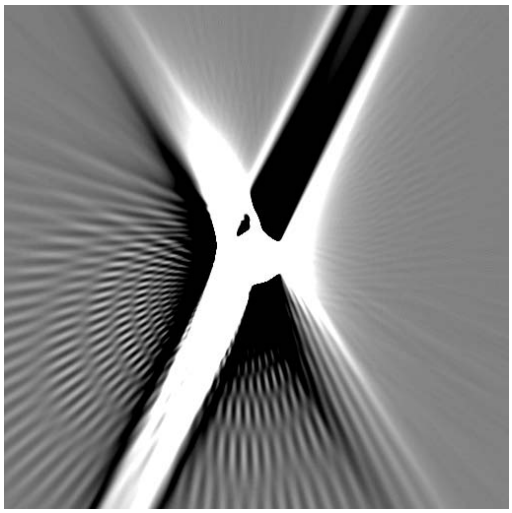


Figure 10: FBP cardiac short-scan reconstruction of the second phantom, window: [-1050 HU, -950 HU]

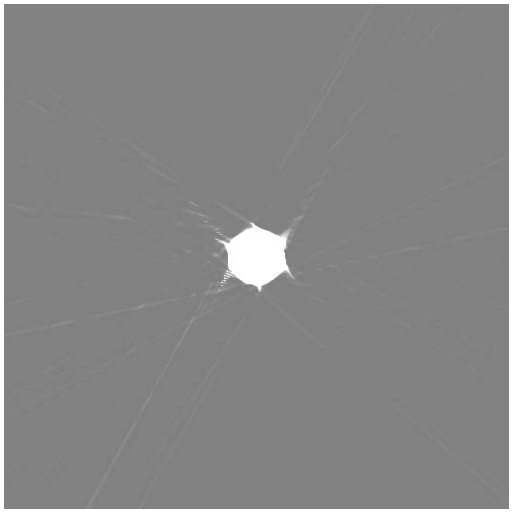


Figure 11: SART reconstruction with positivity constraint of the second phantom from moving data, window: [-1050 HU, -950 HU]

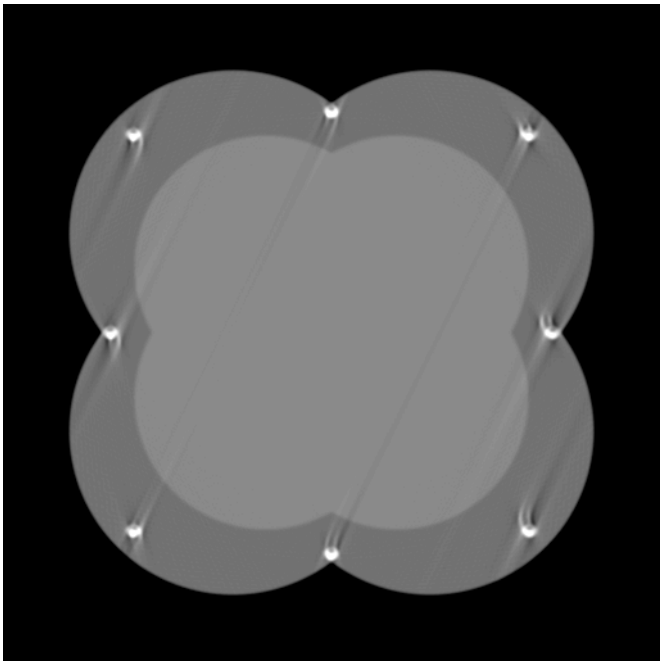


Figure 12: FBP cardiac short scan reconstruction of the moving phantom with background, window: [-700 HU, 700 HU]

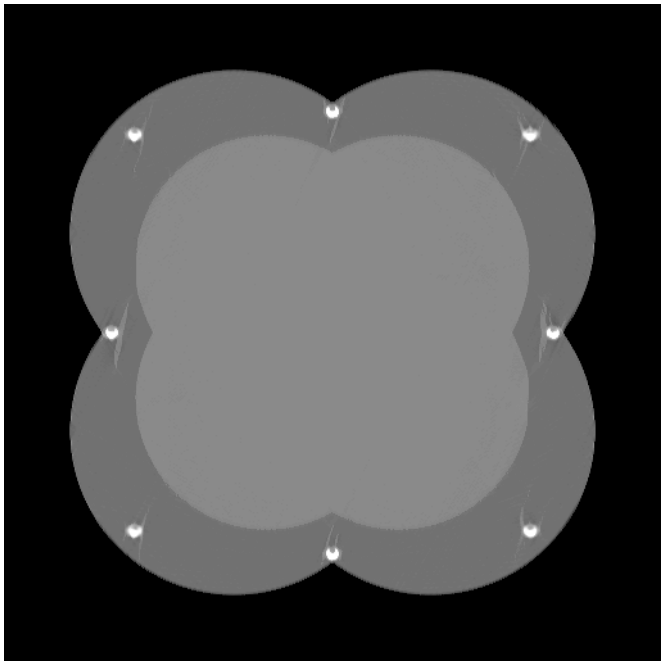


Figure 13: TRIM reconstruction of the moving phantom with background, window: [-700 HU, 700 HU]

B. TRIM Reconstructions

For a preliminary evaluation of the TRIM algorithm we modified the vessel phantom that we used for the first experiment above by adding some background structure at two different attenuation values. Furthermore, we added small off-center inserts to the moving vessels (similarly as in the second phantom above). The high-contrast objects remained again at a fixed position for a scan-range of 120° and then started moving towards the image center at constant speed.

Figure 12 displays a FBP cardiac short-scan reconstruction, while the result of the TRIM algorithm using 120° of rebinned data is shown in figure 13. A clear reduction of motion artifacts in the TRIM image compared with the FBP cardiac short-scan image can be observed.

V. Conclusion

We have presented a new idea to perform a CT reconstruction with enhanced temporal resolution from less than 180° data. Our idea is based on enforcing a positivity or histogram constraint in an iterative reconstruction algorithm for small moving objects. Our primary evaluations have yielded promising results.

The main task is now to further refine the algorithm, especially with a view to processing cone-beam data, as well as to perform further evaluations with simulated and clinical data. However, since this abstract constitutes a report about a work in progress, we hope to be able to report on progression on these tasks by the time of the conference.

Acknowledgment

The authors want to thank Frédéric Noo for valuable discussions.

References

[1] T. G. Flohr, H. Bruder, K. Stierstorfer, M. Petersilka, B. Schmidt, C. H. McCollough, “Image reconstruction and image quality evaluation for a dual source CT scanner”, Med. Phys. **35**, 2008, pp. 5882-5897.

[2] A A Isola, A Ziegler, T Koehler, W J Niessen, M Grass: "Motion-compensated iterative cone-beam CT image reconstruction with adapted blobs as basis functions", Phys. Med. Biol. **53**, 2008, pp.6777-6797

[3] G-H. Chen, J. Tang, J. Hsieh, “Temporal resolution improvement using PICCS in MDCT cardiac imaging”, Med. Phys. **36**, 2009, pp. 2130-2135.

[4] J. Sunnegård, “Iterative Backprojection Methods for Helical Cone-Beam CT”, Dissertation No. 1264, Institute of Technology, Linköping (2009)

[5] H. Kunze, “Iterative Rekonstruktion in der Medizinischen Bildverarbeitung”, Dissertation, University Erlangen – Nürnberg (2007), ISBN 978-3-8322-6866-4

[6] H. Kunze, W. Härer, K. Stierstorfer: Iterative Extended Field of View Reconstruction, Medical Imaging 2007: Physics of Medical Imaging, Proc. of SPIE Vol. 6510

Simple ROI Cone–Beam Computed Tomography

Michael Knaup, Clemens Maaß, Stefan Sawall, and Marc Kachelrieß

***Abstract*—High spatial resolution and large field of measurement are often contradictory demands, especially in x–ray cone–beam CT. On the one hand the number of detector elements are limited to typically 1024×1024 to 4096×4096. On the other hand CT requires to completely cover the lateral field of measurement (FOM) with the detector. If the number of detector elements in the lateral direction is called M and the diameter of the field of measurement is D the spatial resolution that can be achieved is in the order of D/M.**

Zooming into an object by a factor of 10, say, which can be done by decreasing the distance of the focal spot to the isocenter by the same factor, however yields truncated projections. We developed and implemented three methods that use a priori information from a low resolution overview scan to compensate for the data missing in the high resolution scan. These are the data completion, the data filtering, and the data weighting method. Thereby we were aiming at a robust and efficient solution with high image quality and high computational performance.

I. Introduction

AIMING at high spatial resolution in objects with a large transversal diameter requires to challenge the problem of transversal data truncation. For example the specific case we are interested in are objects fitting into a $D_{\mathrm{L}} = 60$ mm diameter field of measurement (FOM) that shall be scanned with a circular cone–beam CT scanner whose flat detector consists of 2000×2000 elements. A standard scan would allow us to achieve a spatial resolution of roughly 30 µm assuming the focal spot size to be small enough. However, we are aiming at high–resolution imaging a $D_{\mathrm{H}} = 6$ mm diameter region of interest, the ROI, with roughly 3 µm spatial resolution. To zoom into the object and increase the scanner's magnification by a factor of 10 the distance of the focal spot to the isocenter can be reduced by the same factor, for example. The projection data p_{H} of this high resolution scan are, however, truncated in the lateral direction. To compensate for the information missing in the high resolution scan we use the projection data p_{L} of the standard or overview scan, which we will refer to as the low resolution scan in the following.

For some objects or scanner geometries, and for large zoom factors, it may happen that the x–ray source or the detector would collide with the object during the high resolution circle scan. To workaround this problem one could use object–dependent non–circular trajectories, as we did propose in reference [1] several years ago. Due to practical reasons and due to the scanner design constraints we are, however, restricted to certain scan trajectories (in our case circle scans or spiral scans). To avoid collision we can neither acquire a full 360° data set nor even a 180° data set, in general. Hence it may happen that the high resolution scan is a limited angle scan that further suffers from lateral data truncation.

Dr. Michael Knaup, Clemens Maaß, Stefan Sawall, Prof. Dr. Marc Kachelrieß: Institute of Medical Physics (IMP), University of Erlangen–Nürnberg, Henkestr. 91, 91052 Erlangen, Germany.

Corresponding author: marc.kachelriess@imp.uni–erlangen.de

Fig. 1. State–of–the–art dimensional CT scanner used for our studies. The scanner is capable of CT rasterization and of ROI tomography. It acquires in circular and spiral mode.

The literature describes several methods to do region–of–interest CT, which means to perform reconstruction from truncated data. All methods deal with the truncation problem and do not allow for a limited angle high resolution scan.

All practical solutions described make use of the low resolution data p_{L} to complete the data missing in p_{H}. This is either done in the rawdata domain itself, e.g. by rebinning, or it is done by reconstructing the low resolution data to obtain a low resolution volume f_{L} which then can be forward projected using the geometry of the high resolution scan to provide rawdata that completes p_{H} [2], [3], [4], [5], [6], [7].

Further on, there are numerous methods that seek reconstructing from the high resolution data only without having low resolution overview scans available. Those are either of truncation–correction type with some kind of extrapolation designed for diagnostic purposes which yield less quantitative results [8], [9], [10], [11], [12], [13] or they are mathematical tweaks including some approximation or highly restrictive a priori assumptions [14], [15], [16]. A promising new interior tomography approach is based on compressed sensing which, however, requires long reconstruction times and the assumption of piecewise constant objects [17], [18], [19]. Due to these undesired properties these methods are considered to be impractical for our purposes and therefore not within the scope of this paper.

In principle, the first class of algorithms, that makes full use of the low resolution overview scan, provides adequate image quality for our purposes. However, one will run into

significant performance problems whenever the focal spot trajectory of the high resolution scan approaches the convex hull of the object. In this case the fan angle of the completed high resolution scan approaches 180° and a very large number of additional detector channels would be needed to complete the truncated region of the high resolution scan.

While one can certainly sidestep those issues, e.g. by using a virtual curved detector in place of a flat detector or by performing the forward projections on a sparse grid instead of using the high resolution grid, we propose two simple rawdata–based methods that require reconstructions of p_{L} but that do not require forward projections at all. It should be noted that our methods can easily be combined with the possibility of laterally shifting the detector to increase the object coverage or the spatial resolution by another factor of two [20], [21], [22], [23]. In this study, however, we did not make use of this option and only deal with data from non–shifted detectors.

II. Materials and Methods

We discuss the reconstruction of cone–beam CT data of the form

$$p(\alpha,\vartheta,\gamma)=\int_0^\infty d\lambda\, f(\boldsymbol{s}+\lambda\boldsymbol{\Theta})$$

with $\boldsymbol{s}=\boldsymbol{s}(\alpha)$ being the source position at time (or angle) α and

$$\boldsymbol{\Theta}=\begin{pmatrix}-\sin\vartheta\cos\gamma\\ \cos\vartheta\cos\gamma\\ \sin\gamma\end{pmatrix}$$

being the direction vector of a ray emerging at $\boldsymbol{s}(\alpha)$ in the direction specified by the angles ϑ and γ. The rawdata are denoted as $p(\alpha,\vartheta,\gamma)$ and the object to be reconstructed is $f(\boldsymbol{r})$. Due to the limited detector size we will conduct two scans of the above type. One overview scan which covers the complete object and thereby results in low resolution data and one high resolution scan that covers only an ROI of the object. We use the subscripts L and H to distinguish between both scans.

Aiming at Feldkamp–type approximate image reconstruction we assume the source trajectory to approximately lie in a plane parallel to the x–y–plane and we assume the rotation axis to be parallel to the z–axis. Note that even if the scanner is performing exact circle scans the rotation axes of the low resolution overview scan and the high resolution scan do not necessarily coincide because the rotation center of the high resolution scan is determined by the position and size of the ROI.

Since we perform approximate and Feldkamp–type image reconstruction the angle γ of the ray is used for length correction only. It does not play a role in the data consistency and weighting criteria discussed below. Therefore we may safely drop γ in the following and restrict ourselves to in–plane considerations, i.e. to the two dimensions x and y.

To achieve data consistency and to define appropriate weighting functions it is necessary to parameterize a ray by its angle ϑ and its distance ξ with respect to the origin of the coordinate system. While ϑ is already well defined we obtain ξ as a function of $\boldsymbol{s}$ and ϑ as

$$\xi=\boldsymbol{s}\cdot\boldsymbol{\vartheta}=s_x\cos\vartheta+s_y\sin\vartheta$$

where we defined

$$\boldsymbol{\vartheta}=\begin{pmatrix}\cos\vartheta\\ \sin\vartheta\end{pmatrix}.$$

The ray specified by ϑ and ξ is the line

$$x\cos\vartheta+y\sin\vartheta=\xi.$$

Image reconstruction can be performed if the data are complete and if data redundancies are properly normalized by defining a weighting function $w(\vartheta,\xi)$ that is zero wherever rays are missing and that fulfills

$$\sum_k w(\vartheta+k\pi,(-)^k\xi)=1, \tag{1}$$

as discussed in reference [23].

Although the methods presented work for general focal spot trajectories the CT scan modi we have in mind rather perform approximate circular scans. Restricting ourselves to such trajectories later allows us to formulate explicit equations for the weighting functions, which is more convenient for the reader. Therefore we introduce the circle trajectories

$$\boldsymbol{s}_{\mathrm{L}}(\alpha)=\begin{pmatrix}R_{\mathrm{FL}}\sin\alpha\\ -R_{\mathrm{FL}}\cos\alpha\\ 0\end{pmatrix}\quad\text{and}\quad\boldsymbol{s}_{\mathrm{H}}(\alpha)=\begin{pmatrix}R_{\mathrm{FH}}\sin\alpha\\ -R_{\mathrm{FH}}\cos\alpha\\ 0\end{pmatrix}+\boldsymbol{o}$$

with R_{F} denoting the radii of the low and the high resolution scan trajectories, respectively, and with $\boldsymbol{o}$ being the isocenter of the high resolution scan. We make use of these when defining the weight functions below.

In total we compare three different methods to perform ROI tomography when sufficient overview data are present: the data completion method which is in wide use already, the data filtering method, and the data weighting method. To our knowledge, the last two methods are new.

A. Data Completion Method

The classical data completion methods perform a forward projection of a low resolution overview volume f_{L} to complete data missing in the high resolution scan. While this is typically done to complete truncated data the same procedure could be used to solve the limited angle problem. As mentioned above the data completion may suffer from low performance and require a significant amount of memory.

To become more formal let us introduce some notation. Let X_{L} denote the x–ray transform corresponding to the overview scan and let $f_{\mathrm{L}}=\mathsf{X}_{\mathrm{L}}^{-1}p_{\mathrm{L}}$ be the said overview volume reconstructed from the measured low resolution data p_{L}. Let $\mathsf{X}_{\mathrm{H}}=\mathsf{X}_{\mathrm{M}}+\mathsf{X}_{\mathrm{U}}$ be a decomposition of the high resolution x–ray transform into rays that have beem measured and those that are unmeasured. The high resolution projection data are denoted as p_{H} which we assume to be zero for all unmeasured rays, i.e. in the truncated region and in the regions where the projection angles are missing.

Using this notation the data completion method is given by the equation

$$f_{\mathrm{H}}=\mathsf{X}_{\mathrm{H}}^{-1}(p_{\mathrm{H}}+\mathsf{X}_{\mathrm{U}}f_{\mathrm{L}}). \tag{2}$$

B. Data Filtering Method

Equation (2) is equivalent to

$$f_{\mathrm{H}} = f_{\mathrm{L}} + \mathsf{X}_{\mathrm{H}}^{-1}(p_{\mathrm{H}} - \mathsf{X}_{\mathrm{M}} f_{\mathrm{L}})$$

because we have $f_{\mathrm{L}} = \mathsf{X}_{\mathrm{H}}^{-1}(\mathsf{X}_{\mathrm{U}} f_{\mathrm{L}} + \mathsf{X}_{\mathrm{M}} f_{\mathrm{L}})$ due to the linearity of the x–ray transform. The latter equation, however, has a different interpretation. It implies that the rawdata obtained by forward projection of the overview scan are subtracted from the measured high resolution rawdata in the measured region(s), and that those rawdata are reconstructed and added to the low resolution overview images.

While this sounds interesting a closer look at the subtraction term

$$p_{\mathrm{H}} - \mathsf{X}_{\mathrm{M}} f_{\mathrm{L}}$$

reveals that it is the difference between high resolution measured rawdata and low resolution virtual rawdata and thereby must be equivalent to a high–pass filtering of the measured data. If h denotes an appropriate high–pass filter then we may summarize our findings as

$$p_{\mathrm{H}} - \mathsf{X}_{\mathrm{M}} f_{\mathrm{L}} = h * p_{\mathrm{H}}.$$

Hence, we can obtain the high resolution reconstruction as

$$f_{\mathrm{H}} = f_{\mathrm{L}} + \mathsf{X}_{\mathrm{H}}^{-1}(h * p_{\mathrm{H}}) \tag{3}$$

which does not involve any forward projections.

To find the appropriate high–pass filter let us go to Fourier domain and let $\mathrm{MTF}_{\mathrm{L}}(u)$ and $\mathrm{MTF}_{\mathrm{H}}(u)$ be the presampling MTFs of the low resolution overview scan and of the high resolution ROI scan, respectively, both being scaled to the isocenter. Then, the high–pass filter $H(u)$ should have the property

$$\mathrm{MTF}_{\mathrm{L}}(u) + \mathrm{MTF}_{\mathrm{H}}(u) H(u) = \mathrm{MTF}_{\mathrm{H}}(u).$$

Solving for $H(u)$ yields

$$H(u) = 1 - \frac{\mathrm{MTF}_{\mathrm{L}}(u)}{\mathrm{MTF}_{\mathrm{H}}(u)}.$$

Note that the simplicity of the data filtering method is appealing. Although the method is analytically equivalent to the completion method this is not true numerically. The downside of the data filtering method is increased image noise and increased susceptibility to artifacts in the high resolution region since the method just adds a correction term to the low resolution data and thereby cannot eliminate image noise or artifacts inherent in f_{L}.

C. Data Weighting Method

Our third method has two advantages: the advantage of the data completion method of not having artifacts propagating into the high resolution ROI and the advantage of the high–pass filtering method of being computationally highly efficient. This is achieved by designing appropriate weighting functions $w_{\mathrm{L}}(\vartheta, \xi)$ and $w_{\mathrm{H}}(\vartheta, \xi)$ that shall be multiplied to the low and high resolution rawdata, respectively.

To start, let us define a redundancy weight $w_{\mathrm{R}}(\vartheta, \xi)$. Whenever the scan range exceeds 180° plus fan angle we need to either weight the data with a short scan weight, which is also known as the Parker weight function, or if the scan exceeds 360° we need to weight the data with the overscan weight function. In case of a shifted detector design the corresponding shifted detector weight can be applied. These redundancy weight functions can be taken from reference [23], for example. Since the focal spot trajectory of the low resolution and of the high resolution scan differ, in general, we need to use two different redundancy weight functions $w_{\mathrm{RL}}(\vartheta, \xi)$ and $w_{\mathrm{RH}}(\vartheta, \xi)$, respectively. The redundancy weights fulfill equation (1) as shown in reference [23].

In addition a weight function $w_{\mathrm{M}}(\vartheta, \xi)$ is needed whose support corresponds to the measured high resolution data. This means that w_{M} is zero wherever the high resolution scan has missing rays and it smoothly increases to one where measured high resolution data are available. Hence w_{M} masks out the truncated regions of the detector and the unmeasured angular positions.

We now set

$$w_{\mathrm{H}}(\vartheta, \xi) = w_{\mathrm{RH}}(\vartheta, \xi) w_{\mathrm{M}}(\vartheta, \xi)$$
$$w_{\mathrm{L}}(\vartheta, \xi) = w_{\mathrm{RL}}(\vartheta, \xi)\big(1 - \sum_k w_{\mathrm{H}}(\vartheta + k\pi, (-)^k \xi)\big).$$

Then

$$f_{\mathrm{H}} = \mathsf{X}_{\mathrm{L}}^{-1} p_{\mathrm{L}} w_{\mathrm{L}} + \mathsf{X}_{\mathrm{H}}^{-1} p_{\mathrm{H}} w_{\mathrm{H}} \tag{4}$$

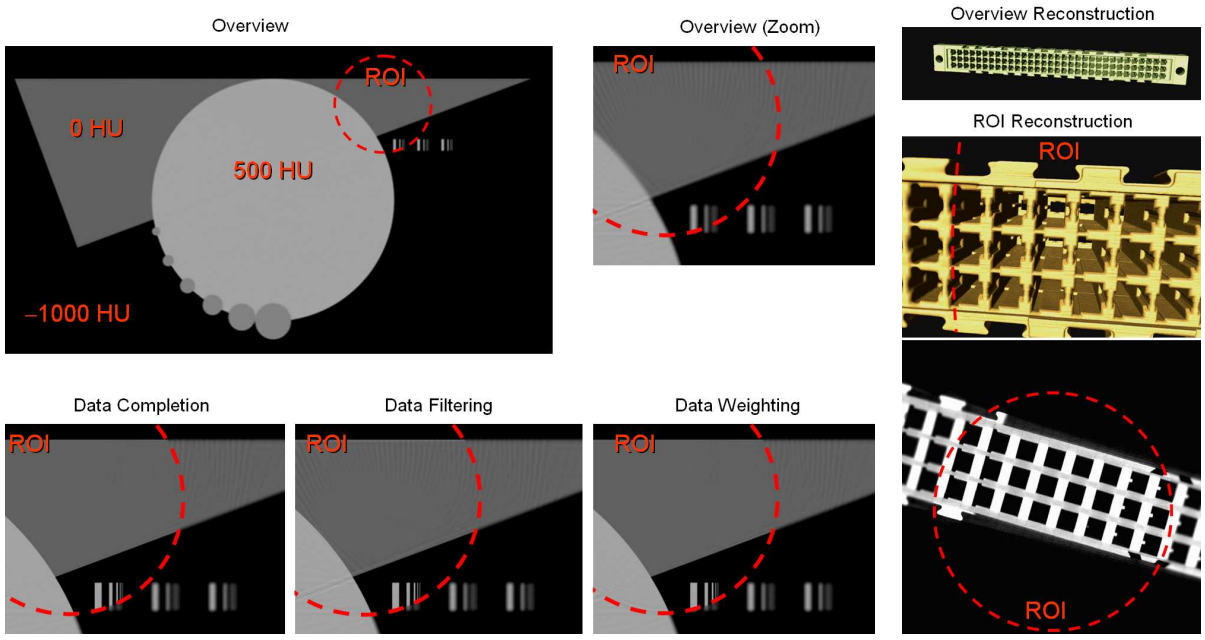
yields the final volume in the high resolution ROI.

To see that everything is properly normalized let us temporarily abbreviate $w(\vartheta + k\pi, (-)^k \xi)$ as $w(k)$. Keeping in mind that w_{RL} is properly normalized, because $\sum_k w_{\mathrm{RL}}(k) = 1$ by definition, we are now ready to check whether $w_{\mathrm{L}} + w_{\mathrm{H}}$ is properly normalized. We find

$$\begin{aligned}
&\sum_k \big(w_{\mathrm{L}}(k) + w_{\mathrm{H}}(k)\big) = \\
&\sum_k \big(w_{\mathrm{RL}}(k)\big(1 - \sum_l w_{\mathrm{H}}(k+l)\big) + w_{\mathrm{H}}(k)\big) = \\
&\sum_k w_{\mathrm{RL}}(k)\big(1 - \sum_l w_{\mathrm{H}}(k+l)\big) + \sum_k w_{\mathrm{H}}(k) = \\
&\sum_k w_{\mathrm{RL}}(k)\big(1 - \sum_l w_{\mathrm{H}}(l)\big) + \sum_l w_{\mathrm{H}}(l) = \\
&\sum_k w_{\mathrm{RL}}(k) + \sum_l w_{\mathrm{H}}(l)\big(1 - \sum_k w_{\mathrm{RL}}(k)\big) = 1.
\end{aligned}$$

To conclude this section, let us give an example of how to define the weight function $w_{\mathrm{M}}(\vartheta, \xi)$ that is smooth and masks out the unmeasured regions for the simple case of scanning a cylindrical high resolution ROI of radius R_{MH} and center $\boldsymbol{o}$. Note that a ray through the center of the ROI has the lateral coordinate $\xi_{\mathrm{C}} = \boldsymbol{o} \cdot \boldsymbol{\vartheta}$ which is a function of the angle ϑ, in general.

Let $s(x)$ have the properties $s(-x) = -s(x)$ and $s(1) = 1$. Given that we apply the data weighting method to sampled data it is also recommended to choose s to be smooth. We use $s(x) = \sin(\frac{1}{2}\pi x)$ for our numerical experiments. Now we can define the mask function which in this case is a simple



(a) Simulation (b) Measurement

Fig. 2. a) The simulated test phantom is shown in a low resolution overview reconstruction, a low resolution zoom version where the FOV corresponds to the FOV used for the high resolution reconstructions. The high resolution reconstructions (bottom row) indicate that all three methods, the data completion, the high–pass filtering and the data weighting method achieve equivalent results. b) Reconstructions of a measured connector element showing a low resolution overview volume rendering, and a high resolution ROI reconstruction using the data weighting method. The high resolution volume is shown as a volume rendering and as a CT slice. Within the ROI the desired high spatial resolution is achieved.

radial transition weight function

$$w_{\mathrm{M}}(\vartheta,\xi)=\frac{1}{2}\begin{cases}0 & \text{if } \xi<\xi_A\\ 1+s\big(2\frac{\xi-\xi_A}{\xi_B-\xi_A}-1\big) & \text{else if } \xi<\xi_B\\ 2 & \text{else if } \xi<\xi_D\\ 1-s\big(2\frac{\xi-\xi_D}{\xi_E-\xi_D}-1\big) & \text{else if } \xi<\xi_E\\ 0 & \text{else}\end{cases}$$

that is zero in the outer region and smoothly increases to one in the inner region of the high resolution scan. The parameters $\xi_{\mathrm{A}}<\xi_{\mathrm{B}}<\xi_{\mathrm{C}}<\xi_{\mathrm{D}}<\xi_{\mathrm{E}}$ are functions of ϑ and specify the lateral detector limits of the high resolution scan, and the transition regions where the weight smoothly increases from zero to one, and the center of the detector, as discussed above. With $\xi_{\mathrm{C}}=\boldsymbol{o}\cdot\boldsymbol{\vartheta}$ we use $\xi_{\mathrm{A}}=\xi_{\mathrm{C}}-R_{\mathrm{MH}}$, $\xi_{\mathrm{B}}=\xi_{\mathrm{C}}-R_{\mathrm{MH}}+\Delta R$, $\xi_{\mathrm{D}}=\xi_{\mathrm{C}}+R_{\mathrm{MH}}-\Delta R$, and $\xi_{\mathrm{E}}=\xi_{\mathrm{C}}+R_{\mathrm{MH}}$ with ΔR being the size of the transition zone, $0<\Delta R\ll R_{\mathrm{MH}}$.

III. Experiments

To evaluate the three methods defined in equations (2), (3) and (4) we conducted several simulations and measurements. The simulations were carried out by the analytical projection simulator RayConStruct PS (RayConStruct GmbH, Nürnberg, Germany). The measurements were done using the Tomo-Scope HV 500 cone–beam CT scanner (Werth Messtechnik GmbH, Gießen, Germany). The tube voltage was 200 kV.

The geometry is a flat detector cone–beam geometry with about 1000 projections per full rotation and a detector with about 1000 by 1000 detector pixels of size 0.4 mm. The distance of the focal spot to the detector surface was chosen as 2400 mm and the radii of the circle scans were set to $R_{\mathrm{FL}}=1200$ mm and $R_{\mathrm{FH}}=150$ mm for the low resolution overview and the high resolution ROI scan, respectively. This results in a radius of the field of measurement of $R_{\mathrm{ML}}=100$ mm for the overview scan and of $R_{\mathrm{MH}}=12.5$ mm for the ROI scan. Thus, the ROI scan zooms into the object by a factor of eight.

IV. Results

Figure 2a) shows various reconstructions of the simulated test phantom. The overview image depicts all objects within the low resolution field of measurement. A dashed circle indicates the region of interest. There is also a zoom version of the overview image which was produced by reconstructing the overview data p_{L} on the same voxel grid as will be used for the reconstructions of the ROI data. The three images at the bottom row of subfigure a) are the ROI reconstructions that use the three methods data completion, data filtering and data weighting to combine p_{L} with p_{H}.

The measured connector element is presented in figure 2b). The overview scan p_{L} can be used to reconstruct the complete object (here, a volume rendering of the connector is shown). The ROI reconstructions (one volume rendering and one transversal slice) instead only show a portion of the connector with high spatial resolution within the cylindrical ROI and low spatial resolution outside the ROI. Due to space limitations the measurement is only shown using the data weighting method.

Figure 3 is the scan of a chromatography column. Only a square section centered around the ROI is shown in four

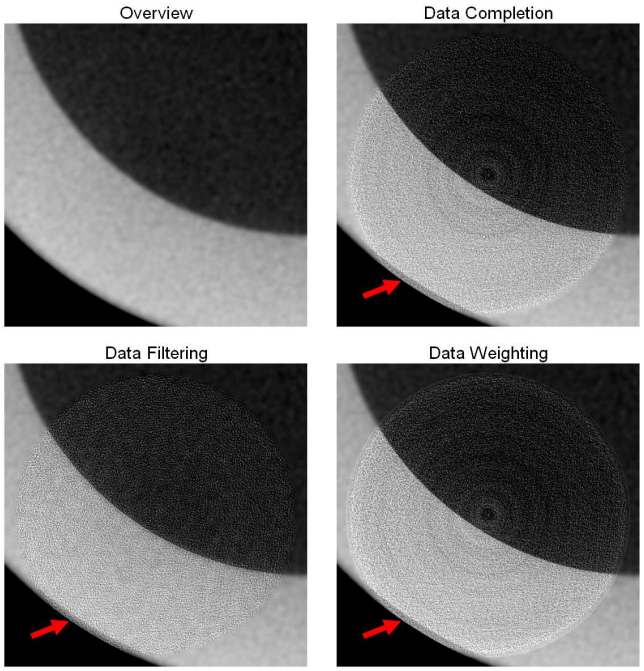


Fig. 3. Chromatography column filled with silica gel. The figures show the zoomed FOV of the ROI. The 50 mm inner diameter is surrounded by a 9 mm thick glas cylinder which is surrounded by a 0.5 mm splinter shield (scatter protection) foil. Only the ROI reconstructions show the scatter protection foil (arrows). The ROI diameter is 25 mm.

versions: overview image, data completion method, data filtering method, and data weighting method. The properties of the filtering method are as expected because noise is increased. Rather unexpected but quite evident is the fact that the only ROI method that does not show ring artifacts is the data filtering method while the completion and the weighting methods show some ring artifacts.

V. Discussion

Whenever an overview scan is available it is relatively simple to perform local tomography. To improve the computational performance and to reduce the memory requirements we proposed two methods that do not need to complete the truncated data of the high resolution scan and that do not need to perform forward projections of an overview volume. In fact these two methods do not even need to reconstruct the overview volume. While our preliminary results shown here give the impression that the data filtering method is of equal image quality as the data weighting method a further analysis using simulations and measurements shows that this is not always the case. Noise and artifacts that propagate from the low resolution reconstruction into the ROI cannot be removed by the data filtering method. The data weighting method, however, does not suffer from those artifacts because only very low frequencies propagate from the low resolution data into the high resolution scan.

Summarizing, we evaluated three highly promising approaches that are readily applicable for industrial tomography.

Acknowledgments

This work was supported in parts by the AiF under grant KF2336201FO9. The scans were carried out by Werth Messtechnik GmbH, Gießen, Germany. The chromatography column was provided by the Lehrstuhl für Thermische Verfahrenstechnik, University of Erlangen–Nürnberg, Germany. Parts of the reconstruction software were provided by RayConStruct GmbH, Nürnberg, Germany. We thank the Intel Corporation and Fujitsu Siemens Computers GmbH for providing the highly performant compute hardware.

References

[1] C. Penßel, W. A. Kalender, and M. Kachelrieß, "ROI–driven CT trajectories," *IEEE Medical Imaging Conference Record*, pp. M06–181, 1969–1972, 2006.

[2] D. Gentle and N. Spyrou, "Region of interest tomography in industrial applications," *Nuclear Instruments and Methods in Physics Research*, vol. A299, pp. 534–537, 1990.

[3] S. Azevedo, P. Rizo, and P. Grangeat, "Region-of-interest cone-beam computed tomography," Lawrence Livermore National Lab., CA (United States), Tech. Rep., June 1995.

[4] I. K. Chun, M. H. Cho, S. C. Lee, M. H. Cho, and S. Y. Lee, "X–ray micro–tomography for small–animal imaging with zoom–in imaging capability," *Phys. Med. Biol.*, vol. 49, pp. 3889–3902, 2004.

[5] J. Wiegert, M. Bertram, T. Netsch, J. Wulff, J. Weese, and G. Rose, "Projection extension for region of interest imaging in cone–beam CT," *Academic Radiology*, vol. 12, pp. 1010–1023, 2005.

[6] S. Y. Lee, M. H. Cho, S. C. Lee, I. K. Chun, and J. J. Park, "Small animal x–ray micro–CT with zoom–in imaging capability," *IEEE Medical Imaging Conference Record*, pp. M14–156, 3102–3105, 2006.

[7] L. Chen, Y. Shen, C.-J. Lai, T. Han, Y. Zhong, S. Ge, X. Liu, T. Wang, W. T. Yang, G. J. Whitman, and C. C. Shaw, "Dual resolution cone beam breast CT: A feasibility study," *Med. Phys.*, vol. 36, no. 9, pp. 4007–4914, Sep. 2009.

[8] R. Lewitt, "Processing of incomplete measurement data in computed tomography," *Med. Phys.*, vol. 6, no. 5, pp. 412–417, 1979.

[9] D. Kadrmas, R. Jaszczak, J. McCormick, R. Colemen, and C. Lim, "Truncation artifact reduction in transmission CT for improved SPECT attenuation compensation," *Phys. Med. Biol.*, vol. 40, pp. 1085–1104, 1995.

[10] B. Ohnesorge, T. Flohr, K. Schwarz, J. Heiken, and K. Bae, "Efficient correction for CT image artifacts caused by objects extending outside the scan field of view," *Med. Phys.*, vol. 27, no. 1, pp. 39–46, Jan. 2000.

[11] J. Hsieh, E. Chao, J. Thibault, B. Grekowicz, A. Horst, S. McOlash, and T. Myers, "Algorithm to extend reconstruction field–of–view," *IEEE International Symposium on Biomedical Imaging*, pp. 1404–1407, Apr. 2004.

[12] C. Penßel, M. Kachelrieß, and W. A. Kalender, "Hybrid detruncation HDT algorithm for the reconstruction of CT data," *Radiology*, vol. 233(P), p. 293, Nov. 2004.

[13] K. Sourbelle, M. Kachelrieß, and W. A. Kalender, "Reconstruction from truncated projections in CT using adaptive detruncation," *European Radiology*, vol. 15, no. 5, pp. 1008–1014, May 2005.

[14] X. Pan, Y. Zou, and D. Xia, "Peripheral and central ROI–image reconstruction from and data–redundancy exploitation in truncated fan–beam data," *Med. Phys.*, vol. 32, no. 3, pp. 673–684, Mar. 2005.

[15] M. Courdurier, F. Noo, M. Defrise, and H. Kudo, "Solving the interior problem of computed tomography using a priori knowledge," *Inverse Problems*, vol. 24, p. 065001, 2008.

[16] H. Kudo, M. Coururier, F. Noo, and M. Defrise, "Tiny a priori knowledge solves the interior problem in computed tomography," *Phys. Med. Biol.*, vol. 53, pp. 2207–2231, 2008.

[17] X. Pan, E. Sidky, and M. Vannier, "Why do commercial CT scanners still employ traditional, filtered back–projection for image reconstruction?" *Inverse Problems*, vol. 25, no. 12, pp. 1–36, Dec. 2009.

[18] H. Yu and G. Wang, "Compressed sensing based interior tomography," *Phys. Med. Biol.*, vol. 54, pp. 2791–2805, 2009.

[19] H. Yu, J. Yang, M. Jiang, and G. Wang, "Supplemental analysis on compressed sensing based interior tomography," *Phys. Med. Biol.*, vol. 54, pp. N425–N432, 2009.

[20] P. S. Cho, A. D. Rudd, and R. H. Johnson, “Cone–beam CT from width–truncated projections,” *Computerized Medical Imaging and Graphics*, vol. 20, no. 1, pp. 49–57, 1996.
[21] G. Wang, “X–ray micro–CT with a displaced detector array,” *Med. Phys.*, vol. 29, no. 7, pp. 1634–1636, Jul. 2002.
[22] V. Liu, N. R. Lariviere, and G. Wang, “X–ray micro–CT with a displaced detector array: Application to helical cone–beam reconstruction,” *Med. Phys.*, vol. 30, no. 10, pp. 2758–2761, Oct. 2003.
[23] C. Maaß, M. Knaup, R. Lapp, M. Karolczak, W. A. Kalender, and M. Kachelrieß, “A new weighting function to achieve high temporal resolution in circular cone–beam CT with shifted detectors,” *Med. Phys.*, vol. 35, no. 12, pp. 5898–5909, Dec. 2008.

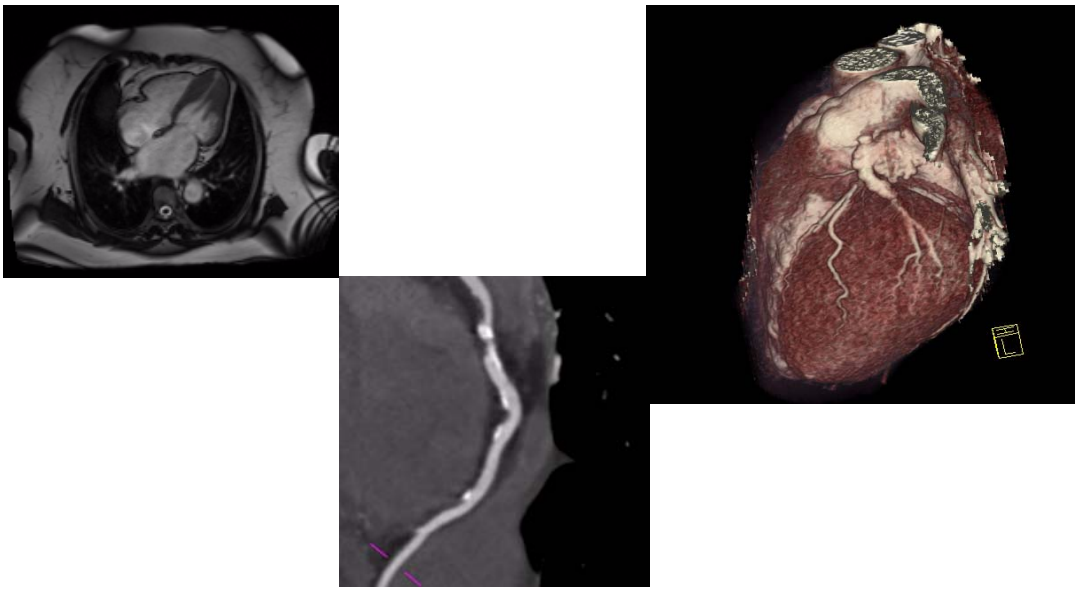
Plenary Talk

Wednesday, June 9, 8:00am - 9:00am

Redefining Cardiac Imaging: The Stunning Rise of Cardiac Computed Tomography

Sheldon E. Litwin, M.D.
Amundsen Professor of Cardiology
Director of Cardiovascular Imaging
University of Utah Hospitals and Clinics

Noninvasive visualization of the coronary arteries has been a highly sought after goal for several decades. Barriers to reaching this goal include the small caliber of the coronary vessels (< 4 mm) and motion of the heart and thorax. These issues have been largely overcome by recent advances in computed tomography (CT), including: electrocardiographic gating, helical acquisition, multi row detectors and new reconstruction algorithms. Today, a complete cardiac study can be performed in < 15 seconds. Stunning images of the heart and great vessels are obtained almost instantly. Both calcified and noncalcified atherosclerotic plaque can be detected in the coronary arteries. The accuracy for diagnosing obstructive coronary artery disease with CT is substantially higher than that for conventional approaches using stress testing. Ongoing studies are addressing the safety, economics, prognostic value and ability to affect outcomes in a variety of patient populations. I believe that in the near future, CT will become the first line imaging technique for diagnosing many common cardiac conditions. The future for this rapidly evolving modality is extraordinarily bright.



Cone beam reconstruction for the distorted circle and line trajectory

Souleymane Konate

Department of Radiology
University of Massachusetts Medical School
Worcester, MA 01655

Alexander Katsevich

Department of Mathematics
University of Central Florida
Orlando, FL 32816

Abstract

We propose an exact filtered backprojection algorithm for inversion of the cone beam data in the case when the trajectory is composed of a distorted circle and a line segment. The length of the line scan is determined by the region of interest, and it is independent of the size of the object. Assuming that the distorted circle is planar and convex, we show that we have an exact reconstruction. Numerical experiments demonstrate good image quality.

1 Introduction

Image reconstruction from projections is important from the theoretical (inverse problems) and applied (computed tomography (CT)) standpoints. Inversion of the cone beam transform has always generated a great deal of interest among scientists. For instance, early general inversion formulae were proposed in Kirillov (1961), Tuy (1983), Smith (1985), Gelfand and Goncharov (1987), Grangeat (1991). Over the years, different types of theoretically exact cone beam CT algorithms have been proposed. Among them the Filtered Backprojection (FBP) algorithms are generally considered the fastest. On the other hand, FBP algorithms have to be developed for specific source trajectories on a case by case basis. For instance, they are available for a wide range of source trajectories, which include circle-and-line, circle-and-arc, constant pitch helix, just to mention a few. A problem of interest is the circle-and-line source trajectory. This trajectory can be obtained by first moving the patient along a line and then by rotating the C-arm. In practice, due to the heavy weight of the C-arm, the presumed circular trajectory of the C-arm is frequently perturbed. If not corrected, the distortions on the source trajectory lead to noticeable artifacts in the reconstructed image. An FBP-type algorithm was developed for the ideal circle and line source trajectory in [2]. In this paper the number of intersection points between Radon planes and the source trajectory was established to be at most three. The main goal of this research is to utilize the algorithm of [2] as a building block for constructing a new inversion algorithm for a broader class of curves. The problem is solved in the following manner. The first step is to consider an entire class of curves on which a set of natural geometric restrictions is imposed. The curves are assumed to be planar, smooth, non self-intersecting with positive curvature. Also, the curves have to satisfy an extra condition referred to as 'convexity with respect to the origin'. A given curve satisfies 'convexity with respect to the origin' if the number of intersection points between the curve and a line passing through its initial point is at most 2. The next step is to study the geometry of the intersection points (IPs) between planes and the source trajectory. The distribution of the IPs over different sections of the source trajectory is also taken into consideration in the analysis. The plan is to limit the number of IPs between planes and the source trajectory to maximum three as prescribed by the ideal circle-and-line algorithm. Third, a set of lemmas is derived proving that the inversion formula of [2] for the ideal circle-and-line can be applied to our class of curves. Finally, the algorithm is implemented in order to demonstrate good image quality. In summary, the derived results apply not only to the ideal circle, but they are also applicable to the entire class of curves defined earlier. In other words, any curve or source trajectory satisfying the prescribed conditions admits an FBP-type reconstruction. This paper is organized in the following manner. In section 2, we give some general definitions and an overview of the general inversion formula described in [1]. In section 3, we present the set of lemmas that make the inversion formula of [2] applicable to our class of curves. Finally, some numerical results are presented in section 4.

2 General Definitions

Definition 1 *Let Γ be a finite union of smooth curves in $\mathbb{R}^3$:*

$$I := \bigcup [a_l, b_l] \to \mathbb{R}^3, I \in s \to y(s) \in \mathbb{R}^3, |\dot{y}(s)| \neq 0 \text{ on } I \tag{1}$$

where

$$-\infty < a_l < b_l < \infty, \dot{y}(s) := dy/ds$$

$$\beta(s,x) := \frac{x-y(s)}{|x-y(s)|}, x \in \mathbb{R}^3 \setminus \Gamma, s \in I;$$

$$\Pi(x,\xi) := \{z \in \mathbb{R}^3 : (z-x)\cdot\xi = 0\}.$$

$g(y,\beta)$ is the cone beam transform of f, and $\beta(s,x)$ is the unit vector directed from the source towards the reconstruction point x.

In what follows, f is assumed to be smooth and compactly supported. Additionally, f is identically zero in a neighborhood of the trajectory.

Definition 2 *Let $x \in \mathbb{R}^3$ and $\xi \in \mathbb{R}^3 - \{0\}$. The intersection points of $\Pi(x,\xi)$ with Γ are denoted $y(s_j)$ where $s_j = s_j(\xi, \xi\cdot x), j = 1, 2, ...$*
For $\beta \in S^2$, $\beta^\perp$ denotes the circle $\{\ \alpha \in S^2 : \alpha\cdot\beta = 0\ \}$ consisting of unit vectors perpendicular to β. Let $Crit(s,x)$ be the set of all directions α is in $\beta^\perp(s,x)$ such that the plane $\Pi(x,\alpha)$ is tangent to Γ or contains an endpoint of Γ. Denote by I_{reg} the set of all parameters s in I, for which the set $Crit(s,x)$ is included (but not equal) in $\beta^\perp(s,x)$. Finally, define $Crit(x)$ to be the union over all s in I of all $Crit(s,x)$. We can concisely reformulate the definitions as:

$$\begin{aligned} Crit(s,x) &:= \Big\{\alpha \in \beta^\perp(s,x) : \Pi(x,\alpha) \text{ is tangent to } \Gamma \\ &\qquad \text{or } \Pi(x,\alpha) \text{ contains an endpoint of } \Gamma\Big\}, \\ I_{reg} &:= \Big\{s \in I : Crit(s,x) \subsetneq \beta^\perp(s,x)\Big\}, \\ Crit(x) &:= \bigcup_{s\in I} Crit(s,x). \end{aligned} \tag{2}$$

Conditions on the trajectory For any given x in $\mathbb{R}^3$, where the function f needs to be computed, the trajectory Γ must satisfy the following main assumptions:

Property 1 (Tuy's Completeness Condition.) *Any plane through x intersects Γ at least at one point.*

Property 2 *For any $s \in I_{reg}(x)$, the number of directions in $Crit(s,x)$ is finite.*

Property 3 *For any $\alpha \in S^2 \setminus Crit(x)$, the number of points in $\Pi(x,\alpha)\cap\Gamma$ is finite.*

Additionally, consider a weight function $n(s,x,\alpha), s \in I_{reg}(x), \alpha \in \beta^\perp(s,x) \setminus Crit(s,x)$. The main assumptions on the function n are the following:

Property 4 *For almost all $\alpha \in S^2$,*

$$\sum_{j:y(s_j)\in C\cap\Pi(x,\alpha)} n(s_j,x,\alpha) = 1. \tag{3}$$

Property 5 *$n(s,x,\alpha)$ is a piecewise constant function.*

3 Inversion Formula

In this section, the main results are stated without proof.

Definition 3 *Let $y(s), s \in [0, s_{max}]$ be a planar curve. The curve C is said to be convex with respect to $y(0)$ its initial point, if any line passing through $y(0)$ intersects C at most twice.*

Definition 4 *$\mathscr{C}$ is defined as the class of non self-intersecting smooth planar curves with positive curvature satisfying the convexity with respect to $y(0)$.*

Theorem 1 *Let $C \in \mathscr{C}$, $y(s) \in C, s \in [0, s_{max}]$. Then, there exists a well defined smooth function R such that $y(s) = (R(s)\cos(s), R(s)\sin(s), 0)$.*

The theorem states that any curves $C \in \mathscr{C}$ can be parametrized as described above. Note the choice of $R(s)$ is not unique. In our case, we choose the function R in the following manner such that we have the lemma below.

Lemma 1 *$R(0) - R(s)\cos(s) > 0$ for all s in $[0, s_{max}]$.*

Definition 5 *Let $C \in \mathscr{C}$, and $y(s) \in C$. We define $\Omega = \{x | x = \lambda y(0) + (1-\lambda)y(s), s \in [0, s_{max}]\}$ to be the set of points located on all chords from the origin $y(0)$ to $y(s)$, a point on the curve, when the parameter $s \in [0, s_{max}]$.*

Definition 6 *A region of interest (ROI) noted U is admissible if its projection $\hat{U}$ on the x-y plane is contained in Ω.*

Description of the trajectory The source trajectory consists of a distorted circle C and a line L attached to C at some point $y(0)$. The curve C is relatively close to a complete circle and the line L is sufficiently long. Let the following be respective parametrizations of the line and circle, $s \in I_1 : s \to y(s) \in L$ and $s \in I_2 : s \to y(s) \in C$. Consider the following two parametric intervals. The first one is $I_1(x) \subset I_1$ which corresponds to the section of L between $y(0)$ and $y_L(x)$. The second interval $I_2(x)$ corresponds to the section of the circle between $y(0)$ and $y_C(x)$. $\Gamma_{1\pi}(x)$ denotes the section of $C \bigcup L$ bounded by $L_{1\pi}(x)$. $\Gamma_{1\pi}(x)$ satisfies Properties 1, 2 similarly to the ideal circle and line case. We would like to prove that $\Gamma_{1\pi}(x)$ also satisfies Property 3, and the weight function $n(s,x,\alpha)$ satisfies Properties 4 and 5. The idea is to demonstrate that the number of intersection points between $\Gamma_{1\pi}(x)$ and a plane $\Pi(x)$ passing through x in the ROI is at most three. If this is the case, then the weight distribution will be analogous to the one in the ideal circle and line problem.

The results follow if we show that the trajectory is not too exotic.

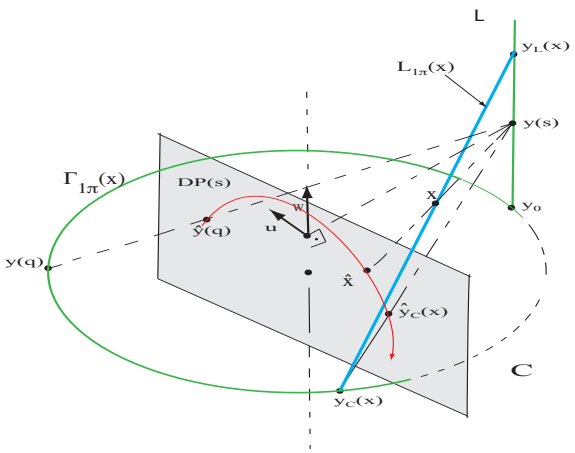


Figure 1: Description of the source trajectory

Theorem 2 *The projection of the curve C on the detector plane $DP(s)$ is a convex curve.*

PROOF: Let $y(s)$ be a source position on the curve C, and let $(u(s), w(s))$ be the projection of $y(s)$ on $DP(s_0)$, $s_0 \in I_1$. The proof follows if we show that $\dot{u}(s) < 0$ and $\frac{d^2w}{du^2} < 0$. Below is a sketch of the proof. First, we show that $\dot{u}(s) \neq 0$. In order to prove that $\dot{u}(s) < 0$ for all s, it suffices to show that $\dot{u}(s) < 0$ as $s \to 0$. Second, with some computation we can derive a formula (see below) for $\frac{d^2w}{du^2}$, and establish that $\frac{d^2w}{du^2} < 0$

$$\frac{d^2w}{du^2} = \frac{C(s)\kappa(s)\Big(R(0) - R(s)\cos(s)\Big)}{\dot{u}(s)}, \qquad (4)$$

where $\dot{u}(s) < 0, C(s) > 0$, the curvature $\kappa(s) > 0$, $R(0) - R(s)\cos(s) > 0$, by Lemma 1. ■

Definition 7 *Let $C \in \mathscr{C}$, and x be a reconstruction point in the ROI. Consider the plane $\Pi(x)$ passing through x and the line L. The plane $\Pi(x)$ intersects the distorted circle at 2 points, $y(0)$ and y_C. Then the PI–line denoted by $L_{1\pi}(x)$ (Figure 1) is defined to be the line segment containing the point x connecting $y_C(x)$ to the line L at some point $y_L(x)$.*

Theorem 3 *For every x in the ROI, $L_{1\pi}(x)$ is unique.*

PROOF: Let x be a point in the ROI. Because $C \in \mathscr{C}$, we can always find a plane $\Pi(x)$ passing through x and containing the line L. According to Definition 7, $\Pi(x)$ determines the $L_{1\pi}(x)$. To prove the uniqueness, let $L_{1\pi}(x)$ and $L'_{1\pi}(x)$be two different Pi-lines. Let P_L be a vertical plane containing both the line L and the reconstruction point x. Then P_L intersects the circle C at y_C and y'_C. Additionally, the corner of the circle and line $y(0)$ belongs to P_L. Now if we consider P_C the plane of the circle, then the points $y(0)$ and y_C belong to P_C. Also recall that the plane P_L contains the points $y(0)$, and y_C and y'_C, so $y(0), y_C, y'_C$ are contained in the intersection $P_C \cap P_L$. In other words, $y(0), y_C, y'_C$ are not only on the curve C, but they are also colinear. This statement violates the fact that the curve C satisfies the convexity with respect to $y(0)$. Therefore $L_{1\pi}(x)$ must be unique. ■

Theorem 4 *Let $C \in \mathscr{C}$, and x be a point in the ROI. Then the following hold true:*

1. *There is no plane tangent to C in the interior of $\Gamma_{1\pi}(x)$ and passing through $L_{1\pi}(x)$.*
2. *There is no plane containing x, tangent to C in the interior of $\Gamma_{1\pi}(x)$, and passing through the corner $y(0)$ of the distorted circle and line.*

Conclusion: The conditions that we just derived state that the trajectory is not too exotic. In other words, we have proven that any plane passing through a reconstruction point x in the ROI cannot intersect the curve C at more than 3 IPs. The distortions on the circle do not violate the number of intersection points (3 IPs at most) referred to in the ideal circle and line case [2]. Consequently, the weight distribution remains the same, and a description of the weight function $n(s, x, \alpha)$ is summarized in Table 1 below.

Table 1: Definition of the Weight Function $n(s, x, \alpha)$

Case	Weight Function $n(s, x, \alpha)$
1IP $s_1 \in I_1(x)$	$n(s_1, x, \alpha) = 1$
1IP $s_1 \in I_2(x)$	$n(s_1, x, \alpha) = 1$
3 IPs $s_1 \in I_1(x)$	$n(s_1, x, \alpha) = -1$
$s_2, s_3 \in I_2(x)$	$n(s_k, x, \alpha) = 1, k = 1, 2$

We have shown that our curve satisfies Properties $1 - 5$, so we can use Katsevich's inversion formula for the ideal circle and line case. The inversion formula is stated below.

Theorem 5 *Let $C \in \mathscr{C}$.For $f \in C_0^\infty(U)$,*

$$f(x) = -\frac{1}{2\pi^2} \int_{I_k(x)} \sum_{k=1}^{2} \frac{\delta_k(s, x)}{|x - y(s)|} \qquad (5)$$

$$\int_0^{2\pi} \frac{\partial}{\partial q} g(y(q), \Theta_k(s, x, \gamma))|_{q=s} \frac{d\gamma}{\sin\gamma} ds, \qquad (6)$$

where

$$\Theta(s,x,\gamma) = \cos\gamma\beta(s,x) + \sin\gamma e_k(s,x), \quad (7)$$
$$e_k(s,x) := \beta(s,x) \times u_k(s,x), \quad (8)$$

and δ_k is defined as follows:

$$\delta_1(s,x) = -\mathrm{sgn}(u_1(s,x)\cdot\dot{y}(s)),\ s \in I_1(x); \quad (9)$$
$$\delta_2(s,x) = 1,\ s \in I_2(x). \quad (10)$$

4 Numerical Experiments

The steps of the numerical implementation of the algorithm on a flat detector geometry are described in [8]. We conducted some numerical experiments with the clock phantom which was originally described in[6]. The background cylinder was at 0 HU, the spheres were at 1000 HU, and the air at -1000 HU. Under the assumption that the plane of the circle is at $z = 0$, the phantom is shifted by $\Delta z = +20.0$. The purpose of this shift is to better illustrate how well the algorithm reconstructs cross-sections which are located away from the plane of the circle. The size of the image is 512 x 512, which corresponds to the following ROI $|x| \leq 250$ and $|y| \leq 250$. In order to make small artifacts visible, we introduce a highly compressed grey level window level and window width of $[1.0, 0.1d]$, for a density $d = 1$. Besides small numerical discretization artifacts, the quality of the reconstructed cross section is good. The values of the parameter used are in mm. The radius of rotation is 570, the height of L is 160, the detector pixel size is 0.7, number of detector rows (551), number of columns (1001), number of source positions on C (600), and L (160), number of filtering lines for source points (200). The variable radius is chosen to be $R(s) = R(0) - \frac{1}{2}\epsilon s^2$, where the distortion parameter $\epsilon = 5.0$.

Conclusion We presented an exact filtered backprojection cone beam inversion formula when the source trajectory is composed of a distorted circle and a line. With some mild geometric conditions on the curve, we are able to show that Katsevich's algorithm for the perfect circle and line trajectory can be extended to our class of curves.

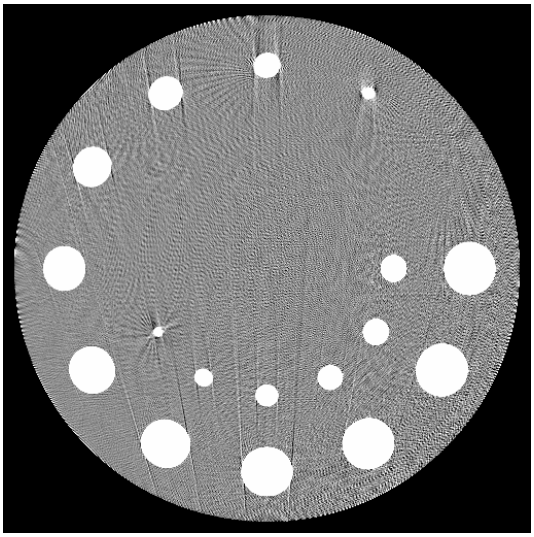


Figure 2: Cross section $z = 20$ mm through the reconstructed clock phantom. The region $|x| \leq 250$,$|y| \leq 250$ mm is show for a distortion parameter of $\epsilon = 5.0$

References

[1] A. Katsevich, "A general scheme for reconstructing inversion algorithms for cone beam CT," *International Journal of Mathematics and Mathematical Sciences*, Vol. 2003, pp. 1305-1321, 2003.

[2] A. Katsevich, "Image reconstruction for the circle and line trajectory," *Institute of Physics Publishing*, Vol. 49, pp. 5059–5072, 2004.

[3] I. M. Gelfand and A. B. Goncharov, "Recovery of a compactly supported function starting from its integrals over lines intersecting a given set of points in space," *Soviet Math. Dokl.*, Vol. 45, pp. 373–376, 1987.

[4] M. Defrise and R. Clack, " A cone-beam reconstruction algorithm using shift-variant filtering and cone-beam backprojection," *IEEE Trans. on Medical Imaging*, Vol. 13, pp. 186–195, 1994.

[5] M. Defrise and F. Noo and H.Kudo, "A solution to the long-object problem in helical cone-beam tomography, " *Phys. Med. Biol.*, Vol. 45, pp. 623–643, 2000.

[6] Turbell H Danielsson P-E, "Helical cone beam tomography," *Int. J. Imaging Syst. Technol.*, Vol. 11, pp. 91–100, 2000

[7] P. Grangeat, "Mathematical framework of cone–beam reconstruction via the first derivative of the Radon transform," *JCP*, Vol. 1497, pp. 66–97, 1991.

[8] F. Noo, J. Pack and D. Heuscher, "Exact helical reconstruction using native cone-beam geometries," *Institute of Physics Publishing*, Vol. 48, pp. 3787–3818, 2003.

[9] G. H. Chen, "An alternative derivation of Katsevich's cone beam reconstruction," *Med. Phys.*, Vol. 30 pp. 3017–3027, 2003.

BPF versus FBP: A comparison of reconstruction methods for circular X-ray tomography with off-center detector

Dirk Schäfer, Michael Grass

***Abstract*— We present two new variants of back-projection filtration (BPF) algorithms for circular cone-beam X-ray tomography with off-center detector in divergent beam geometry. The first one applies redundancy weighting of overlapping opposite projections before differentiation in a single projection. The second one uses the Katsevich-type differentiation involving two neighboring projections followed by redundancy weighting and back-projection. An averaging scheme is presented to mitigate streak artifacts along the Hilbert filter lines in the reconstructed volume inherent to circular BPF algorithms. The BPF-type algorithms are compared to three filtered back-projection (FBP) algorithms for off-center detector acquisitions investigated earlier. In a simulation study using the Forbild head phantom, the best image quality is obtained with the BPF algorithm based on Katsevich-type differentiation.**

I. Introduction

Circular X-ray tomography with off-center detectors saves detector area while keeping a big field-of-view. The detector is positioned asymmetrically in fan direction with respect to the central ray passing through the iso-center. Typically, the projections are heavily truncated in fan direction causing problems for filtered back-projection (FBP) algorithms. The source rotates on a full circle of 360° and only the overlapping detector part measures redundant data in the central axial plane on opposite source positions. The concept of redundancy is extended as an approximation to higher cone-angles.

We investigated three different FBP algorithms in a previous paper [9] using redundancy weighting before ramp filtering [2], redundancy weighting after ramp filtering [1] and redundancy weighting after Katsevich-type differentiation and Hilbert filtering [9]. In this paper, we present to new variants of BPF-type algorithms applied to off-center detector acquisitions. The first one is based on the circular BPF algorithm with differentiation in a single projection and redundancy weighting for short scan acquisitions by Yu et al. [11], modified for the off-center scenario and combined with the method of inverse Hilbert filtering presented by You and Zeng [10]. The second one uses the Katsevich-type differentiation to calculate the differentiated back-projection (DBP) [5][11] and combines it with the inverse Hilbert filtering of You and Zeng [10]. Additionally, we present a simple but effective method to mitigate the streak artifacts along the Hilbert filter lines in the reconstructed volume inherent to circular BPF algorithms by averaging several reconstructions along different filter line directions.

DS and MG are with Philips Research Europe - Hamburg, Röntgenstraße 24–26, 22335 Hamburg, Germany, dirk.schaefer@philips.com

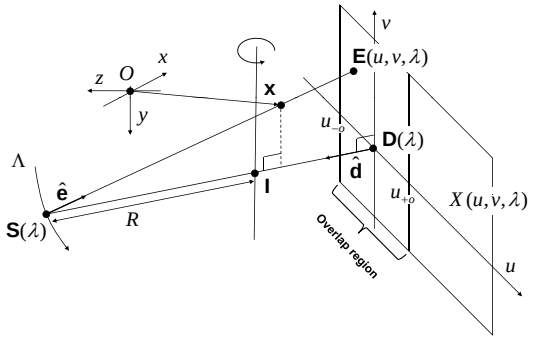


Fig. 1. Geometry for off-center circular X-ray tomography.

This paper shortly presents the reconstruction algorithms and compares the resulting image quality by means of reconstructions of simulated projection data.

II. Reconstruction Methods

A schematic view of the acquisition geometry is shown in Fig. 1. The planar detector and the X-ray source are rotated around the y-axis. The distance between source and detector is given by D. The distance from the source to the rotation axis is denoted R, and $\mathbf{I}$ represents the iso-center of the imaging system. The circular orbit is parameterized by the path length $\lambda \in \Lambda = [0, 2\pi R)$. The projected iso-center on the detector is located at $\mathbf{D}(\lambda)$ and defines the origin of the detector system. The detector v-axis is parallel to the rotational axis. Accordingly, the u-axis is parallel to the trajectory tangent vector with $u_{min} \leq u \leq u_{max}$. The cone beam projection data is denoted by $\mathcal{X}(u, v, \lambda)$:

$$\mathcal{X}(u, v, \lambda) = \int_0^\infty f(\mathbf{S}(\lambda) + l\hat{\mathbf{e}}(u, v, \lambda))\mathrm{d}l, \qquad (1)$$

where $\hat{\mathbf{e}}(u, v, \lambda)$ is the unit vector from the source position $\mathbf{S}(\lambda)$ to the detector element $\mathbf{E}(u, v, \lambda)$. The corresponding length is denoted by $\overline{SE}$. The flat detector is positioned in off-center geometry. The overlap region $\mathbb{O}(\lambda) = \{(u, v) \in \mathbb{R}^2 \mid u_{-o} \leq u \leq u_{+o}, v_{min} \leq v \leq v_{max}\}$ is defined as the symmetric region around $\mathbf{D}(\lambda)$ with measured projection values $\mathcal{X}(u, v, \lambda)$. The width of the overlap region is $\Delta u = u_{+o} - u_{-o}$.

Following the idea first introduced by Cho et al. [1], a redundancy weight $w(u)$ is introduced according to:

$$w(u) = \begin{cases} 0, & u_{min} \leq u < u_{-o} \\ \sin^2\left(\frac{\pi}{2}\frac{u - u_{-o}}{\Delta u}\right), & u_{-o} \leq u \leq u_{+o} \\ 1, & u_{+o} < u \leq u_{max}. \end{cases} \qquad (2)$$

Five different reconstruction methods are used in this article and presented in the following.

A. Pre-weighting FDK: f^{FBP1}

Cho et al. [2] have used a simple pre-weighting of the projection data with a redundancy weight similar to Eq. 2, followed by standard FDK reconstruction [3]. The weighting function mitigates artifacts that are caused by ramp filtering across the boundary of measured data and the detector edge. These pre-weighted projections are reconstructed according to the modified FDK formula given in Eq. 3:

$$f^{FBP1}(\mathbf{x}) = \int_{\Lambda} \frac{D}{|(\mathbf{x}-\mathbf{S})\cdot\hat{\mathbf{d}}|^2} \int_{-\infty}^{\infty} w(u)\,\frac{D}{\overline{SE}(\mathbf{x})} \mathcal{X}(u',v,\lambda)\; h_R(u-u')\;\mathrm{d}u'\mathrm{d}\lambda \quad (3)$$

$$\text{with:} \qquad h_R(\rho) = \int_{-\infty}^{\infty} |P| e^{j2\pi\rho P}\,\mathrm{d}P.$$

The pre-weighting FDK is exact in the axial mid-plane, where filtering and redundancy are correctly handled. However, at higher cone-angles no really redundant data exist due to the divergent geometry. Combined with the approximative FDK reconstruction, this leads to significant artifacts especially for asymmetric objects [9].

B. Post-weighting FDK with extended projections: f^{FBP2}

Redundancy weighting after filtering combined with the FDK algorithm leads to a non-exact reconstruction algorithm in the axial mid-plane. First, the truncated projections from off-center geometry are re-binned to a complete projection data set using complementary rays. The fan angle α of a specific ray u is given by $\alpha(u) = \text{atan}(u/D)$ and the source angle by $\beta = \lambda/R$. Rewriting the projection data with these coordinates gives $\tilde{\mathcal{X}}(\alpha, v, \beta) = \mathcal{X}(\text{atan}(u/D), v, \lambda/R)$. The projections are extended with the complementary rays in the region $u_{min} \le u < u_{-o}$ with $u_{min} = -u_{max}$:

$$\tilde{\mathcal{X}}_1(\alpha,v,\beta) = \tilde{\mathcal{X}}(-\alpha, v, \beta+\pi\pm 2\alpha) \quad \text{for } \alpha(u_{min}) \le \alpha \le \alpha(u_{-o}), \quad (4)$$

where the sign depends on the rotation direction. To guarantee a smooth transition of the extended data and the originally measured data, a faded additive offset correction is applied:

$$\mathcal{X}_2(u,v,\lambda) = \quad (5)$$
$$\begin{cases} \mathcal{X}_1(u,v,\lambda), & u_{min} \le u < (u_{-o}-\Delta) \\ \mathcal{X}_1(u,v,\lambda) + \delta\cos\left(\frac{\pi}{2}\frac{u_{-o}-u}{\Delta}\right), & (u_{-o}-\Delta) \le u \le u_{-o} \\ \mathcal{X}(u,v,\lambda), & u_{-o} < u \le u_{max}, \end{cases}$$

where $\delta = \mathcal{X}(u_{-o},v,\lambda) - \mathcal{X}_1(u_{-o},v,\lambda)$ defines the offset and the fading region is chosen as $\Delta = u_{+o} - u_{-o}$.
The post-weighting FDK with extended projections as presented in Eq. 6 has been proposed by Cho et al. [1][2].

$$f^{FBP2}(\mathbf{x}) = \int_{\Lambda} \frac{D}{|(\mathbf{x}-\mathbf{S})\cdot\hat{\mathbf{d}}|^2}\, w(u) \int_{-\infty}^{\infty} \frac{D}{\overline{SE}(\mathbf{x})} \mathcal{X}_2(u',v,\lambda)\; h_R(u-u')\;\mathrm{d}u'\mathrm{d}\lambda. \quad (6)$$

C. Katsevich-type FBP with extended projections: f^{FBP3}

The problem of incorrect handling of redundancy with the post-weighting FDK method can be removed by using an algorithm that correctly applies redundancy weights after the filtering step [9]:

$$f^{FBP3}(\mathbf{x}) = \frac{1}{2\pi}\int_{\Lambda} \frac{w(u)}{R-(\mathbf{x}-\mathbf{I})\cdot\hat{\mathbf{d}}} \int_{-\infty}^{\infty} \frac{D}{\overline{SE}(\mathbf{x})} \mathcal{X}_2^{KD}(u',v,\lambda)\; h_H(u-u')\;\mathrm{d}u'\,(1/R)\,\mathrm{d}\lambda, \quad (7)$$

$$\text{with} \quad h_H(\rho) = -\int_{-\infty}^{\infty} i\;\mathrm{sgn}(P) e^{j2\pi\rho P}\,\mathrm{d}P,$$

$$\text{and} \quad \mathcal{X}_2^{KD}(u,v,\lambda) = \left(\frac{\partial\mathcal{X}_2}{\partial\lambda} + \frac{\partial\mathcal{X}_2}{\partial u}\frac{\partial u}{\partial\lambda} + \frac{\partial\mathcal{X}_2}{\partial v}\frac{\partial v}{\partial\lambda}\right)$$

is the Katsevich-type derivative along the source trajectory with fixed ray direction [4]. This derivative is computed using the blended chain rule for arbitrary detector orientations derived by Noo et al. [8].

D. BPF with differentiation in a single projection

Let $\mathcal{L}(t,\mathbf{s},\hat{\mathbf{m}})$ be the line in direction of $\hat{\mathbf{m}}$ through $\mathbf{s}$. Then,

$$\mathbf{x}(t,\mathbf{s},\hat{\mathbf{m}}) = \mathbf{s} + t\hat{\mathbf{m}} \quad (8)$$

is a parametrization of the points on this line, with $t \in (-\infty,\infty)$. For those lines $\mathcal{L}$ intersecting the object support Ω, there is a finite interval $[t_{min}, t_{max}]$ corresponding to this intersection.
A BPF algorithm with differentiation in a single projection for circular acquisitions has been proposed by Yu et. al [11]. The differentiated back-projection (DBP) onto parallel lines $\mathcal{L}(t,\mathbf{s},\hat{\mathbf{m}})$ is expressed as:

$$b^{SD}(\mathbf{x}(t,\mathbf{s},\hat{\mathbf{m}}),\lambda_1,\lambda_2) = \quad (9)$$
$$\int_{\lambda_1}^{\lambda_2} \frac{D^2}{|R-(\mathbf{x}-\mathbf{I})\cdot\hat{\mathbf{d}}|^2}\cdot w_{sig}(\hat{\mathbf{m}},u,\lambda) \times \frac{\partial}{\partial u}\left[\frac{R}{\overline{SE}(\mathbf{x})}\overline{w}(\lambda,u)\mathcal{X}(u,v,\lambda)\right](1/R)\,\mathrm{d}\lambda + \frac{1}{|\mathbf{x}-\mathbf{S}|} w_{sig}(\hat{\mathbf{m}},u,\lambda)\overline{w}(\lambda,u)\mathcal{X}(u,v,\lambda)\Big|_{\lambda_1}^{\lambda_2} + \frac{2}{|\mathbf{x}-\mathbf{S}|}\overline{w}(\lambda,u)\mathcal{X}(u,v,\lambda)\Big|_{\lambda(\hat{\mathbf{m}}_1)}^{\lambda(\hat{\mathbf{m}}_2)},$$

where in the case of a full circular scan with offset detector $\lambda_1 = 0$ and $\lambda_2 = 2\pi$, so the second term in eq. 9 cancels out. The redundancy weight $\overline{w}(\lambda,u) = w(u)$ is given by the off-center weighting function from Eq. 2. The source positions $\lambda(\hat{\mathbf{m}}_1), \lambda(\hat{\mathbf{m}}_2)$ correspond to the intersections of the line $\mathcal{L}(t,\mathbf{s},\hat{\mathbf{m}})$ (projected in the central axial plane) and the source

trajectory. The weight w_{sig} accounts for the change of sign when the back-projected ray crosses the direction $\hat{\mathbf{m}}$.

$$w_{sig}(\hat{\mathbf{m}},u,\lambda)=\begin{cases}+1, & \text{if } 0 \le \operatorname{acos}(\hat{\mathbf{m}}\cdot\hat{\mathbf{S}(\lambda)}) \\ & \quad -\operatorname{atan}(u/D(\lambda)) < \pi, \\ -1 & \text{else.}\end{cases} \tag{10}$$

E. BPF with Katsevich-style differentiation

The DBP of the Katsevich-type differentiated projections $\mathcal{X}^{KD}$ onto parallel lines $\mathcal{L}(t,\mathbf{s},\hat{\mathbf{m}})$ is expressed as [5][11]:

$$b^{KD}(\mathbf{x}(t,\mathbf{s},\hat{\mathbf{m}}),\lambda_1,\lambda_2)=\int_{\lambda_1}^{\lambda_2}\frac{w(u)\cdot w_{sig}(\hat{\mathbf{m}},u,\lambda)}{|\mathbf{x}-\mathbf{S}(\lambda)|}\,\mathcal{X}^{KD}(u,v,\lambda)\,(1/R)\,\mathrm{d}\lambda. \tag{11}$$

F. Inverse finite Hilbert transform of DBP

The DBP for a non-truncated full scan or minimum data schemes is equivalent to the Hilbert transform $\mathcal{H}_{\hat{\mathbf{m}}}[f(\mathbf{x})]$ of the object function along a set of lines $\mathcal{L}(t,\mathbf{s},\hat{\mathbf{m}})$ as shown in Refs. [12][11]. Short scan or off-center detector redundancies can also be incorporated in the BPF scheme [7][6][11].

$$\begin{aligned}\mathcal{H}[f](\mathbf{x}(t,\mathbf{s},\hat{\mathbf{m}})) &= \int_{-\infty}^{\infty}\frac{1}{\pi(t-t')}f(\mathbf{x}(t,\mathbf{s},\hat{\mathbf{m}}))\mathrm{d}t' \\ &= \frac{1}{2\pi^2}b(\mathbf{x}(t,\mathbf{s},\hat{\mathbf{m}}),\lambda_1,\lambda_2).\end{aligned} \tag{12}$$

The DBP b (which can be either b^{KD} or b^{SD}) has to be known along the line $\mathcal{L}(t,\mathbf{s},\hat{\mathbf{m}})$ within an interval $[t_{L2},t_{U2}]$, that is slightly larger than the support Ω of f, such that the following condition holds [10]:

$$t_{L2} < t_{L1} < t_{min} < t_{max} < t_{U1} < t_{U2}. \tag{13}$$

Then the object f can be recovered by computing the finite inverse Hilbert Transform $\bar{\mathcal{H}}$ [10]:

$$\begin{aligned}f(\mathbf{x}(t,\mathbf{s},\hat{\mathbf{m}})) &= \frac{1}{2\pi}\bar{\mathcal{H}}[b] \\ &= \frac{1}{2\pi^2[k(t,t_{L2},t_{U2})-k(t,t_{L1},t_{U1})]} \\ &\times \int_{t_{L2}}^{t_{U2}}\frac{\mathrm{d}t'}{t'-t}\left[k(t',t_{L2},t_{U2})-k(t',t_{L1},t_{U1})\right] \\ &\times b(\mathbf{x}(t',\mathbf{s},\hat{\mathbf{m}}),\lambda_1,\lambda_2)\end{aligned} \tag{14}$$

with

$$k(t,t_L,t_U)=\begin{cases}\sqrt{(t-t_L)(t_U-t)}, & \text{if } t_L < t < t_U \\ 0 & \text{else.}\end{cases}$$

The reconstruction is denoted f^{SD} and f^{KD} when using b^{SD} and b^{KD} for the DBP. Typically, the resulting reconstruction suffers from streak artifacts along the filter line direction in regions of sharp intensity transitions. Therefore, it is beneficial to reconstruct the same object N-times along different filter directions $\hat{\mathbf{m}}_i$ and average the results to suppress the streak artifacts:

$$f_N(\mathbf{x})=\frac{1}{N}\sum_{i=1}^{N}f_i(\mathbf{x}(t,\mathbf{s},\hat{\mathbf{m}}_i)). \tag{15}$$

The object function is resampled at the final grid positions using tri-linear interpolation.

III. Results

Simulated projection data of the Forbild head phantom[1] positioned offcenter ($x=40$ mm) have been generated using 600 source positions equally sampled on 360 degree. The detector is symmetric with respect to the projected rotation axis and has a width of 756.6 mm and a height of 397.3 mm with 975×512 pixels. The distance from the source to the rotation axis is $R=881.8$ mm and to the detector $D=1325.1$ mm. The influence of the number of reconstructions with filter lines along different directions is shown for a full detector acquisition in Fig. 2. In the axial slice with an offset of 26.5 mm strong streak artifacts in the horizontal filter direction can be observed for f_1^{SD}, i.e. when using only one reconstruction with a single filter line. In the central sagittal slice the streaks originating from the ear and the nose are visible. All these artifacts are significantly reduced by averaging reconstructions over 9 different filter line directions.

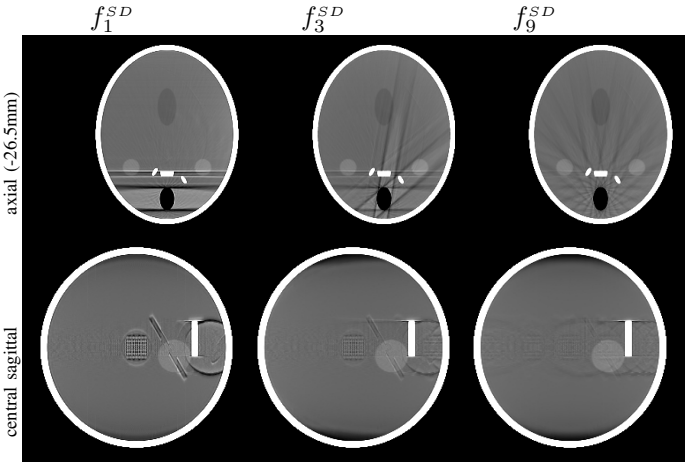


Fig. 2. Reconstructions from **full detector projections** of shifted (x=40 mm) Forbild head phantom with 1,3 and 9 reconstructions along different filter line directions, Level/Window=0/100 HU,

The off-center detector acquisitions are generated by setting all projection data with $u < u_{-o} = -15.5$ mm to zero. Reconstructions of the phantom using the three different FBP algorithms and the two BPF algorithms are shown in Figures 3,4,5 for the central axial, coronal and sagittal slices. The BPF reconstructions are averaged over 9 reconstructions with equi-angular spaced filter directions. The central axial slice is almost identical for all methods except for f^{FBP2} using the incorrect combination of ramp filtering and redundancy weighting after filtering in divergent geometry leading to low frequent shading artifacts. The alternating artifacts for the FBP methods reported in Ref. [9] for different objects can be retrieved in the central coronal and especially in the central sagittal slice (Fig. 5). The best FBP method clearly is f^{FBP3} with reduced artifacts at higher cone angles and without shading artifacts as observed for f^{FBP2}.

[1] http://www.imp.uni-erlangen.de/forbild/deutsch/results/head/head.html

Interestingly, the BPF reconstructions f_9^{SD} using the off-center detector redundancy weighting before the differentiation in a single projection similar to FBP1 show the same type of artifacts at higher cone angles as f^{FBP1}. This artifact is not present in the central sagittal slice of the full detector reconstruction f_9^{SD} (see Fig. 2). The best image quality is obtained using the BPF algorithm with Katsevich-type differentiation and off-center detector redundancy weighting after the differentiation, i.e. for f_9^{KD}. No alternating artifacts are observed, neither in the central coronal nor in the sagittal plane. Only a slightly increased low intensity drop in the coronal plane at highest cone angles close to the calotte is visible compared to the FBP methods.

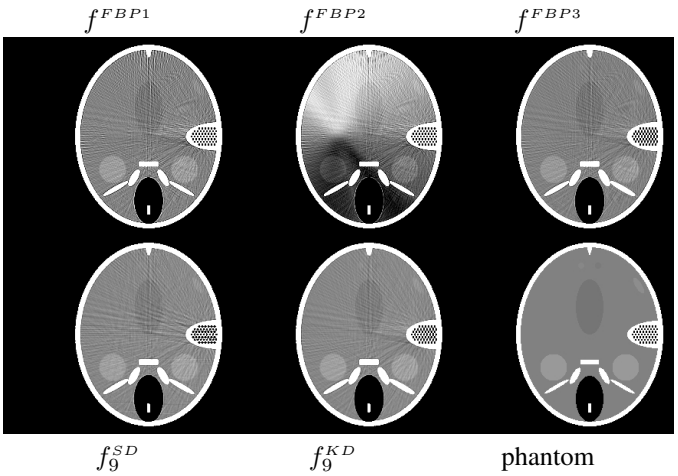


Fig. 3. Central axial (y) slice: Reconstructions from **off-center detector projections** of shifted (x=40 mm) Forbild head phantom, Level/Window=0/100 HU.

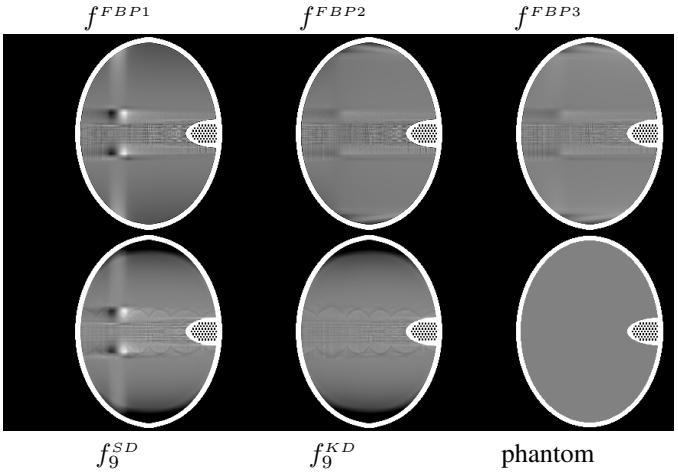


Fig. 4. Central coronal (z) slice: Reconstructions from **off-center detector projections** of shifted (x=40 mm) Forbild head phantom, Level/Window=0/100 HU.

IV. Conclusions

Two BPF algorithms for off-center detector acquisitions have been presented and compared to three existing FBP algorithms. Those algorithms that apply the off-center detector redundancy

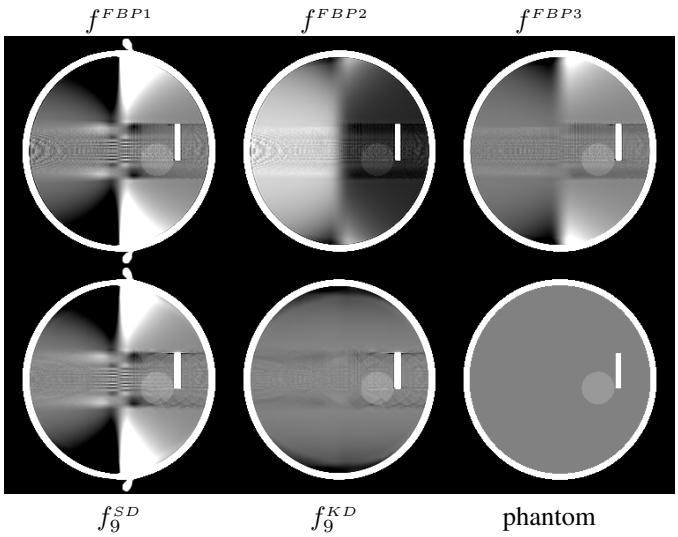


Fig. 5. Central sagittal (x) slice: Reconstructions from **off-center detector projections** of shifted (x=40 mm) Forbild head phantom, Level/Window=0/100 HU.

weights after the filtering show less artifacts. This holds for FBP and BPF algorithms. The FBP algorithms using the redundancy weights after the filtering, however, need approximated (re-binned) projection data from the opposite side of the circular trajectory for the filtering step. This drawback is overcome using BPF methods, which apply only the short range differentiation on the projection and the long range Hilbert filter on the DBP volume, and hence do not use any re-binned or approximated projection data. Surprisingly, the BPF framework in the off-center detector setting is also sensitive to the order of differentiation and redundancy weighting. The best image quality is obtained using the BPF algorithm with Katsevich-style differentiation. Averaging of several reconstructions using different filter lines mitigates the streak artifact problem inherent to the BPF methods.

References

[1] P. Cho, R. Johnson, T. Griffin, Phys. Med. Biol. Vol. 40, 1863-83, 1995.
[2] P. Cho, A. Rudd, R. Johnson, Computerized Medical Imaging and Graphics, Vol. 20, 49-57, 1996.
[3] L. Feldkamp, L. Davis, J. Kress, J. Opt. Soc. Am. A, 612-19, 1984.
[4] A. Katsevich, Phys. Med. Biol. Vol. 47, 2583–97, 2002.
[5] L. Li, ZQ. Chen, L. Zhang, YX. Xing, KJ. Kang, High Energy Physics and Nuclear Physics - Chinese Edition, Vol. 30(8), 812-17, 2006.
[6] F. Noo, M. Defrise, R. Clackdoyle, H. Kudo, Phys. Med. Biol., Vol. 47, 2525–46, 2002.
[7] F. Noo, R. Clackdoyle, JD. Pack, Phys. Med. Biol., Vol. 49, 3903, 2004.
[8] F. Noo, S. Hoppe, F. Dennerlein, G. Lauritsch, J. Hornegger, Phys. Med. Biol., Vol. 52,5393–414, 2007.
[9] D. Schäfer, M. Grass, Proc. Fully 3D Conference, Beijing, China, 86-89, 2009.
[10] J. You, GL. Zeng, Inverse Problems 22, L7–L10, 2006.
[11] L. Yu, Y. Zou, EY. Sidky, CA. Pelizzari, P. Munro, X. Pan, IEEE TMI, Vol. 25(7), 869-881, 2006.
[12] Y. Zou, X. Pan, Phys. Med. Biol. 49, 941–959, 2004.

A Novel Implementation of Backprojected Filtration Algorithm for Helical Cone-beam CT

Yunsong Zhao, Yining Zhu, and Peng Zhang*

***Abstract*—To solve the long-object problem, helical cone-beam scanning configuration was proposed, and several exact or approximate reconstruction algorithms were invented. Compared with approximate algorithms (like FDK algorithm), exact algorithms are free of cone-beam artifacts. Fast implementation of exact reconstruction algorithms are of high importance in clinical and industrial applications. In the present paper, we propose a new implementation of the backprojected filtration algorithm for helical cone-beam CT. The advantages of our implementation are: (i) it can directly reconstruct horizontal slices (z-perpendicular slices), avoiding 3-D interpolation; (ii) it needs not to calculate the backprojection interval for any point; (iii) it is of high degree of parallelism and is suitable for acceleration on GPU (Graphic Processing Unit) or other parallel systems.**

***Index Terms*—Cone-beam tomography, helical CT, horizontal slice reconstruction, fast implementation.**

I. Introduction

Compared with fan beam scanning configuration, cone-beam scanning configuration has many advantages, including higher dose usage and consistency spatial resolution. Since the 1980s, research on cone-beam CT has been drawing researchers' attention. Especially in recent years, with the rapid development of panel detector technology and computer technology, research on cone-beam CT has become an international hot topic.

Helical cone-beam scanning configuration, which could be easily carried out in engineering, was proposed to solve the long-object problem. From Tuy's result [1], theoretically, the density function of the long object can be exactly reconstructed from projections acquired from this scanning configuration. But it was not until 2002 that Katsevich gave the first exact filtered backprojection (FBP) reconstruction algorithm [2], which permits axial data truncation and has a higher efficiency than the algorithms based on Radon transform [3], [4]. Another type of exact reconstruction algorithm, called backprojected filtration (BPF) algorithm, was achieved by Zou and Pan [5] in 2004. A special feature of this algorithm is exact reconstruction of ROI (regions of interest). Pack *et al* [6] extended Zou and Pan's BPF algorithm to general scanning trajectories and gave a formula that revealed the relation between the cone-beam projections and the Hilbert transform of the density function of the inspected object. Based on this formula, density function can be reconstructed along R-line or M-line [6], where an R-line is any line segment that connects two points of the helix, and an M-line is any line segment that connects the X-ray source and a cell of the panel detector. An R-line is a so-called PI-line [7] when the two end points of the R-line are in one helix turn. Comprehensive surveys on cone-beam reconstruction algorithms were given in [8] and [9].

To our knowledge, R-line(including PI-line)-based BPF reconstruction algorithms need first backproject derivatives of projections to a set of R-lines to obtain the Hilbert transform of the density function along the R-lines; and then perform a finite inverse Hilbert transform along the R-lines to reconstruct the density function on the R-lines. In order to obtain CT images of horizontal slices (i.e. slices vertical to the rotation axis), a 3-D interpolation procedure is required, which not only increases the computational cost but also affects the accurateness of the reconstructed images. Pack *et al* [6] reconstructed a set of images on M-line surfaces. But the backprojection interval for each point to be reconstructed needs to be determined, which leads to solving a nonlinear equation numerically that is time-consuming.

In the present paper, we propose a novel implementation of the BPF reconstruction algorithm for helical cone-beam CT. Compared with existing implementations, our implementation has the following advantages: (i) it can directly reconstruct horizontal slices, avoiding 3-D interpolation; (ii) it needs not to calculate the backprojection interval for any point; (iii) it is of high degree of parallelism and is suitable for acceleration on GPU or other parallel systems, since the reconstruction processes for different slices are independent.

The remaining parts of this paper are organized as follows. In section 2 we introduce the basic theory on the BPF reconstruction algorithm for helical cone-beam CT. In section 3 we describe our fast implementation of the BPF reconstruction algorithm. In section 4 we verify the correctness of our implementation using both simulated data and real data. The efficiency of our implementation is also tested. A summary is given in the final section.

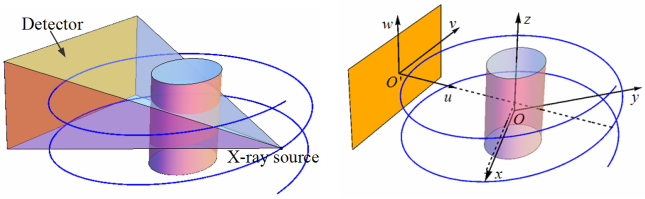
II. Basic theory

In this section we introduce the basic theory on the BPF reconstruction algorithm for helical cone-beam CT. First we give the mathematical descriptions and notations for helical cone-beam scanning configuration. Then we define the Hilbert transform of a 3-D function along a straight line and introduce

This work was supported in part by National Science Foundation of China under Grant 60971131/F010404. *Asterisk indicates corresponding author.*

Yunsong Zhao and Yining Zhu are with the CT Laboratory, School of Mathematics, Capital Normal University, Beijing 100048, China. (e-mail: zhao_yunsong@163.com, zhumilan997@163.com).

*Peng Zhang is with the CT Laboratory, School of Mathematics, Capital Normal University, Beijing 100048, China. (phone: +86-10-68903094; fax: 86-10-68903094; e-mail: pzhangct@sina.com).



(a) Helical cone-beam scanning configuration. (b) Definition of coordinate systems.

Fig. 1. Scanning configuration and definition of cooradinate systems.

its finite inverse transform. Last we derive the relation between the cone-beam projections and the Hilbert transform of a 3-D function, which is the foundation of the next section.

A. Scanning configuration

The helical cone-beam scanning configuration is shown in Fig. 1(a). The cone-beam, which is composed of the X-ray source and the panel detector, rotates around the object and moves axially. The X-ray source emits photons and then the detector collects the photons penetrating the inspected object. Let the rotation axis of the cone-beam be the z axis, we establish the object coordinate system $Oxyz$. The helical trajectory of the X-ray source can be defined as:

$$\vec{a}(\phi) = (R\cos\phi, R\sin\phi, h\frac{\phi}{2\pi}), \phi \in \mathbb{R}, \tag{1}$$

where R denotes the distance from the X-ray source to Oz axis, and h indicates the helical pitch. Denote the orthogonal projection of the X-ray source onto the detector plane as O'. With O' being the origin we establish the cone-beam coordinate system $O'uvw$ such that $O'w$ is parallel to Oz and $O'u$ points to the X-ray source from O', that is the cone-beam coordinate system rotates with the rotation of the X-ray source and the detector. $O'u, O'v, O'w$ forms a right-hand system, as shown in Fig. 1(b). Let D denote the distance from the X-ray source to the detector.

For a given point in 3-D space, its coordinates $\vec{x} = (x, y, z)$ in the object system and coordinates $\vec{u} = (u, v, w)$ in the cone-beam system can be related through

$$\vec{u} = \big(\vec{x} - \vec{k}(\phi)\big)M(\phi) \tag{2}$$

and

$$\vec{x} = \vec{u}M(-\phi) + \vec{k}(\phi), \tag{3}$$

where

$$M(\phi) = \begin{pmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{4}$$

$$\vec{k}(\phi) = \big((R-D)\cos\phi, (R-D)\sin\phi, h\frac{\phi}{2\pi}\big). \tag{5}$$

Let $f(\vec{x})$ (or f) denote the density function of the inspected object and the support of f is denoted as Ω. Furthermore, it is assumed that $f \geq 0$ and is continuously differentiable everywhere.

B. Hilbert transform and its inversion

The Hilbert transform of a 1-D function (say $g(t)$) is defined as a convolution with the kernel $\frac{1}{\pi t}$,

$$Hg(t) = g(t) * \frac{1}{\pi t} = \int_{-\infty}^{+\infty} \frac{g(t')}{\pi(t-t')}\mathrm{d}t'. \tag{6}$$

The Hilbert transform of a 3-D function is an extension of that of a 1-D function. The line passing though the point $\vec{x}_0$ and parallel to $\vec{n}$ (a unit vector) can be represented as

$$L(\vec{x}_0, \vec{n}) : \ \vec{x}_0 + t\vec{n}, t \in (-\infty, +\infty). \tag{7}$$

On the line $L(\vec{x}_0, \vec{n})$, a 3-D function $f(\vec{x})$ can be regraded as a 1-D function $f(\vec{x}_0 + t\vec{n})$ on t. The Hilbert transform of $f(\vec{x})$ along $L(\vec{x}_0, \vec{n})$ is defined as the Hilbert transform of $f(\vec{x}_0 + t\vec{n})$ on the variable t,

$$Hf(t, \vec{x}_0, \vec{n}) = \int_{-\infty}^{+\infty} \frac{1}{\pi(t-t')} f(\vec{x}_0 + t'\vec{n})\mathrm{d}t'. \tag{8}$$

The value of $Hf(t, \vec{x}_0, \vec{n})$ at the point $\vec{x} \in L(\vec{x}_0, \vec{n})$ is $Hf(t, \vec{x}_0, \vec{n})|_{t=(\vec{x}-\vec{x}_0)\cdot\vec{n}}$. As $\vec{x}$ is on the line $L(\vec{x}_0, \vec{n})$, $Hf(t, \vec{x}_0, \vec{n})|_{t=(\vec{x}-\vec{x}_0)\cdot\vec{n}}$ is independent of $\vec{x}_0$. So the Hilbert transform can be simplified as

$$Hf(\vec{x}, \vec{n}) := Hf(t, \vec{x}_0, \vec{n})|_{t=(\vec{x}-\vec{x}_0)\cdot\vec{n}} = Hf(t, \vec{x}, \vec{n})|_{t=0}. \tag{9}$$

Substitute (8) into (9), we have

$$Hf(\vec{x}, \vec{n}) = -\int_{-\infty}^{+\infty} \frac{1}{\pi t'} f(\vec{x} + t'\vec{n})\mathrm{d}t'. \tag{10}$$

Note from (10) that $Hf(\vec{x}, \vec{n})$ is odd in $\vec{n}$, that is

$$Hf(\vec{x}, -\vec{n}) = -Hf(\vec{x}, \vec{n}). \tag{11}$$

If $f(\vec{x})$ has a impact support, there is a finite inverse Hilbert transform [10] which can be used to recover $f(\vec{x})$ from the values of $Hf(\vec{x}, \vec{n})$ in the support Ω: If there are $t_{\min}(\vec{x}_0, \vec{n})$ and $t_{\max}(\vec{x}_0, \vec{n})$ (or $t_{\min}, t_{\max}$ for simplicity) such that $Hf(t, \vec{x}_0, \vec{n})$ is known for $t \in [t_{\min}, t_{\max}]$, and exists a $\varepsilon(\vec{x}_0, \vec{n})$ (ε for simplicity) such that

$$f(\vec{x}_0 + t\vec{n}) = 0, t \notin [t_{\min} + \varepsilon, t_{\max} - \varepsilon], \tag{12}$$

then

$$\begin{aligned} f(\vec{x}) = & \frac{-1}{\sqrt{(t-t_{\min})(t_{\max}-t)}} \\ & \times\Bigg(\int_{t_{\min}}^{t_{\max}} \sqrt{(t'-t_{\min})(t_{\max}-t')}\frac{Hf(t', \vec{x}_0, \vec{n})}{\pi(t-t')}\mathrm{d}t' \\ & + C(\vec{x}_0, \vec{n})\Bigg)\Bigg|_{t=(\vec{x}-\vec{x}_0)\cdot\vec{n}}, \end{aligned} \tag{13}$$

where $C(\vec{x}_0, \vec{n})$ is a constant, and is the same for all points $\vec{x}_0 + t\vec{n}, t \in [t_{\min} + \varepsilon, t_{\max} - \varepsilon]$, which can be determined from the knowledge of $f(\vec{x}_0 + t\vec{n})$ at some point $t \in [t_{\min}, t_{\min} + \varepsilon] \cup [t_{\max} - \varepsilon, t_{\max}]$ [11].

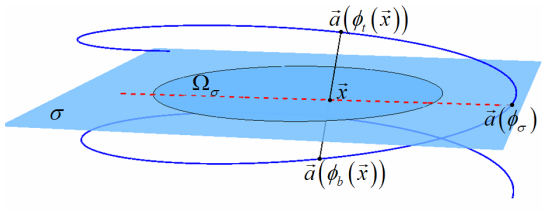


Fig. 2. PI-line and M-line: the solid line segment is a PI-line, the dashed half-line is a M-line.

C. The relation between the cone-beam projections and the Hilbert transform of the density function

The ray that emits from the X-ray source and projects onto the address (v, w) can be expressed in cone-beam coordinate system as

$$\vec{u}(t) = (D, 0, 0) + t\frac{(-D, v, w)}{|(-D, v, w)|}, t > 0. \tag{14}$$

by (3), the ray in the object system can be expressed as

$$\vec{x}(t) = \big((D, 0, 0) + t\frac{(-D, v, w)}{|(-D, v, w)|}\big)M(-\phi) + \vec{k}(\phi). \tag{15}$$

Let $p(\phi, v, w)$ denote the cone-beam projection, then we have

$$p(\phi, v, w) = \int_0^\infty f\Big(\big((D, 0, 0) + t\frac{(-D, v, w)}{|(-D, v, w)|}\big) \times M(-\phi) + \vec{k}(\phi)\Big)\mathrm{d}t. \tag{16}$$

The relation between the cone-beam projections and the Hilbert transform of f can be derived as

$$Hf(\vec{x}, \vec{n}(\phi, \vec{x}))|_{\phi_1}^{\phi_2} = \frac{1}{\pi}\Big(\int_{\phi_1}^{\phi_2} \frac{|(-D, v^*, w^*)|^2|\vec{l}|}{|\vec{x} - \vec{a}(\phi)|^2} \frac{\mathrm{d}}{\mathrm{d}\vec{l}}\frac{p(\phi, v, w)}{|(-D, v, w)|}\Big|_{\substack{v=v^*\\w=w^*}} \mathrm{d}\phi + \frac{p(\phi, v^*, w^*)}{|\vec{x} - \vec{a}(\phi)|}\Big|_{\phi_1}^{\phi_2}\Big), \tag{17}$$

where $\vec{n}(\phi, \vec{x})$ is the unit vector from the X-ray source to $\vec{x}$, $(v^*, w^*) = (v^*(\phi, \vec{x}), w^*(\phi, \vec{x}))$ is the projection address of $\vec{x}$ at angle ϕ, $\frac{\mathrm{d}}{\mathrm{d}\vec{l}}(\cdot)$ is the directional derivative along $\vec{l} = (R, \frac{h}{2\pi})$ in the detector plane. The left hand side of (17) is the difference between the two Hilbert transforms of $f(\vec{x})$ at $\vec{x}$ along $\vec{n}(\phi_2, \vec{x})$ and $\vec{n}(\phi_1, \vec{x})$. We have

$$\vec{n}(\phi, \vec{x}) = \frac{\vec{x} - \vec{a}(\phi)}{|\vec{x} - \vec{a}(\phi)|}, \tag{18}$$

$$v^*(\phi, \vec{x}) = \frac{-Dy\cos\phi + Dx\sin\phi}{x\cos\phi + y\sin\phi - R}, \tag{19}$$

$$w^*(\phi, \vec{x}) = \frac{D[2\pi z - h\phi]}{2\pi[R - x\cos\phi - y\sin\phi]}. \tag{20}$$

III. The implementation of direct horizontal slice reconstruction

In this section we derive a reconstruction formula which can directly reconstruct horizontal slices of the inspected object. In addition, we propose strategies for fast numerical implementation of the reconstruction formula.

A. Reconstruction formula

Consider horizontal slice σ. As shown in Fig. 2, the intersection region of σ with the inspected object is denoted as Ω_σ and the intersection point of σ with the helix is denoted as $\vec{a}(\phi_\sigma)$, where ϕ_σ is the angular parameter of the intersection point. For any given σ, the fan-beam emitted from $\vec{a}(\phi_\sigma)$ covers the whole Ω_σ. The derivation of the horizontal reconstruction formula is based on the following two observations: (i) The Hilbert transform of $f(\vec{x})$ along the fan-beam can be calculated from the projections; (ii) the CT image of the slice can be obtained by the finite inverse Hilbert transform.

Now we analysis how to get the Hilbert transform of the density function of the inspected object. It is pointed out in [7] that there exists one and only one PI-line passing though $\vec{x} \in \Omega_\sigma$. Denote the intersection points of the PI-line passing though $\vec{x}$ with the helix as $\vec{a}(\phi_b(\vec{x}))$ and $\vec{a}(\phi_t(\vec{x}))$ (or $\vec{a}(\phi_b)$ and $\vec{a}(\phi_t)$ for simplicity, but beware that ϕ_b and ϕ_t are functions of $\vec{x}$). On the PI-line $\overline{\vec{a}(\phi_b)\vec{a}(\phi_t)}$, $\vec{x}$ is located between the two end points, so the two vector from $\vec{a}(\phi_b)$ and $\vec{a}(\phi_t)$ to $\vec{x}$ are of opposite directions, that is

$$\vec{n}(\phi_b, \vec{x}) = -\vec{n}(\phi_t, \vec{x}). \tag{21}$$

Noting that $Hf(\vec{x}, \vec{n})$ is odd in $\vec{n}$, we have

$$Hf(\vec{x}, \vec{n}(\phi_b, \vec{x})) = -Hf(\vec{x}, \vec{n}(\phi_t, \vec{x})). \tag{22}$$

By the above formula, we have

$$Hf(\vec{x}, \vec{n}(\phi_\sigma, \vec{x})) = \frac{1}{2}\big(Hf(\vec{x}, \vec{n}(\phi, \vec{x}))|_{\phi_b}^{\phi_\sigma} + Hf(\vec{x}, \vec{n}(\phi, \vec{x}))|_{\phi_t}^{\phi_\sigma}\big). \tag{23}$$

Let $\phi_1 = \phi_b, \phi_2 = \phi_\sigma$ and $\phi_1 = \phi_t, \phi_2 = \phi_\sigma$, substitute (17) into (23), we have

$$\begin{aligned} Hf(\vec{x}, \vec{n}(\phi_\sigma, \vec{x})) = {} & \frac{1}{2\pi}\Big(\int_{\phi_b}^{\phi_\sigma} \Big(\frac{|(-D, v, w)|^2|\vec{l}|}{|\vec{x} - \vec{a}(\phi)|^2}\frac{\mathrm{d}}{\mathrm{d}\vec{l}}\frac{p(\phi, v, w)}{|(-D, v, w)|}\Big)\Big|_{\substack{v=v^*\\w=w^*}} \mathrm{d}\phi + \frac{p(\phi, v^*, w^*)}{|\vec{x} - \vec{a}(\phi)|}\Big|_{\phi_b}^{\phi_\sigma}\Big) \\ & + \frac{1}{2\pi}\Big(\int_{\phi_t}^{\phi_\sigma} \Big(\frac{|(-D, v, w)|^2|\vec{l}|}{|\vec{x} - \vec{a}(\phi)|^2}\frac{\mathrm{d}}{\mathrm{d}\vec{l}}\frac{p(\phi, v, w)}{|(-D, v, w)|}\Big)\Big|_{\substack{v=v^*\\w=w^*}} \mathrm{d}\phi + \frac{p(\phi, v^*, w^*)}{|\vec{x} - \vec{a}(\phi)|}\Big|_{\phi_t}^{\phi_\sigma}\Big). \end{aligned} \tag{24}$$

By the above formula we can get the Hilbert transform of $f(\vec{x})$ from $p(\phi, v, w)$. As the arbitrariness of $\vec{x}$ and that the fan-beam covers the whole Ω_σ, we can obtain all the values of the Hilbert transform of $f(\vec{x})$ along the fan-beam in Ω_σ,

then by the finite inverse Hilbert transform we can reconstruct the image of the slice σ.

There are some difficulties in calculating $Hf(\vec{x},\vec{n}(\phi_\sigma,\vec{x}))$ by (24) numerically: the limits of the integral $\phi_b(\vec{x})$ and $\phi_t(\vec{x})$ are different for different points $\vec{x}\in\Omega_\sigma$. So we need to calculate $\phi_b(\vec{x})$ and $\phi_t(\vec{x})$ for all $\vec{x}\in\Omega_\sigma$, which can be done by solving non-linear equations numerically, but the computation is large.

To obtain an efficient implementation, we rewrite (24) into a new form. It is mainly based on the following findings: (i) The shape of Tam-Danielsson window [7] on the detector plane, the dashed area shown in Fig. 3, does not change with views; (ii) For any given point $\vec{x}\in\Omega_\sigma$, its projection addresses (v^*,w^*) on the detector plane form a continuous curve as ϕ changes, as shown in Fig. 4. (iii) (v^*,w^*) lies in the Tam-Danielsson window iff $\phi\in[\phi_b,\phi_t]$. (iv) $w^*>0$, if $\phi\in[\phi_b,\phi_\sigma)$, $w^*<0$ if $\phi\in(\phi_\sigma,\phi_t]$ and $w^*=0$, if $\phi=\phi_\sigma$. (v) the projection of $\vec{x}$ is just on the top border of the Tam-Danielsson window if $\phi=\phi_b$, and on the bottom border of the Tam-Danielsson window if $\phi=\phi_t$. Define

$$\bar{p}(\phi,v,w)=\mathrm{sgn}(w)\chi_{\mathrm{TD}}(v,w)\frac{d}{d\vec{l}}\frac{p(\phi,v,w)}{|(-D,v,w)|} \tag{25}$$

and

$$\tilde{p}(\phi,v,w)=\chi_{\mathrm{TDB}}(v,w)p(\phi,v,w), \tag{26}$$

where $\mathrm{sgn}(w)$ is a sign function, $\chi_{\mathrm{TD}}(v,w)$ is the characteristic function of Tam-Danielsson window and $\chi_{\mathrm{TDB}}(v,w)$ is the characteristic function of the border of the Tam-Danielsson window. Now we can rewrite (24) as

$$\begin{aligned}
&Hf(\vec{x},\vec{n}(\phi_\sigma,\vec{x}))\\
&=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\left(\frac{|(-D,v^*,w^*)|^2|\vec{l}|\bar{p}(\phi,v^*,w^*)}{|\vec{x}-\vec{a}(\phi)|^2}\right.\\
&\quad\left.-\frac{p(\phi,v^*,w^*)}{|\vec{x}-\vec{a}(\phi)|}(\delta(\phi-\phi_b)+\delta(\phi-\phi_t))\right)\mathrm{d}\phi\\
&\quad+\frac{p(\phi_\sigma,v^*(\phi_\sigma,\vec{x}),w^*(\phi_\sigma,\vec{x}))}{\pi|\vec{x}-\vec{a}(\phi_\sigma)|}\\
&=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\left(\frac{|(-D,v^*,w^*)|^2|\vec{l}|\bar{p}(\phi,v^*,w^*)}{|\vec{x}-\vec{a}(\phi)|^2}\right.\\
&\quad\left.-\frac{\tilde{p}(\phi,v^*,w^*)}{|\vec{x}-\vec{a}(\phi)|}(\delta(\phi-\phi_b)+\delta(\phi-\phi_t))\right)\mathrm{d}\phi\\
&\quad+\frac{p(\phi_\sigma,v^*(\phi_\sigma,\vec{x}),w^*(\phi_\sigma,\vec{x}))}{\pi|\vec{x}-\vec{a}(\phi_\sigma)|},
\end{aligned} \tag{27}$$

where $\delta(\cdot)$ presents the Dirac function. Although the integral interval in (27) is $(-\infty,+\infty)$, the integrand is 0, if $\phi\notin[\phi_b(\vec{x}),\phi_t(\vec{x})]$. Take a proper Δ such that

$$\phi_b(\vec{x}),\phi_t(\vec{x})\in[\phi_\sigma-\Delta,\phi_\sigma+\Delta],\forall\vec{x}\in\Omega, \tag{28}$$

then for any $\vec{x}\in\Omega_\sigma$, $Hf(\vec{x},\vec{n}(\phi_\sigma,\vec{x}))$ can be obtained by

$$\begin{aligned}
&Hf(\vec{x},\vec{n}(\phi_\sigma,\vec{x}))\\
&=\frac{1}{2\pi}\int_{\phi_\sigma-\Delta}^{\phi_\sigma+\Delta}\left(\frac{|(-D,v^*,w^*)|^2|\vec{l}|\bar{p}(\phi,v^*,w^*)}{|\vec{x}-\vec{a}(\phi)|^2}\right.\\
&\quad\left.-\frac{\tilde{p}(\phi,v^*,w^*)}{|\vec{x}-\vec{a}(\phi)|}(\delta(\phi-\phi_b)+\delta(\phi-\phi_t))\right)\mathrm{d}\phi\\
&\quad+\frac{p(\phi_\sigma,v^*(\phi_\sigma,\vec{x}),w^*(\phi_\sigma,\vec{x}))}{\pi|\vec{x}-\vec{a}(\phi_\sigma)|}.
\end{aligned} \tag{29}$$

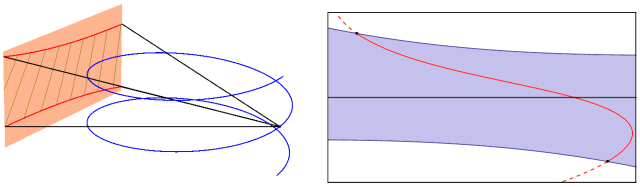
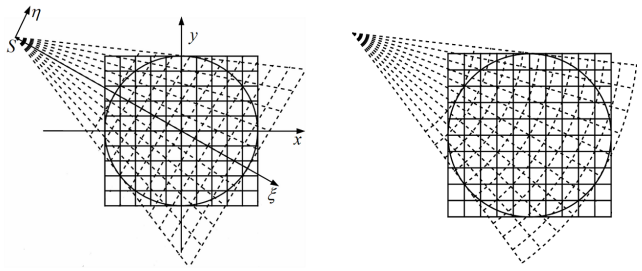


Fig. 3. The Tam-Danielsson window: the dashed area on the detector plane.

Fig. 4. Projection curve of a given point on the detector plane.



(a) Trapezoidal sampling. (b) Polar sampling.

Fig. 5. Sampling modes.

The advantage of (29) is that the computation of the backprojection intervals for different points is no longer needed. We use the same backprojection interval for all $\vec{x}\in\Omega_\sigma$, so the parallelism of the implementation is improved significantly. Of course, using the same backprojection interval for different points will add some extra costs. But it is worthwhile compared with computing the backprojection interval for each point. With the scanning parameters given in Section IV, the added computation for backprojection is no more than 3%.

The minimum Δ satisfying (28) for a given FOV with radius r can be obtained by solving the following extreme value problem:

$$\Delta=\max_{\alpha\le\phi\le2\pi-\alpha}\frac{1+\sqrt{\tau^2+(\tau^2-1)\cot^2(\frac{\phi}{2})}}{2}\phi, \tag{30}$$

where $\alpha=2\cos^{-1}\tau$, $\tau=\frac{r}{R}$. Note that different slices have the same Δ, so we need only to calculate Δ once.

B. Sampling and interpolation

When implementing the algorithm numerically, we need to sample the backprojection points. As a finite inverse Hilbert transform along the fan-beam is needed after backprojection, we suggest two sampling modes, called trapezoidal sampling mode and polar sampling mode respectively, as shown in Fig. 5. On the figure, the intersection point of dashed lines are sampling points and the intersection points of solid lines are Cartesian grids.

After reconstructing values of $f(\vec{x})$ on sampling points, we can get the values of $f(\vec{x})$ on Cartesian grids via a 2-D interpolation procedure. We take trapezoidal sampling mode as an example to show this interpolation procedure. Set up coordinate systems as shown in Fig. 5(a), then according to

the coordinate transform from Oxy to $S\xi\eta$

$$\begin{cases} \xi = R - x\cos\phi - y\sin\phi \\ \eta = -y\cos\phi + x\sin\phi \end{cases} \tag{31}$$

we can get values of $f(\vec{x})$ on Cartesian grids by interpolation from its values on trapezoidal grids.

C. Steps of numerical implementation

We describe the numerical implementation procedure to reconstruct the image of slice σ as follows:

Step1. Initialization:

i) Calculate Δ by solving (30).

ii) Initialize $f(\vec{x}) := 0$ for every sampling point $\vec{x}$ on the slice.

Step2. Image reconstruction:

i) For every discrete point $\phi \in [\phi_\sigma - \Delta, \phi_\sigma + \Delta]$ calculate $\bar{p}(\phi, v, w)$ and $\tilde{p}(\phi, v, w)$ by (25) and (26).

ii) For every discrete point $\phi \in [\phi_\sigma - \Delta, \phi_\sigma + \Delta]$, do

$$f(\vec{x}) := f(\vec{x}) + \frac{1}{2\pi}\frac{|(-D, v^*, w^*)|^2 |\vec{l}| \bar{p}(\phi, v^*, w^*)}{|\vec{x} - \vec{a}(\phi)|^2}\Delta\phi - \frac{1}{2\pi}\frac{\tilde{p}(\phi, v^*, w^*)}{|\vec{x} - \vec{a}(\phi)|}. \tag{32}$$

iii) Do

$$f(\vec{x}) := f(\vec{x}) + \frac{p(\phi_\sigma, v^*(\phi_\sigma, \vec{x}), w^*(\phi_\sigma, \vec{x}))}{\pi|\vec{x} - \vec{a}(\phi_\sigma)|}. \tag{33}$$

iv) Do finite inverse Hilbert transform by (13) to get the density value of the inspected object on sampling points, which can be speeded up by FFT.

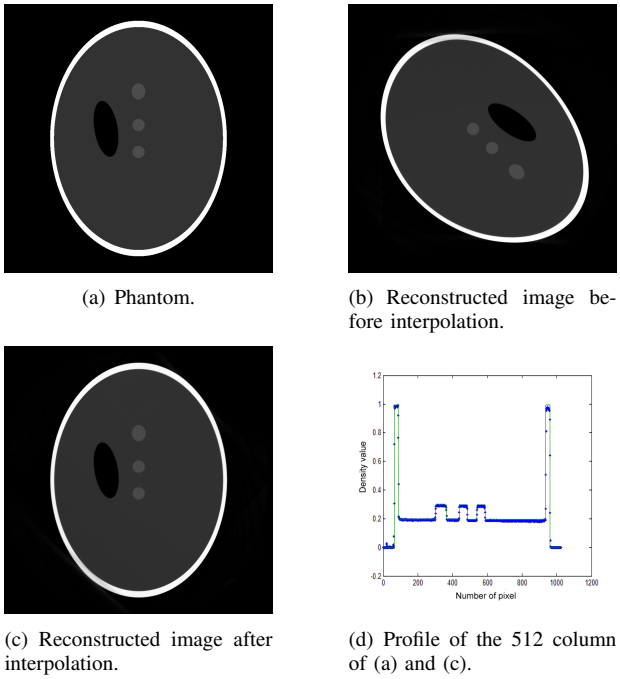
v) Do interpolation according to (31) to get the density value of the inspected object on Cartesian grids.

IV. Numerical experiments

In this section we perform numerical experiment with both simulated data and real data to verify the validation of our implementation. The parameters of the scanning configuration are as follows: the distance from the X-ray source to the z axis $R = 1000$ mm, the distance from the X-ray source to detector plane $D = 1230$ mm, and the helical pitch $h = 195$ mm. The panel detector is composed of 1920×1536 detector cells , with each cell covering an area of 0.127×0.127 mm^2. 720 projections are collected per turn.

A. Reconstruction results from simulated data

We take 3-D Shepp-Logan phantom as an example to get simulated data. The phantom has an ellipsoid support with half axes of 68.0 mm, 90.7 mm and 71.9 mm along the x, y and z axis, respectively. Projections are computed using analytical formula for line integrals through all objects composing the phantom. Fig. 6 illustrates the reconstructed images of slice $z = 24.375$ mm and the size of the images is 1024×1024.



(a) Phantom.

(b) Reconstructed image before interpolation.

(c) Reconstructed image after interpolation.

(d) Profile of the 512 column of (a) and (c).

Fig. 6. Reconstructed images at $z = 24.375$ mm. Gray scale: $[0, 1]$.

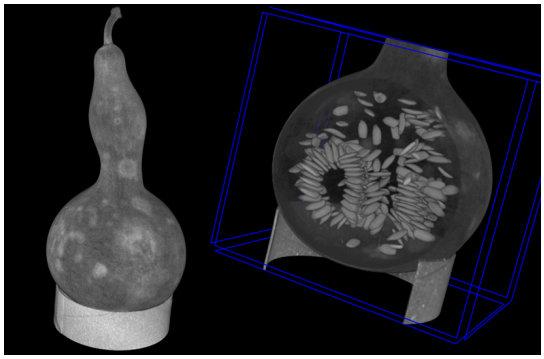


Fig. 7. 3-D image of the calabash and its partial cutaway view.

B. Reconstruction results from real data

The real data are acquired from the CT devices that are developed by our laboratory. The inspected object is a calabash whose height is 255 mm. The height of the detector is 195.072 mm, so the cone-beam can not cover the whole calabash. With the above parameters, the scan with two turns is enough to collect projections needed to reconstruct the whole calabash. The reconstructed 3-D image of the calabash is shown in Fig. 7. From the 3-D image and its partial cutaway view, we can clearly see the fungus spots on the surface of the calabash and the distribution of the seeds.

C. Efficiency test

The proposed implementation is of high degree of parallelism. We have tested our implementation using two Quadro FX5800 graphics cards (NVIDIA, Sanata Clara, CA) with 4GB video memory each. CUDA (2.3) is used to program the GPUs. The CPU host codes were compiled with Microsoft

TABLE I
TEST RESULTS FOR EACH STEP OF THE IMPLEMENTATION.

Operations	Time (sec)
Load projections to main memory	68.75
Transfer projections to graphic memory	6.66
Compute $\bar{p}(\phi, v, w)$ and $\tilde{p}(\phi, v, w)$	6.32
Backprojection	**189.82**
Inverse Hilbert transform and interpolation	61.75
Transfer reconstructed image to main memory	3.10
Save image to hard disk	119.77
Total time	468.43 (7.8 min)

Visual Studio 9.0 running on a Dell Precision T7500 with two Intel Xeon X5570 running at 2.93 GHz and 24 GB of RAM. The tested data consists of 1440 projections, each of which is of size 1920×1536; the size of the reconstructed image is 1024^3. 32 bits float are used during the computation. TABLE I gives the test results.

V. SUMMARY

In this paper we have proposed a fast implementation of the BPF reconstruction algorithm for helical cone-beam CT, which can directly reconstruct the images of the horizontal slices of the inspected object. Our implementation is of high degree of parallelism and is suitable for acceleration on GPU and other parallel systems. Numerical results from both simulated data and real data have verified the correctness of our implementation and the reconstruction speed on GPU is satisfactory.

ACKNOWLEDGMENT

The authors would like to thank Associate Professor Xing Zhao for his help on GPU programming. We also thank laboratory technician Defeng Chen for acquiring real data for our numerical experiments.

REFERENCES

[1] H. Tuy, "An inversion formula for cone-beam reconstruction," *SIAM J. Appl. Math.*, no. 43, pp. 546-552, 1983.

[2] A. Katsevich, "Theoretically exact filtered backprojection-type inversion algorithm for spiral CT," SIAM J. Appl. Math., vol. 62, pp. 2012-2026, 2002.

[3] P. Grangeat, "Mathematical framework of cone-beam 3-D reconstruction via the first derivative of the Radon transform," in *Mathematical Methods in Tomography*, G. T. Herman, A. K. Louis, and F. Natterer, Eds. Berlin, Germany: Springer-Verlag, 1991, vol. 1497, Lecture Notes in Mathematics, pp. 66-97.

[4] B. D. Smith, "Image reconstruction from cone-beam projections: necessary and sufficient conditions and reconstruction methods," *IEEE Trans. Med. Imag.*, vol. MI-4, pp. 14-25, 1985.

[5] Y. Zou and X. Pan, "Exact image reconstruction on PI-lines from minimum data in helical cone-beam CT," *Phys. Med. Biol.*, vol. 49, no. 6, pp. 941-959, 2004.

[6] J. D. Pack, F. Noo, and R. Clackdoyle, "Cone-beam reconstruction using the backprojection of locally filtered projections," *IEEE Trans. Med. Imaging*, vol. 24, pp. 70-85, 2005.

[7] P. E. Danielsson, P. Edholm, J. Eriksson, and M. Magnusson Seger, "Toward exact reconstruction for helical cone-beam scanning of long objects. A new detector arrangement and a new completeness condition," in *Proc. 1997 Meeting Fully 3-D Image Reconstruction in Radiology and Nuclear Medicine*, D.W. Townsend and P. E. Kinahan, Eds., Pittsburgh, PA, 1997, pp. 141-144.

[8] G. Wang, Y. Ye, and H. Yu, "Appropriate and exact cone-beam reconstruction with standard and nonstandard spiral scanning," *Physics in Medicine and Biology*, vol. 52, no. 6, pp. 1-13, 2007.

[9] X. Pan, J. Siewerdsen, P. J. La Riviere, and W. A. Kalender, "Anniversary Paper: Development of x-ray computed tomography: The role of Medical Physics and AAPM from the 1970s to present," *Med. Phys.*, vol. 35, no. 8, pp. 3728-3739, 2008.

[10] S. G. Mikhlin, *Integral Equations and Their Applications to Certain Problems in Mechanics*, New York, Pergamon Press, 1957.

[11] F. Noo, R. Clackdoyle, and J. D. Pack, "A two-step Hilbert transform method for 2-D image reconstruction," *Physics in Medicine and Biology*, vol. 49, no. 17, pp. 3903-3923, 2004.

3D Forward and Back-Projection for X-Ray CT Using Separable Footprints with Trapezoid Functions

Yong Long and Jeffrey A. Fessler

***Abstract*—The greatest impediment to practical adoption of iterative methods for X-ray CT is the computation burden of cone-beam forward and back-projectors. Moreover, forward and back-projector accuracy is also crucial to iterative reconstruction methods. We previously described a computationally efficient projector that approximates the voxel footprint functions by the 2D separable products of trapezoid functions in the transaxial plane and rectangular functions in the axial direction [1], [2]. The separability of these footprint functions simplifies calculating their integrals over rectangular detector cells. We showed that this separable footprint (SF-TR) method was more accurate than the distance-driven (DD) method but with comparable computation time. This paper describes a new extension of that projector, called the SF-TT projector, that uses trapezoid functions in both directions. We show that using a trapezoid along the axial direction improves projector accuracy for voxels associated with larger cone angles. However, this improved accuracy requires increased computation compared to the rectangular approximation. Having both options available facilitates evaluation of the trade offs between accuracy and computation for different cone-beam geometries.**

***Index Terms*—Cone-beam tomography, iterative tomographic image reconstruction, forward and back-projection, separable footprint**

I. Introduction

Iterative statistical methods for 3D tomographic image reconstruction offer the potential for improved image quality and reduced X-ray dose, compared to conventional filtered back-projection (FBP) methods. The primary computational bottleneck in iterative reconstruction methods is forward and back-projection operations. The forward projection is roughly a discretized evaluation of the Radon transform, and the back-projector is its adjoint. Mathematically, an accurate forward projector must compute the convolution of the footprint of an image basis function with some detector blur model, such as a 2D rectangular function that represents the finite size of detector cells.

Numerous 3D forward and back-projection methods have been proposed [1]–[8]. Each method compromises between computational complexity and accuracy. Spherically symmetric basis functions (blobs) [5], [6] have radially symmetric footprints that conveniently are independent of the viewing angle, except for a magnification factor. However, when high accuracy is desired, blob footprints intersect many more detector cells than voxel footprints, increasing computation.

This work was supported in part by NIH grant P01-CA59827.

Y. Long and J. Fessler are with Dept. of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109.

The distance-driven (DD) method [4] maps the horizontal and vertical boundaries of the image voxels and detector cells onto a common plane such as xz or yz plane, approximating their shapes by rectangles. It calculates the lengths of overlap along the transaxial direction and along the axial direction, and then multiplies them to get the area of overlap. The DD method has the largest errors when the X-ray source's azimuthal angle is near odd multiples of $\pi/4$, where the transaxial footprint is approximately triangular rather than rectangular.

We proposed previously a separable footprint (SF-TR) projector [1] that approximates the voxel footprint functions as 2D separable functions with trapezoid and rectangle functions in the transaxial and axial directions respectively. The separability of these footprint functions greatly simplifies the calculation of their integrals over detector cells leading to an efficient implementation. The SF-TR method has similar computation speed as the DD projector, but is more accurate, reducing particularly the errors around odd multiples of $\pi/4$. The rectangle approximation in the axial direction is reasonable for smaller CT cone angles such as multi-slice detector geometries. However, for CT systems with larger cone angles ($> 10^\circ$), such as flat-panel detector geometries, the rectangle approximation becomes less accurate.

This paper describes a new separable footprint method for forward and back-projection called the SF-TT method. It approximates the voxel footprint functions using 2D separable functions with trapezoid functions in *both* the transaxial and axial directions. We show that the SF-TT projector is more accurate than the SF-TR projector, but requires more computation. To balance computation and accuracy, one may use the SF-TR projector for voxels associated with small cone angles (*i.e.*, near the plane of the X-ray source) where the rectangle approximation is adequate, and use the new SF-TT projector for voxels associated with larger cone angles.

The organization of this paper is as follows. Section 2 describes the cone-beam 3D system model, and introduces the SF-TT projector. Section 3 gives simulation results, including accuracy and speed comparison between the SF-TT and SF-TR projector as stand alone modules and within iterative image reconstruction. Finally, conclusions are in Section 5.

II. Method

A. Cone-Beam 3D System Model

For iterative image reconstruction, we forward project a discretized approximation of the continuous-space object $f(\vec{x})$

represented by a common basis function $\beta_0(\vec{\mathrm{x}})$ superimposed on a $N_1 \times N_2 \times N_3$ Cartesian grid as follows:

$$f(\vec{\mathrm{x}}) = \sum_{\vec{n}} f[\vec{n}]\ \beta_0\Big((\vec{\mathrm{x}} - \vec{c}[\vec{n}]) \oslash \vec{\Delta}\Big), \tag{1}$$

where the sum is over the $N_1 \times N_2 \times N_3$ lattice that is estimated and $\vec{c}[\vec{n}] = (\mathrm{c}_1[\vec{n}], \mathrm{c}_2[\vec{n}], \mathrm{c}_3[\vec{n}])$ denotes the center of the $\vec{n}$th basis function and $\vec{n} = (n_1, n_2, n_3) \in \mathbb{Z}^3$. The grid spacing is $\vec{\Delta} = (\Delta_1, \Delta_2, \Delta_3)$, and $\oslash$ denotes element-wise division. We consider the case $\Delta_1 = \pm\Delta_2$ hereafter, but we allow $\Delta_1 \neq \Delta_3$, because voxels are often not cubic.

Axial cone-beam projection space is characterized by three independent indices (s, t, β) and two distance parameters $(D_{\mathrm{sd}}, D_{\mathrm{s0}})$, where β denotes the angle of the source point counter-clockwise from the y axis, (s, t) denote the local coordinates on the 2D detector plane where the s-axis is perpendicular to the z-axis and the t-axis is parallel to the z-axis D_{sd} denotes the source to detector distance and D_{s0} denotes the source to rotation center distance.

The cone-beam projections of $f(\vec{\mathrm{x}})$ are given by

$$p(s, t; \beta) = \int_{\mathcal{L}(s,t,\beta)} f(x, y, z)\, \mathrm{d}\ell, \tag{2}$$

where

$$\mathcal{L}(s, t, \beta) = \left\{ \vec{p}_0 + \alpha \vec{e}_3 : \alpha \in \left[0, \sqrt{D_{\mathrm{sd}}^2 + s^2 + t^2}\right] \right\},$$

and $\vec{e}_3$ denotes the direction vector of a ray from the source position $\vec{p}_0$ to a point $\vec{p}_1$ on the detector plane.

Assume that the detector blur $h(s, t)$ is shift invariant, independent of β, and acts only along the s and t coordinates. Then the mean projections satisfy

$$\bar{y}_\beta[s_k, t_l] = \iint h(s_k - s, t_l - t)\, p(s, t; \beta)\, \mathrm{d}s\, \mathrm{d}t, \tag{3}$$

where (s_k, t_l) denotes the center of detector cell specified by indices (k, l) for $k = 0, \ldots, N_s - 1$ and $l = 0, \ldots, N_t - 1$.

Substituting the basis expansion model (1) for the object into (3) and using (2) leads to the linear model

$$\bar{y}_\beta[s_k, t_l] = \sum_{\vec{n}} a_\beta[s_k, t_l; \vec{n}]\, f[\vec{n}], \tag{4}$$

where the elements of cone-beam system matrix $\mathbf{A}$ are samples of the following cone-beam projection of a single basis function centered at $\vec{c}[\vec{n}]$:

$$a_\beta[s_k, t_l; \vec{n}] = F(s_k, t_l; \beta; \vec{n}), \tag{5}$$

where the "blurred footprint" function is

$$F(s, t; \beta; \vec{n}) \triangleq \iint h(s - s', t - t')\, q(s', t'; \beta; \vec{n})\, \mathrm{d}s'\, \mathrm{d}t',$$

and $q(s, t; \beta; \vec{n})$ denotes the cone-beam footprint of basis function $\beta_0\Big((\vec{\mathrm{x}} - \vec{c}[\vec{n}]) \oslash \vec{\Delta}\Big)$, *i.e.*,

$$q(s, t; \beta; \vec{n}) = \int_{\mathcal{L}(s,t,\beta)} \beta_0\Big((\vec{\mathrm{x}} - \vec{c}[\vec{n}]) \oslash \vec{\Delta}\Big)\, \mathrm{d}\ell\,. \tag{6}$$

The goal of forward projectors is to compute (4) rapidly but accurately.

A simple model for the detector blur is

$$h(s, t) = \frac{1}{r_s r_t} \operatorname{rect}\left(\frac{s}{r_s}\right) \operatorname{rect}\left(\frac{t}{r_t}\right), \tag{7}$$

where r_s and r_t denote the width along s and t respectively. This model accounts for the finite size of the detector elements.

B. Separable Footprint Projector with Trapezoid Functions (SF-TT)

The footprints of voxel basis functions can be computed analytically for cone-beam geometries [9, p. 104]. Fig. 1 shows an example of a true footprint and its profiles. This 2D function is approximately separable except for small areas at the upper left and lower right corner.

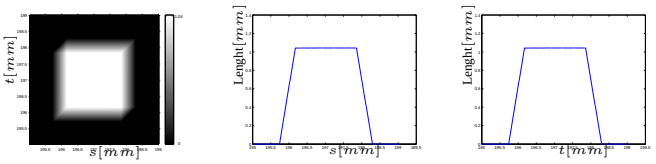


Fig. 1. The exact footprint function $q(s, t; \beta; \vec{n})$ and its profiles of a voxel with $\Delta_1 = \Delta_2 = \Delta_3 = 1$mm centered at $(93, 93, 93)$mm under a flat-detector cone-beam geometry with $D_{\mathrm{sd}} = 949$mm and $D_{\mathrm{s0}} = 541$mm when $\beta = 0°$. The azimuthal and polar angle of the ray connecting the source and the voxel center are $11.7°$ and $11.5°$ respectively.

Inspired by the shape of the true footprint, we approximate voxel footprints as 2D separable functions with trapezoid functions in *both* the transaxial and axial direction as follows,

$$\begin{aligned} q(s, t; \beta; \vec{n}) \quad &\approx \quad \triangle_{\mathrm{x}} l(\beta; \vec{n}) \operatorname{trap}(s; \tau_0, \tau_1, \tau_2, \tau_3) \\ &\quad \cdot \operatorname{trap}(t; \xi_0, \xi_1, \xi_2, \xi_3)\,, \end{aligned} \tag{8}$$

where

$$l(\beta; \vec{n}) \triangleq \frac{1}{|\cos(\theta_0)| \cdot \max\{|\cos(\varphi_0)|, |\sin(\varphi_0)|\}},$$

$$\operatorname{trap}(a; b_0, b_1, b_2, b_3) \triangleq \begin{cases} \frac{a - b_0}{b_1 - b_0}, & b_0 < a < b_1 \\ 1, & b_1 \leq a \leq b_2 \\ \frac{b_3 - a}{b_3 - b_2}, & b_2 < a < b_3 \\ 0, & \text{otherwise}, \end{cases} \tag{9}$$

where θ_0 and φ_0 denote the polar and azimuthal angles of the ray connecting the source and center of the $\vec{n}$th voxel respectively, τ_0, τ_1, τ_2 and τ_3 denote vertices of the trapezoid function which are at the exact locations as those of the true footprint function in the s direction, and ξ_0, ξ_1, ξ_2 and ξ_3 denote vertices of the trapezoid function in the t direction which are the projected t coordinates of four axial boundaries of the voxel.

Using the projections of boundaries of the voxel basis function as the boundaries of the approximate separable footprints ensures the depth-dependent magnification of the cone-beam geometry is modeled accurately. It also allows the approximated separable footprints to adapt their shapes according to relative positions of the source, detector and voxels, as true footprints do. For example, for a voxel centered at the origin, its axial footprint is approximately a rectangular function as

shown in [1, Fig. 1], instead of a trapezoid function. For this voxel $\mathrm{trap}(t;\xi_0,\xi_1,\xi_2,\xi_3)$ is almost a rectangle because $\xi_0 \approx \xi_1$ and $\xi_2 \approx \xi_3$ because ξ_0, ξ_1, ξ_2 and ξ_3 are the projected t coordinates of four axial boundaries of the voxel.

To accelerate the computation of the SF-TT projector, we adopt the acceleration method used by the SF-TR projector [1]. We initially ignore $l(\beta;\vec{n})$ by setting $l(\beta;\vec{n}) = 1$ for all the voxels at any projection view, and then scale the "blurred footprint" functions by multiplying them by a ray-dependent scale factor.

III. Result

A. Forward and Back-Projector as Single Modules

We simulated an axial cone-beam flat-detector X-ray CT system with a detector size of $N_s \times N_t = 512 \times 512$ cells spaced by $\Delta_{\mathrm{s}} = \Delta_{\mathrm{t}} = 1$mm with $N_\beta = 984$ angles over $360°$. The source to detector distance D_{sd} is 949mm, and the source to rotation center distance D_{s0} is 541mm. We included a rectangular detector response (7) with $r_s = \Delta_{\mathrm{s}}$ and $r_t = \Delta_{\mathrm{t}}$.

We implemented the SF-TR and SF-TT projector in ANSI C using single precision. The DD projector was provided by De Man *et al.*, also implemented in ANSI C.

1) Maximum Errors of Forward Projectors: We define the maximum error as

$$e(\beta;\vec{n}) = \max_{s,t\in\mathbb{R}} \left| F(s,t;\beta;\vec{n}) - F_{ap}(s,t;\beta;\vec{n}) \right|, \qquad (10)$$

where $F_{ap}(s,t;\beta;\vec{n})$ is any of the approximate blurred footprints by the SF-TR, SF-TT and DD methods. We generated the true footprint $F(s,t;\beta;\vec{n})$ in (5) by linearly averaging 1000×1000 analytical line integrals of rays sampled over each detector cell.

We compared the maximum errors of the forward SF-TR, SF-TT and DD projectors for a voxel with $\triangle_{\mathrm{x}} = \triangle_{\mathrm{y}} = \triangle_{\mathrm{z}} = 1$ mm centered at $(0, 0, -100)$ mm. Since the voxel is centered at the origins of x and y axes, we choose $N_\beta = 180$ angles over only $90°$ rotation. Fig. 2 shows the results on a logarithmic scale. The maximum errors of the SF-TT projector are smaller than those of the SF-TR and DD projector, *e.g.*, the maximum errors of the DD and SF-TR projector are about 18 and 3 times larger than that of the SF-TT projector when $\beta = 45°$. We also compared the maximum errors for a voxel centered at $(100, 150, -100)$ mm. We choose $N_\beta = 720$ angles over $360°$ rotation due to the offsets of this voxel in the x and y direction. Fig. 3 shows the results. The maximum errors over $360°$ rotation of the DD and SF-TR projector are about 13 and 3 times of that of the SF-TT projector.

2) Speed of Forward and Back-projectors: We compared the computation times of the SF-TR, SF-TT and DD projectors using an image whose size is $N_1 = 512, N_2 = 512, N_3 = 128$ and spacing is $\Delta_1 = \Delta_2 = \Delta_3 = 0.5$ mm. We evaluated the elapsed time using the average of 5 projector runs on a 8-core Sun Fire X2270 server with 2.66 GHz Xeon X5500 processors. Because of the "hyperthreading" of these Nehalem cores, we used 16 POSIX threads to parallelize the projection operation across views. (We found that using 16 threads reduced computation time by only about 10% compared to using 8 threads.)

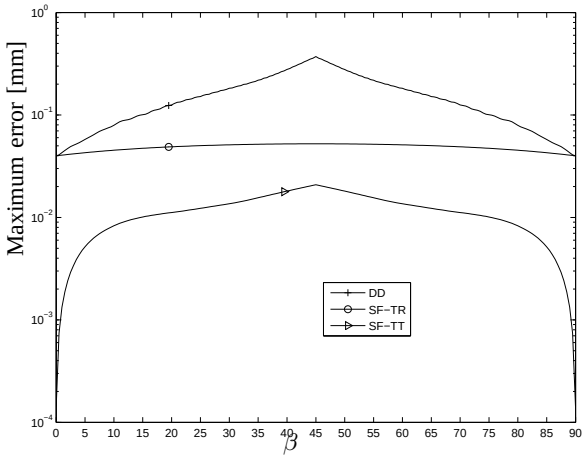


Fig. 2. Maximum error comparison between the forward DD, SF-TR and SF-TT projector for a voxel centered at $(0, 0, -100)$mm.

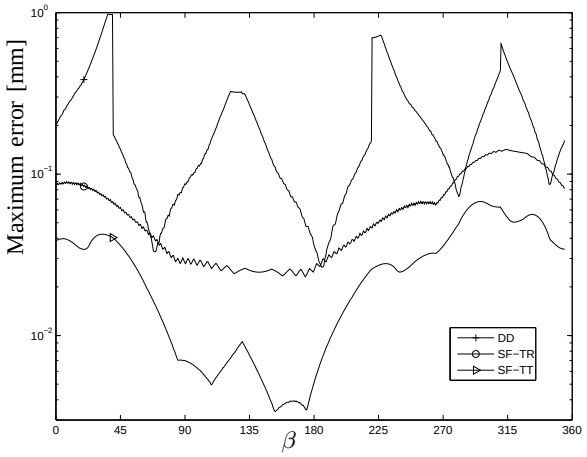


Fig. 3. Maximum error comparison between the forward DD, SF-TR and SF-TT projector for a voxel centered at $(100, 150, -100)$mm.

Table I summarizes computation times. The computation times of the SF-TR and DD projector are about the same, whereas the SF-TT projector is about 2 times slower. Of course execution times depend on code implementation.

B. Forward and Back-projectors within Iterative Reconstruction

We also compare the SF-TT projector and the SF-TR projector within an iterative reconstruction method. (We already showed the SF-TR method provides less artifacts in the reconstructed images than the DD method in [1]).

We simulated a X-ray axial cone-beam CT system with a flat-panel detector of 512 detector channels for 512 slices

Projectors	SF-TT	SF-TR	DD
Forward time (seconds)	91	35	46
Backward time (seconds)	92	44	49

TABLE I
SPEED COMPARISON OF THE SF-TT, SF-TR AND DD FORWARD AND BACK PROJECTORS.

($N_s = 512, N_t = 512$) by $N_\beta = 984$ views over $360°$. The size of each detector cell is $\Delta_{\mathrm{s}} \times \Delta_{\mathrm{T}} = 1 \times 1$ mm^2. The source to detector distance is $D_{\mathrm{sd}} = 949.075$ mm, and the source to rotation center distance is $D_{\mathrm{s0}} = 541$ mm. A quarter detector offset in the s direction is included to reduce aliasing.

We modified the 3D Shepp-Logan digital phantom to include several ellipsoids centered at the $z = 112.5$ plane because the trapezoid approximation of the SF-TT method is more realistic than the rectangle approximation of the SF-TR method especially for voxels far away from the origin. The field of view (FOV) is $250 \times 250 \times 250$ mm^3, implying $256 \times 256 \times 256$ voxels with a resolution of $0.9766 \times 0.9766 \times 0.9766$ mm^3. We simulated noiseless cone-beam projection measurements from the Shepp-Logan phantom by linearly averaging 8×8 analytical rays [9, p. 104] sampled across each detector cell. To focus on the projector accuracy, we used noiseless projection data.

We implemented iterative image reconstruction with these two projector/backprojector methods. We ran 50 iterations of the ordered subsets method with 82 subsets [10], initialized with reconstruction by the FDK method [11], for the following penalized weighted least-squares cost function with an edge-preserving "hyperbola" penalty function (PWLS-OS):

$$\Phi(\boldsymbol{x}) = \sum_i w_i \frac{1}{2}(y_i - [\mathbf{A}\boldsymbol{x}]_i)^2 + \beta R(\boldsymbol{x}) \qquad (11)$$

$$R(\boldsymbol{x}) = \sum_k \psi([\mathbf{C}\boldsymbol{x}]_k), \qquad (12)$$

where y_i is the negative $\log$ of the measured cone-beam projection, w_i values are statistical weighting factors, $\mathbf{A}$ is the system matrix, $\mathbf{C}$ is a finite differencing matrix and $\psi(t)$ is the potential function. Here we used the hyperbola:

$$\psi(t) = \frac{\delta^2}{3}\left(\sqrt{1 + 3\,(t/\delta)^2} - 1\right). \qquad (13)$$

For this simulation, we used $w_i = \exp(-[\mathbf{A}\boldsymbol{x}]_i)$, $\beta = 2$ and $\delta = 5$ Hounsfield units (HU).

For this iterative reconstruction experiment, we did not see obvious visual differences between reconstructions by the SF-TT and SF-TR method, and the normalized root-mean-square (NRMS) errors were similar. It appeared that the axial cone-beam artifacts due to poor sampling (not truncation) at the off-axis slices dominated other effects in the reconstructed images, such as the errors caused by rectangle approximation. Further research will evaluate these two projectors within iterative reconstruction methods under other CT geometries where the off-axis sampling is better, such as helical scans, yet where the cone angle is large enough to differentiate the SF-TT and SF-TR method .

IV. Conclusion

We have presented a 3D forward and back projector, named the SF-TT projector for X-ray CT. Our results have shown that the SF-TT projector is more accurate but computationally slower than the SF-TR projector. We demonstrated previously that the SF-TR projector is more accurate than the well-known DD projector but with similar computation speed in [1].

The SF-TT projector uses trapezoid functions in both the transaxial and axial directions. Using trapezoid functions in the axial direction involves more computation compared with using simple rectangular functions, such as projecting four axial boundaries of each voxel instead of two and evaluating the weight (contribution of a voxel to a detector cell) in two additional triangle areas of each trapezoid function. Thus it is reasonable that the computation time of the SF-TT projector was about 2 times that of the SF-TR projector. To save computation and maintain relative accuracy, one may use the original SF-TR projector for voxels that are near the X-ray source plane, where the cone angles are small and the rectangle approximation is reasonable, and use the SF-TT projector for other voxels.

Acknowledgment

We thank Dr. James Balter for ongoing discussions about cone-beam CT and GE for the use of their DD projector/backprojector code.

References

[1] Y. Long, J. A. Fessler, and J. M. Balter, "A 3D forward and back-projection method for X-ray CT using separable footprint," in *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, 2009, pp. 146–9, winner of poster award.

[2] ——, "3D forward and back-projection for X-ray CT using separable footprints," *IEEE Trans. Med. Imag.*, 2009, submitted as TMI-2009-0397.

[3] R. L. Siddon, "Fast calculation of the exact radiological path for a three-dimensional CT array," *Med. Phys.*, vol. 12, no. 2, pp. 252–5, Mar. 1985.

[4] B. De Man and S. Basu, "Distance-driven projection and backprojection in three dimensions," *Phys. Med. Biol.*, vol. 49, no. 11, pp. 2463–75, Jun. 2004.

[5] A. Ziegler, T. Köhler, T. Nielsen, and R. Proksa, "Efficient projection and backprojection scheme for spherically symmetric basis functions in divergent beam geometry," *Med. Phys.*, vol. 33, no. 12, pp. 4653–63, Dec. 2006.

[6] S. Matej and R. M. Lewitt, "Practical considerations for 3-D image reconstruction using spherically symmetric volume elements," *IEEE Trans. Med. Imag.*, vol. 15, no. 1, pp. 68–78, Feb. 1996.

[7] C. Riddell and Y. Trousset, "Rectification for cone-beam projection and backprojection," *IEEE Trans. Med. Imag.*, vol. 25, no. 7, pp. 950–62, Jul. 2006.

[8] C. Schretter, "A fast tube of response ray-tracer," *Med. Phys.*, vol. 33, no. 12, pp. 4744–8, Dec. 2006.

[9] A. C. Kak and M. Slaney, *Principles of computerized tomographic imaging*. New York: IEEE Press, 1988.

[10] H. Erdoğan and J. A. Fessler, "Ordered subsets algorithms for transmission tomography," *Phys. Med. Biol.*, vol. 44, no. 11, pp. 2835–51, Nov. 1999.

[11] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone beam algorithm," *J. Opt. Soc. Am. A*, vol. 1, no. 6, pp. 612–9, Jun. 1984.

Tomographic Reconstruction Through View-based Implementation of Least-Square algorithms

Cyril Riddell

GE Healthcare, 283 rue de la Minière, 78533 Buc cedex

I. Introduction

Iterative reconstruction of tomographic problems is a demanding computation task. According to the system geometry, the size of the data, or the interpolation model, the requirements in terms of computation power and memory are not the same. With C-arm based X-ray rotational angiography, the data is generally truncated, angular sampling is limited by the detector read-out rate, and the circular trajectory induces so-called cone-beam artifacts. Regions-of-interest are reconstructed using data extrapolation and analytical reconstruction. Iterative least-square technique may be considered as an alternative to data extrapolation as well as for correcting for the cone-beam artifacts. This implies a significant increase in computation. It is common to trade memory versus computation. For instance, the use of look-up tables where pre-calculated values are stored [1], or the use of over-sampling with nearest neighbor to replace linear interpolation [2], reduce computation while increasing memory needs. Here, we look at the opposite direction, where we decrease memory needs, even if it increases computation. If we consider the Landweber iteration [3-4], it has a slow convergence. The conjugate gradient has a much faster convergence [5], but requires several volumes to be kept in memory at once. In addition, the full volume must be considered in an iterative reconstruction process, not just the region-of-interest. In [6], it is noted that the Landweber iteration can be expressed in a way such that the iterates are views rather than volumes. This allowed for *approximating* the iterative algorithm by a filtering step, thus removing the computational burden.

In the present work, we further explore this change of perspective and provide expressions for the conjugate gradient and the SART [7] algorithms, where the iterates are views rather than volumes. In the common case of under-determined problem, where the volume size is significantly larger than the data size, a view-based scheme requires proportionally less memory for about the same amount of computation. However, it allows for reconstructing arbitrary large volumes since volumes are never stored entirely. This approach suits specialized hardware such as GPU, where the computation power is much higher than CPU-based system, while memory is comparatively scarce.

II. Theory

A. Image-based Landweber iteration

We consider the resolution of a linear system $Rf = s$ through the use of the application A defined as:

$$A[f] = (I - \rho R^t R) f + \rho R^t s \qquad (1)$$

If we set ρ s.t. $\|I - \rho R^t R\| < 1$, application A is a contraction, we have:

$$\lim_{n\to\infty} A^n[f^{(0)}] = f^* = \min_f \|Rf - s\|^2 \qquad (2)$$

The repeated use of the contraction defines an algorithm that converges towards the least square approximation of the solution of the system. In the context of tomography, this algorithm can be referred to as the Landweber iteration [3]. It computes image f from the view data s.

B. View-based Landweber iteration

We replace the image-based application A with the view-based application V defined for p, vector of the space of views as:

$$V[p] = (I - \rho R R^t) p + \rho s \qquad (3)$$

We note that if $f = R^t p$:

$$\begin{aligned} R^t(V[p]) &= R^t(I - \rho R R^t) p + \rho R^t s \\ &= (I - \rho R^t R) f + \rho R^t s = A[f] \end{aligned} \qquad (4)$$

Therefore, by simple recurrence, if $f^{(0)} = R^t p^{(0)}$ both the image-domain and the view-domain schemes will produce the same iterates $f^{(n)} = R^t p^{(n)}$.

This approach has been used in [6] to approximate the iterative reconstruction by a filtering step: the full iterative computation is performed on basis objects from which a filter is designed such that approximate reconstruction of any object is obtained from this filtering. Here we look at an exact computation for any object. Both schemes are equivalent because they compute products $f = R^t p$ (i.e. backprojection) and $p = Rf$ (i.e. reprojection) in the same order. Implementation of the view-domain algorithm relies on the computation of $RR^t p$. If it is implemented as the computation of $R^t p$, which is the back-projection of all views into one volume, followed by the computation of $R(R^t p)$, which is the reprojection of the volume resulting from the backprojection, there is neither difference nor gain to expect from the view-domain approach. Gain is expected if one considers the product matrix RR^t and applies it in the view domain as $(RR^t)p$.

For so doing, we denote $(R)_{ij} = r_{ij}$ the coefficient of matrix R that relates image pixel j to view measurement i. The coefficient that relates measurement j to measurement j' through RR^t is:

$$(RR^t)_{j'j} = \sum_i r_{ij'} r_{ij} \qquad (5)$$

which is the sum over all pixels of the combined interactions between pixel i and measurements j and j'. Memory requirements are now independent of the sampling in i, while, on the contrary, computation is proportional to i.

The final volume after N iterations is obtained computing: $f^{(N)} = R^t s^{(N)}$. This step may be delayed to visualization time, when only part of the volume is under study, or concern the region-of-interest only, so that no volume is entirely stored at any time.

C. View-based conjugate-gradient:

The contraction-based algorithm provides a solution to a quadratic optimization problem that can be reached in fewer iterations using the conjugate gradient algorithm. We verify here that its implementation can be set indifferently in the view or the image domain.

The image-based algorithm may be implemented as follows:

$$\begin{cases} f^{(0)} = 0, \quad r^{(0)} = d^{(0)} = -R^t s \\ f^{(n)}, r^{(n)}, d^{(n)}, \text{compute } R^t R d^{(n)} \\ \alpha^{(n)} = -\left\langle r^{(n)}, d^{(n)} \right\rangle \Big/ \left\langle R^t R d^{(n)}, d^{(n)} \right\rangle \\ f^{(n+1)} = f^{(n)} + \alpha^{(n)} d^{(n)} \\ r^{(n+1)} = r^{(n)} + \alpha^{(n)} R^t R d^{(n)} \\ \beta^{(n)} = \left\langle r^{(n+1)}, r^{(n+1)} \right\rangle \Big/ \left\langle r^{(n)}, r^{(n)} \right\rangle \\ d^{(n+1)} = r^{(n+1)} + \beta^{(n)} d^{(n)} \end{cases} \tag{6}$$

which requires the storage in memory of 4 images at once: $f^{(n)}, r^{(n)}, d^{(n)}, R^t R d^{(n)}$.

The view-based implementation has intermediate steps that are kept in sets of views rather than images, using the following change in variables:

$$\begin{cases} p^{(n)} \quad s.t. \quad f^{(n)} = R^t p^{(n)} \\ q^{(n)} \quad s.t. \quad r^{(n)} = R^t q^{(n)} \\ t^{(n)} \quad s.t. \quad d^{(n)} = R^t t^{(n)} \end{cases} \tag{7}$$

We obtain

$$\begin{aligned} \alpha^{(n)} &= -\left\langle r^{(n)}, d^{(n)} \right\rangle \Big/ \left\langle R^t R d^{(n)}, d^{(n)} \right\rangle \\ &= -\left\langle R^t q^{(n)}, R^t t^{(n)} \right\rangle \Big/ \left\langle R^t R R^t t^{(n)}, R^t t^{(n)} \right\rangle \\ &= -\left\langle q^{(n)}, R R^t t^{(n)} \right\rangle \Big/ \left\langle R R^t t^{(n)}, R R^t t^{(n)} \right\rangle \end{aligned} \tag{8}$$

and

$$\begin{aligned} \beta^{(n)} &= \left\langle r^{(n+1)}, r^{(n+1)} \right\rangle \Big/ \left\langle r^{(n)}, r^{(n)} \right\rangle \\ &= \left\langle R^t q^{(n+1)}, R^t q^{(n+1)} \right\rangle \Big/ \left\langle R^t q^{(n)}, R^t q^{(n)} \right\rangle \\ &= \left\langle q^{(n+1)}, R R^t q^{(n+1)} \right\rangle \Big/ \left\langle q^{(n)}, R R^t q^{(n)} \right\rangle \end{aligned} \tag{9}$$

which, given $f^{(n)} = R^t p^{(n)}, r^{(n)} = R^t q^{(n)}, d^{(n)} = R^t t^{(n)}$ and the computation of $RR^t t^{(n)}, RR^t q^{(n)}$, leads to

$$\begin{cases} q^{(n)}, RR^t t^{(n)} \Rightarrow \alpha^{(n)} \\ p^{(n+1)} = p^{(n)} + \alpha^{(n)} t^{(n)} \Rightarrow f^{(n+1)} = f^{(n)} + \alpha^{(n)} d^{(n)} \\ q^{(n+1)} = q^{(n)} + \alpha^{(n)} RR^t t^{(n)} \Rightarrow r^{(n+1)} = r^{(n)} + \alpha^{(n)} R^t R d^{(n)} \\ RR^t q^{(n+1)} \Rightarrow \beta^{(n)} \\ t^{(n+1)} = q^{(n+1)} + \beta^{(n)} t^{(n)} \Rightarrow d^{(n+1)} = r^{(n+1)} + \beta^{(n)} d^{(n)} \\ RR^t t^{(n+1)} = RR^t q^{(n+1)} + \beta^{(n)} RR^t t^{(n)} \Rightarrow \alpha^{(n+1)} \end{cases} \tag{10}$$

We have

$$\begin{cases} p^{(0)} = 0 \Rightarrow f^{(0)} = 0 \\ q^{(0)} = t^{(0)} = -s \Rightarrow r^{(0)} = d^{(0)} = -R^t s, \\ \text{compute } RR^t t^{(0)} \Rightarrow \alpha^{(0)}, RR^t q^{(0)} = RR^t t^{(0)} \\ p^{(1)} = p^{(0)} + \alpha^{(0)} t^{(0)} \Rightarrow f^{(1)} = f^{(0)} + \alpha^{(0)} d^{(0)} \\ q^{(1)} = q^{(0)} + \alpha^{(0)} RR^t t^{(0)} \Rightarrow r^{(1)} = r^{(0)} + \alpha^{(0)} R^t R d^{(0)} \\ \text{compute } RR^t q^{(1)} \Rightarrow \beta^{(0)} \\ t^{(1)} = q^{(1)} + \beta^{(0)} t^{(0)} \Rightarrow d^{(1)} = r^{(1)} + \beta^{(0)} d^{(0)} \\ RR^t t^{(1)} = RR^t q^{(1)} + \beta^{(1)} RR^t t^{(1)} \Rightarrow \alpha^{(1)} \end{cases} \tag{11}$$

Therefore, by recurrence, both schemes are mathematically equivalent. It is therefore possible to replace the 4 intermediate vectors $f^{(n)}, r^{(n)}, d^{(n)}, R^t R d^{(n)}$ in the image space with 5 intermediate vectors in the view space: $s^{(n)}, q^{(n)}, r^{(n)}, RR^t q^{(n)}, RR^t t^{(n)}$. No extra-computation is added inside the loop, but an extra vector is required for storage.

D. View-based SART:

In the block-iterative approach, application A is replaced by a series of application A_Θ derived from the partition of matrix R and vector s into blocks of rows R_Θ and sub-vector s_Θ. Θ therefore denotes a set of row indices, and application A_Θ is such that

$$A_\Theta[f] = \left(I - \rho R_\Theta^t R_\Theta\right) f + \rho R_\Theta^t s_\Theta \tag{12}$$

Each block of the partition is applied successively so that all measurements are used once per iteration, leading to the product application:

$$A_\Pi[f] = \left(\prod_\Theta A_\Theta \right)[f] \tag{13}$$

Convergence of the series $\lim_{n\to\infty} A_\Pi^n \left[f^{(0)} \right] = f^\Pi$ has been proven in [8]. For tomography, a few iterations give satisfactory results.

We now denote p a vector in the space of views.

We denote Φ the set of indices that complement Θ. We can write:

$$R = \begin{bmatrix} R_\Theta \\ R_\Phi \end{bmatrix}, \quad p = \begin{bmatrix} p_\Theta \\ p_\Phi \end{bmatrix} \quad s = \begin{bmatrix} s_\Theta \\ s_\Phi \end{bmatrix} \tag{14}$$

We note that $R^t s = R_\Theta^t s_\Theta + R_\Phi^t s_\Phi$ and that $R_\Theta^t s = R_\Theta^t s_\Theta$ and $R_\Phi^t s = R_\Phi^t s_\Phi$.

We now replace the image-based application A_Θ with the view-based application:

$$V_\Theta[p] = \begin{bmatrix} (I - \rho R_\Theta R^t)p + \rho s_\Theta \\ p_\Phi \end{bmatrix} \qquad (15)$$

that modifies only the part of the measurements indexed by Θ.

We now show that if $f = R^t p = R^t_\Theta p_\Theta + R^t_\Phi p_\Phi$, we have

$$\begin{aligned} R^t(V[p]) &= R^t_\Theta\left((I - \rho R_\Theta R^t)p + \rho s_\Theta\right) + R^t_\Phi p_\Phi \\ &= R^t_\Theta p + R^t_\Phi p_\Phi - \rho R^t_\Theta R_\Theta R^t p + \rho R^t_\Theta s_\Theta \\ &= f - \rho R^t_\Theta R_\Theta f + \rho R^t_\Theta s_\Theta = A_\Theta[f] \end{aligned} \qquad (16)$$

Therefore, by simple recurrence, if $f^{(0)} = R^t p^{(0)}$ both the image-domain and the view-domain schemes will produce the same iterates $f^{(n)} = R^t p^{(n)}$.

III. Experiments

As an illustration, a simple simulation of a parallel geometry was performed. A 256^2 image of a skull phantom was obtained with pixel values scaled to shifted Hounsfied Unit (HU), i.e. with 0 for air, 1000 for water, leading to a maximum bone value of 2477. This image was projected using parallel-beam geometry with 128 projection angles and 256 detector bins. The size of the data was therefore 50% of the size of the image to be reconstructed. This data set was reconstructed with the Landweber algorithm. The image-based algorithm is denoted ibL, the view-based version vbL. Fifty iterations were computed using $\rho = 9.10^{-3}$ for both versions (a value of $\rho = 1.10^{-2}$ did not provide converge). The same data set was reconstructed with the conjugate gradient algorithm. We denote ibCG the image-based version, and vbCG the view-based version. Twenty iterations were performed for each version. Image-based results were compared to view-based images by subtraction of the generated images.

IV. Results

Fig.1 shows the images obtained with the image-based (left) and view-based (middle) algorithms. The top row shows the results obtained with ibL (top left) and vbL (top center), the bottom row the reconstruction of the same data set with ibCG (bottom left) and vbCG (bottom center). The right column shows the difference between image-based and view based implementations (top ibL – vbL, bottom ibCG – ivCG). All images are scaled to their own maximum. For the Landweber iteration, the greatest error in the difference image ibL – vbL was -0.023HU. For the conjugate gradient, after 20 iterations, the error was up to 10HU. The pattern of the difference image shows that these high error values were located at edges. The qualitative appearance of the reconstructed image does not demonstrate difference of resolution. Indeed, fig. 2 shows a profile taken at the centre row of the conjugate gradient reconstructed images: errors are hardly noticeable. Similar results were obtained with image-based and view-based SART (not shown).

V. Discussion

View-based implementations of least-square algorithms have been proposed as a mean for reducing the memory requirements when computing tomographic reconstruction from limited data sets. From a practical perspective, application to rotational angiography is of interest. Typically, 512^3 volumes are reconstructed from 150 truncated views on a circular orbit. Potential improvements are expected from using iterative reconstruction in order to better handle truncation by avoiding the need for data extrapolation, as well as a better handling of cone-beam artifacts due to the incomplete circular orbit. Such a task would benefit from the use of powerful graphic cards, but the need for storing four volumes of 512^3 32-bit floating values would pre-empt 2 Gb of memory. The proposed scheme with 150 views would "only" require 750 Mb. Furthermore, if the volume is a region-of-interest, bigger volume and associated memory are required, while the view-based scheme requirements remain constant. Although the provided numerical simulation only cover parallel beam noise-free data, we see no reason why the algorithm should not provide equivalent solution (whatever they are) when applied to cone-beam acquisition from a real system.

On a more theoretical point of view, view-based implementations allow for reconstructing volumes with arbitrary small voxels, and thus explore the cost associated with the limited sampling of the volume.

VI. Conclusion

Extensions of the view-based expression for the Landweber iteration have been proposed for the conjugate-gradient and the SART algorithm. They are equivalent to their image-based counterparts, but in the case under-determined problems, they allow for storing intermediate results in the view domain, therefore using less memory storage. This can lead to significant memory savings when considering iterative reconstruction of under-sampled tomographic problems.

VII. References

[1] S. Steckmann, M. Knaup, M. Kachelrieß. "High performance cone-beam spiral backprojection with voxel-specific weighting." *Phys. Med. Biol.* 21;54(12):3691-3708, 2009

[2]C. Riddell and Y. Trousset, "Rectification for cone-beam projection and backprojection." *IEEE Trans. Med. Imaging*. vol. 25, pp.950-962, 2006

[3] L. Landweber "An iterative formula for Fredholm integral equations of the first kind," A*mer. J. Math.,* vol. 73, pp. 615-624, 1951

[4] Y. Censor, "Finite series-expansion reconstruction methods," *Proc. IEEE*, vol. 71, no. 3, pp. 409–419, Mar. 1983.

[5] P. Ciarlet, *Introduction to numerical linear algebra and optimisation,* Cambridge Texts In Applied Mathematics, 1989

[6] H. Kunze, W. Härer, J. Orman, T. Mertelmeier, K. Stierstorfer, "Filter determination for Tomosynthesis aided by iterative reconstruction techniques," *Conf. Proc. of the 9th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, *Lindau*, pp 309-312, 2007

[7] A. Andersen and A. Kak, "Simultaneous Algebraic Reconstruction Technique (SART): A Superior Implementation of ART," *Ultrasonic Imaging*, 1984

[8] M. Jiang and G. Wang "Convergence of the simultaneous algebraic reconstruction technique (SART)," *IEEE Trans. Image Proc.*, vol. 12, pp. 957–961, 2003.

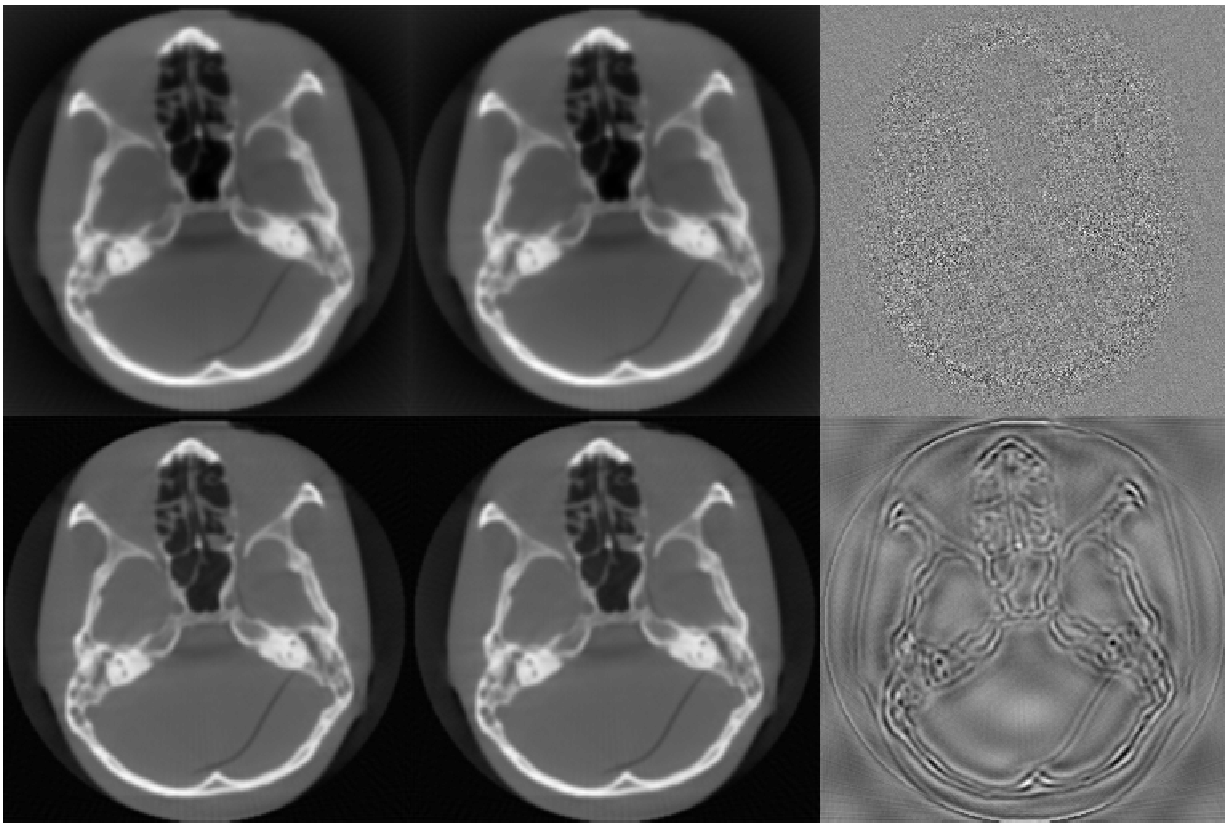


Fig 1: Top row: Landweber, 50 iterations. Bottom row: conjugate gradient, 20 iterations. Left column: image-based implementation. Center column: view-based implementation. Right column: difference image between image-based and view-based implementation. All images are scaled to their own maximum.

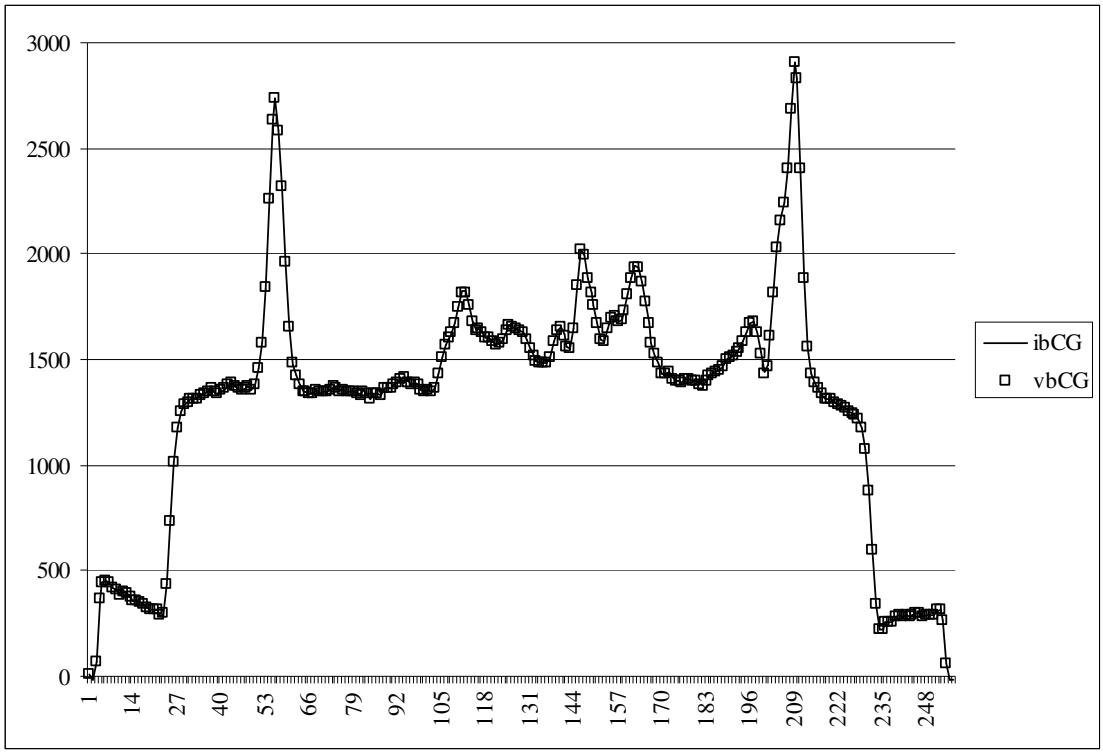


Fig. 2: Profile through the center row of the image-based conjugate gradient (line) and the view-based conjugate gradient (symbols).

Comparison of Sinogram- and Image-Domain Penalized-Likelihood Reconstruction for CT

Phillip A. Vargas* and Patrick J. La Rivière

I. Introduction

Iterative reconstruction methods can provide superior spatial image resolution and noise properties relative to analytic approaches. This is achieved by correcting for geometric effects of the imaging system, engineering artifacts and photon counting statistics. These effects cannot be readily modeled in linear analytic reconstruction algorithms such as filtered back projection (FBP). The main disadvantage of these iterative methods is that they can be computationally intensive since they involve extensive back- and forward-projections. To mitigate this burden, our group has worked on striking a balance between fully iterative image reconstruction and naive FBP. The method we have chosen to accomplish this reconstructs a smoothed and restored sinogram with conventional FBP. In this method the sinogram is smoothed by maximizing a penalized-likelihood objective function that can incorporate many (though not all) of the effects included in fully iterative image-domain reconstruction. By only iterating a few times we reduce a great deal of computation cost. This cost is further lowered by performing these calculations in the sinogram space where we avoid having to perform the back- and forward-projections.

In a previous report [1] we showed that under certain conditions the image-domain and sinogram-domain penalized-likelihood approaches are, in fact, precisely equivalent. These conditions were that

1) The objective function is convex.

2) The degradations can be modeled in the sinogram domain.

3) The geometric projection matrix linking the two domains is invertible and the inverse matrix is available.

The last condition is obviously the most restrictive. More commonly, one has an accurate but approximate inverse available, such as that represented by a discretized analytic algorithm such as FBP. In the previous work, we demonstrated provisionally that even when using such an approximate inverse, one obtained very similar resolution-variance trade-offs from both the sinogram- and image-domain methods. This held for both linear (emission tomography) and non-linear (transmission tomography) reconstruction algorithms.

In this work, we provide a much more thorough comparison between the two approaches. Building on the work of Fessler for image-domain methods [2], we derive analytic expressions for the resolution and variance properties of images reconstructed using the sinogram-domain PL strategy. We compare the results obtained using these expressions to those obtained using brute-force Monte Carlo simulations and find they agree extremely closely. We also compare the sinogram-domain and image-domain methods for a variety of situations, including overdetermined and under-determined system matrices. The sinogram-domain and image-domain approaches continue to produce very similar resolution-variance trade-offs for all of the cases explored. These validated analytic expressions provide a tool for further optimizing the sinogram-domain approach.

This work was supported by NIH Grant RO1-134680CA. *Asterisk indicates corresponding author.*

P. A. Vargas is with the Department of Radiology, The University of Chicago, Chicago, IL 60637 USA (e-mail: vargasp@uchicago.edu).

P. J. La Rivière is with the Department of Radiology, The University of Chicago, Chicago, IL 60637 USA.

II. Methods

A. Measurement model

Let $Y = [Y_1, .., Y_N]'$ denote a random measurement vector (e.g. a noisy sinogram), where N is the total number of measurements in the scan. The $'$ denotes a vector transpose. The composite index i will be used to specify the projection angle and the the location of a particular detector element in the array. Each Y_i measurement corresponds to the measurement of photons passing through the patient along the ith measurement line, which we denote l_i. This measurement line can be represented by the simple discretized line integral

$$l_i = \sum_{j=1}^{M} a_{ij}\theta_j \tag{1}$$

where $\theta = [\theta_1, .., \theta_M]'$ is the discretized linear attenuation map of the object being imaged and a_{ij} represents a weighting element of the the system matrix $\mathbf{A}$. The transmitted number of X-rays Φ_{ij} at an average energy $\bar{E}$ emerging unabsorbed and unscattered from the patient along this line is given by the Beer-Lambert law

$$\Phi_i = I_i e^{-\sum_{j=1}^{M} a_{ij}\theta_j}. \tag{2}$$

The photon fluence Φ_i plus a number s_i scattered photons, are detected in numbers assumed to be governed by Poisson statistics. Each detected photon contributes to the measured detector signal in proportion to its energy E, and with a constant of proportionality by G_i. The detected signal is read out through detector electronics having dark current d_i and electronic readout noise assumed to be normally distributed with variance σ^2. Overall, then, we assume that each y_i is a realization of a random variable Y_i. The energy-weighted combination of Poisson random variables results in compound Poisson statistics. We have shown in previous studies that this polychromatic model can be mapped to a monochromatic model with simple Poisson statistics [3]. In this study our primary focus is noise and will assume the photons are monochromatic and at the average energy which leads to a simple Poisson model. We, also, assume that I_i, G_i, d_i, σ^2 and the average energy of the incident beam are known from calibration scans and dark-current measurements, and the scatter can be estimated leading to the equation

$$Y_i = \text{Poisson}\left\{\Phi_i\right\} + \text{Normal}\left\{d_i, \sigma_i^2\right\}. \tag{3}$$

B. Image-Domain Objective function

In fully iterative statistical image reconstruction from transmission tomography measurements, one seeks to estimate an appropriately discretized version of the attenuation map θ. In this paper this is accomplished by maximizing the penalized-likelihood objective function

$$\hat{\theta}^{PL-I} = \arg\max_{\theta>0} \left[\mathcal{L}(\theta, \mathbf{y}) - \beta R(\theta)\right] \tag{4}$$

where $\mathcal{L}(\theta, \mathbf{y})$ is the Poisson log-likelihood function, $R(\theta)$ is a penalty function, and β is a regularization parameter. For the remainder of this paper we will use the notion $\hat{\theta}^{PL-I}$ to represent our image estimate from the penalized-likelihood approach in the image domain.

C. Sinogram-Domain Objective function

Estimating the attenuation maps for datasets as large as those acquired in CT is extremely computationally intensive. In previous work we considered a more modest but still meaningful goal—that of estimating the set of monochromatic attenuation line integrals l_i from the set of measurements y_i. The estimated line integrals can then be input to a standard analytic reconstruction algorithm. We estimate the set of line integrals needed for reconstruction by maximizing a penalized-likelihood objective function. According to the previous model, the measurements follow statistical distributions given by the sum of a Poisson distribution and a Gaussian distribution. We thus seek an to find the estimate for the line integrals

$$\hat{\mathbf{l}} = \arg\max_{\mathbf{l}>0} \left[\mathcal{L}(\mathbf{l}, \mathbf{y}) - \beta R(\mathbf{l})\right] \tag{5}$$

where $\mathcal{L}(\mathbf{l}, \mathbf{y})$ is the Poisson log-likelihood function, $R(\mathbf{l})$ is a penalty function, and β is a regularization parameter. We then estimate the image from the these line integrals with FBP

$$\hat{\theta}^{PL-S} = \mathbf{B}^{\text{FBP}}\left(\hat{l}\right). \tag{6}$$

where $\mathbf{B}^{\text{FBP}}$ is the discretized FBP matrix and the approximate inverse of the system matrix $\mathbf{A}^{-1}$. For the remainder of this paper we will use the notion $\hat{\theta}^{PL-S}$ to represent our image estimate from the penalized-likelihood approach in the sinogram domain.

D. Local Impulse Response

For an estimator with mean $\hat{\bar{\theta}}$, we we define the local impulse response of the ith pixel to be

$$\text{lir}_j(\theta) = \lim_{\delta \to 0} \frac{\hat{\bar{\theta}}^{Imp} - \hat{\bar{\theta}}}{\delta} = \frac{\partial}{\partial \theta_j}\hat{\bar{\theta}}, j = 1, ..., M \tag{7}$$

where we define a new measurement model $Y^{Imp} = \left[Y_1^{Imp}, .., Y_N^{Imp}\right]'$ which denotes a random measurement vector of $\theta^{Imp} = \left[\theta_1^{Imp}, .., \theta_M^{Imp}\right]'$. Here $\theta^{Imp} = \theta + \delta e^j$ where δe^j is a small impulse placed in the jth pixel.

1) Brute Force Evaluation of Local Impulse Response: Most estimators $\hat{\theta}(y)$ do not have an explicit analytical form. Thus a numerical approach to evaluate the local impulse response is a valid option, or estimating the sample mean by running several realizations in a computer simulation. The following recipe illustrates this brute-force approach. (1)Select an object θ of interest and generate multiple realizations of noisy measurements according to the density $f(y;\theta)$. (2)Apply the estimator of interest to each of the measurement realizations to obtain estimates; estimate the sample mean. (3)Choose a pixel j of interest and small value δ, and generate a second set of noisy measurements according to the density $f\left(y;\theta^{IMP}\right)$. (4)Compute the sample mean from the second set and subtract the two images to obtain the local impulse response.

2) Approximate Brute Force Evaluation of Local Impulse Response: Several investigators have observed that the sample mean of a likelihood-based estimator is approximately equal to the value that one obtains by applying the estimator to noiseless data

$$\hat{\bar{\theta}} = E_\theta\left[\hat{\theta}(Y)\right] \approx \hat{\theta}\left(\overline{Y}(\theta)\right) \triangleq \breve{\theta} \tag{8}$$

where $\breve{\theta}$ denotes the value of the estimator when given noiseless data $\bar{Y}$. This approximation is equivalent to assuming that the estimator is locally linear and leads to not only a simpler recipe for numerically evaluation the local impulse response, but far less computational. The procedure is. (1)Select an object θ of interest, a pixel j of interest and a small value δ. (2)Generate two noiseless measurements vectors: $\overline{Y}(\theta)$ and $\overline{Y}\left(\theta^{Imp}\right)$. (3)Apply the estimator of interest to each of the two noiseless measurements, obtaining estimates $\hat{\theta}\left(\bar{Y}(\theta)\right)$ and $\hat{\theta}\left(\bar{Y}\left(\theta^{Imp}\right)\right)$. (4)Estimate the local impulse response.

3) Analytic Evaluation of Local Impulse Response: Many estimators in tomography are defined implicitly as the maximizer of some objective function. We assume Φ has a unique global maximum, so that $\theta(y)$ is well defined. Fessler has shown [4] that penalized likelihood objective function have linearized local impulse responses of the form

$$\text{lir}_j = \left[\mathbf{A}'\mathbf{D}\left[\frac{\bar{Y}_i^2\left(\breve{\theta}\right)}{\bar{Y}_i(\theta)}\right]\mathbf{A} + \beta\mathbf{R}\left(\breve{\theta}\right)\right]^{-1}\mathbf{A}'\mathbf{D}\left[\frac{\bar{Y}_i^2\left(\breve{\theta}\right)}{\bar{Y}_i(\theta)}\right]\mathbf{A}e^j. \tag{9}$$

When evaluating this expression the local impulse response for transmission tomography using the penalized likelihood methods described become

$$\text{lir}_j^{PL-S} = \mathbf{B}^{\text{FBP}}\left[\mathbf{D}_{\breve{\theta}^2/\theta}^{PL-S} + \beta\mathbf{R}\right]^{-1}\mathbf{D}_{\breve{\theta}^2/\theta}^{PL-S}\mathbf{A}e^j \tag{10}$$

where

$$\mathbf{D}_{\breve{\theta}^2/\theta}^{PL-S} = \text{diag}\left\{\left[\frac{\left(\mathbf{A}\breve{\theta}^{PL-S}\right)_i^2}{\bar{Y}_i(\theta)}\right]\right\} \tag{11}$$

and

$$\text{lir}_j^{PL-I} = \left[\mathbf{A}'\mathbf{D}_{\breve{\theta}^2/\theta}^{PL-I}\mathbf{A} + \beta\mathbf{R}\left(\breve{\theta}\right)\right]^{-1}\mathbf{A}'\mathbf{D}_{\breve{\theta}^2/\theta}^{PL-I}\mathbf{A}e^j. \tag{12}$$

where

$$\mathbf{D}_{\breve{\theta}^2/\theta}^{PL-I} = \text{diag}\left\{\left[\frac{\left(\mathbf{A}\breve{\theta}^{PL-I}\right)_i^2}{\bar{Y}_i(\theta)}\right]\right\}. \tag{13}$$

4) Approximate Analytic Evaluation of Local Impulse Response: For moderate or small values of β, $\breve{\theta}$ is a slightly blurred version of θ. Since the projection operation $\mathbf{A}\theta$ is a smoothing operator, the projections $Y(\theta)$ and $Y\left(\breve{\theta}\right)$ are approximately equal. Also, since $\mathbf{R}$ is θ independent when using a quadratic penalty the above expression simplifies to

$$l_j \approx \left[\mathbf{A}'D\left[\bar{Y}_j(\theta)\right]\mathbf{A}+\beta\mathbf{R}\right]^{-1}\mathbf{A}'\mathbf{D}\left[\bar{Y}_j(\theta)\right]\mathbf{A}e^j \tag{14}$$

When applying this approximation to the two penalized likelihood methods we get

$$\mathrm{lir}_j^{PL-S} \approx \mathbf{B}^{\mathrm{FBP}}\left[\mathbf{D}_\theta+\beta\mathbf{R}\right]^{-1}\mathbf{D}_\theta\mathbf{A}e^j \tag{15}$$

and

$$\mathrm{lir}_j^{PL-I} \approx \left[\mathbf{A}'\mathbf{D}_\theta\mathbf{A}+\beta\mathbf{R}\right]^{-1}\mathbf{A}'\mathbf{D}_\theta\mathbf{A}e^j \tag{16}$$

where

$$\mathbf{D}_\theta = \mathrm{diag}\left\{\bar{Y}_i(\theta)\right\}. \tag{17}$$

This estimation of the local impulse response is desirable since it is independent of the PL estimate of the image and depends only on the noiseless data.

E. Variance

Evaluating just the local impulse response itself is insufficient since resolution has an inherent trade-off with noise. Thus, to better measure the image quality we also evaluated the noise properties of the estimators.

1) Brute Force Evaluation of Variance: Again, when there is no explicit form for $\hat{\theta}(y)$, there is usually no explicit form for the estimator variance either, and a numerical approach becomes a valid option. The following recipe illustrates this brute-force approach. (1)Select an object θ of interest and generate multiple realizations of noisy measurements according to the density $f(y;\theta)$. (2)Apply the estimator of interest to each of the measurement realizations to obtain estimates. (3)Estimate the estimator mean using the sample mean. (4)Estimate the estimator sample variance.

2) Analytic Evaluation of Variance: Fessler has shown [5] that penalized likelihood objective function have covariances of the form

$$cov\{\theta\} = \left[\mathbf{A}'\mathbf{D}_{\breve{\theta}}\mathbf{A}+\beta\mathbf{R}\right]^{-1}\mathbf{A}'\mathbf{D}_\theta\mathbf{A}\left[\mathbf{A}'\mathbf{D}_{\breve{\theta}}\mathbf{A}_{\breve{\theta}}+\beta\mathbf{R}\right]^{-1}. \tag{18}$$

This equation can be used in its current form to evaluate the $cov\left\{\theta^{PL-I}\right\}$. A similar derivation

$$cov\left\{\theta^{PL-S}\right\} = \mathbf{B}^{\mathrm{FBP}}\left[\mathbf{D}_{\breve{\theta}}+\beta\mathbf{R}\right]^{-1}\mathbf{D}_\theta\left[\mathbf{D}_{\breve{\theta}}+\beta\mathbf{R}\right]^{-1}\mathbf{B}^{\mathrm{FBP}\prime} \tag{19}$$

can be used to evaluate the sinogram-domain PL estimate

3) Approximate Analytic Evaluation of Variance: Again, for moderate or small values of β, the projections $Y(\theta)$ and $Y\left(\breve{\theta}\right)$ are approximately equal. Also, since $\mathbf{R}$ is θ independent when using a quadratic penalty the above expression simplifies to

$$cov\left\{\theta^{PL-S}\right\} \approx \left[\mathbf{A}'\mathbf{D}_\theta\mathbf{A}+\beta\mathbf{R}\right]^{-1}\mathbf{A}'\mathbf{D}_\theta\mathbf{A}\left[\mathbf{A}'\mathbf{D}_\theta\mathbf{A}+\beta\mathbf{R}\right]^{-1} \tag{20}$$

and

$$cov\left\{\theta^{PL-I}\right\} \approx \mathbf{B}^{\mathrm{FBP}}\left[\mathbf{D}_\theta+\beta\mathbf{R}\right]^{-1}\mathbf{D}_\theta\left[\mathbf{D}_\theta+\beta\mathbf{R}\right]^{-1}\mathbf{B}^{\mathrm{FBP}\prime}. \tag{21}$$

III. RESULTS

A. Ellipse phantom and sinogram

For the numerical phantom, we used the 1024 × 1024 attenuation image shown in Fig. 1, which has relative linear attenuation coefficients of 0.030, 0.096 and 0.130] cm^{-1} in the hot disk (left), background ellipse and cold disk (right) respectively. The pixel size was .0375 cm. The background ellipse has a major axis of 34.54 cm and minor axis of 19.2 cm while the hot and cold disks have 11.52 cm diameters. This phantom was created to match the phantom used in Fessler's work.

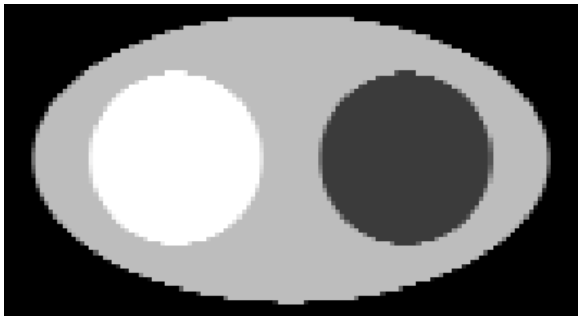


Figure 1. Numerical ellipse phantom used to determine image properties

To examine the performance of the proposed penalized-likelihood algorithms, we simulated both noisy and noiseless projections of this numerical phantom. We computed a sinogram of 1024 parallel bins of extent .0375 cm at the isocenter by 1024 angles acquired over 360 degrees. We did not employ a quarter detector offset. We simulated the data using a monochromatic spectrum at the average energy $\bar{E}$. Due to computation cost the forward projection data was rebinned using bilinear interpolation to sinograms with more modest size dimensions. These dimensions are listed with the experiment that used them. When creating noisy sinograms, Poisson statistics according to the forward model of Eq. 3 were used on the resized sinogram. The electronic noise variance is equivalent to the counting statistics variance corresponding to a mean of 10 detected photons under the same model.

We reconstructed the images using several combinations of smoothing and analytic reconstruction. The PL^{sino} algorithm was applied with 31 beta parameters β and then reconstructed with FBP using ramp and Hanning filters with .5 f_c. The PL^{image} algorithm was used to reconstruct these images at 31 different β values. FBP using the Hanning filter was used as initial guess.

B. System Matrices

In the standard transmission model the system matrix $\mathbf{A}$ is an $M \times N$ matrix that forward projects the object data into the sinogram domain. To reduce computational burdens the projection matrix was calculated once and then stored and accessed. Additionally, symmetries in the projection matrix were exploited to reduce the size. These matrices are also sparse, which further reduced the storage.

C. Approximate Brute Force Evaluation of Resolution

Resolution was measured by the full-width half-max (FWHM) metric. The index and value of the highest intensity pixel was recorded. The index location of the pixels immediately below

and above that value along the x and y-axis were recorded. The half-max location was calculated by linear interpolation of these pixels and used to calculate the full-width for the x-axis and y-axis. Since the impulses are not isotropic, impulse profiles were measured at 1 degree intervals and averaged together. Since the image is discretized, this an unblurred impulse would have a FWHM values between 1 and $1\sqrt{2}$.

The brute force method described earlier was implemented at three location in the phantom; in the center of both discs and in the center of the phantom. 1000 noisy realizations of the simulated phantoms were created at an I_i of 1e6 and then reconstructed. Impulse, δ, of value of .05 cm^{-1} were added to the simulated phantom and 1000 realizations of its sinogram were also reconstructed. The resolutions were compared at each of the 31 β for both the penalized likelihood methods on sinograms rebinned to 128x128 to the approximate brute force evaluation of local impulse response. It was confirmed that the resolution matched precisely and validated the use of the approximate method with noiseless data.

IV. Discussion

The analytic approximations of sinogram-domain penalized likelihood approach agreed very well with each other and the Monte Carlo numerical simulations for a wide range of regularization parameters. These results can be seen in Figure 2. The analytic approximations of image-domain penalized likelihood approach agreed very with each other and relatively well with the Monte Carlo numerical simulation. These results can be seen in Figure 3. The small discrepancy is thought to be due to the propagating numerical errors in matrix operations. The sinogram domain approach is less affected by this since its system matrices are tridiagonal and operating on these matrices requires extensively less mathematical operations. The two approaches, also, match relatively well with each other. These results can be seen in Figure 4. In future work we plan to introduce sinogram- and image-domain degradations and geometry effects in the system matrix.

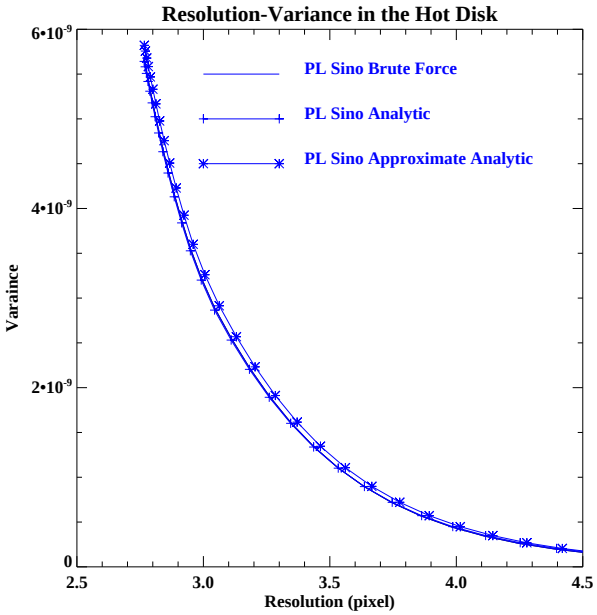


Figure 2. Resolution variance plots for the sinogram domain reconstruction approach.

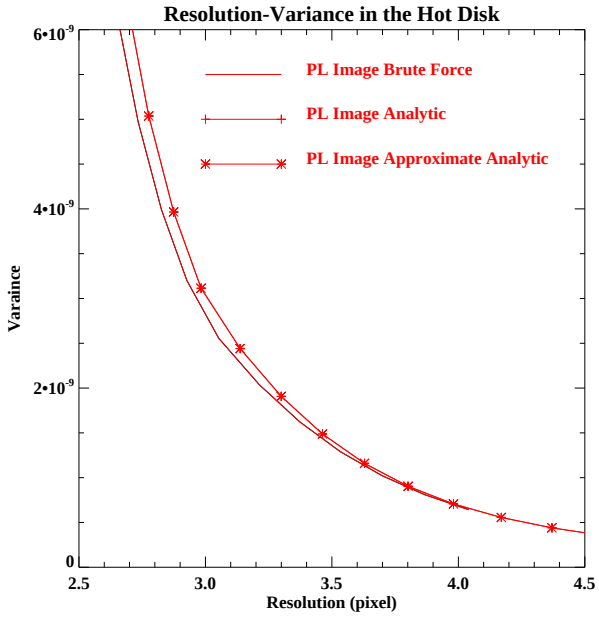


Figure 3. Resolution variance plots for the image domain reconstruction approach.

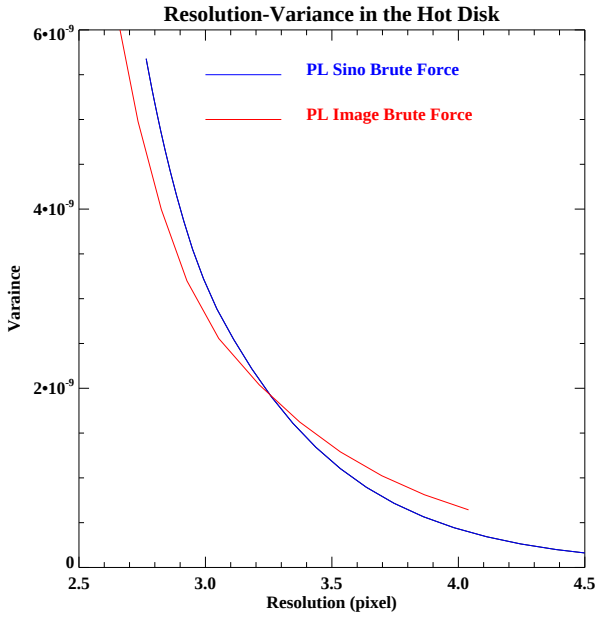


Figure 4. Resolution variance plots for the both approaches.

References

[1] P. J. La Rivière and P. Vargas, "Potential equivalence of sinogram and image-domain penalized likelihood methods," in *Proc. 2007 IEEE Nucl. Sci. Symp. and Med. Imag. Conf.*, 2007. CD-ROM.

[2] J. A. Fessler, "Statistical image reconstruction methods for transmission tomography," in *Handbook of Medical Imaging, Vol. 2* (J. M. Fitzpatrick and M. Sonka, eds.), pp. 1–70, Bellingham, WA: SPIE Press, 2000.

[3] P. J. La Rivière, "Penalized-likelihood sinogram smoothing for low-dose CT," *Med. Phys.*, vol. 32, pp. 1676–1683, 2005.

[4] J. A. Fessler and W. L. Rogers, "Spatial resolution properties of penalized-likelihood image reconstruction methods: Space-invariant tomographs," *IEEE Trans. Image Processing*, vol. 5, pp. 1346–1358, 1996.

[5] J. A. Fessler and W. L. Rogers, "Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): Applications to tomography," *IEEE Trans. Image Processing*, vol. 5, pp. 493–506, 1996.

Iterative Image Reconstruction with Half Precision

Clemens Maaß and Marc Kachelrieß

***Abstract*—Recently, the authors have shown that an upcoming lower precision floating–point data type can be adopted to reduce the memory needs and to shorten reconstruction times of analytic CT image reconstruction. This paper shows how these savings in reconstruction hardware can be used during the computationally even more demanding iterative image reconstruction and where the limits of the new data type are. To do so, the SART algorithm (simulataneous algebraic reconstruction technique) is implemented using the *half* data type and the standard *float* data type. The reconstruction results are compared regarding the voxel noise and the convergence speed as a function of the size of the subset of the projections that is used to calculate one volume update. The results are that the *half* data type introduces only insignificant additional image noise if the subset sizes are not chosen too large.**

***Index Terms*—CT Reconstruction, Image Quality, Fast Reconstruction, Larrabee**

I. Introduction

IMAGES from computed tomography (CT) show highly quantitative values. Typically, the Hounsfield scale is the basis of CT image visualization and the dynamic range of the images ranges from -1000 HU up to 3000 HU. This states a quantization of the dynamic range of CT images in 4000 different steps.

Due to the drawbacks of floating–point processing in former data processors there were experiments to use integer values to handle CT data [1]. Today, those drawbacks are overcome and the motivation to reduce data is the costly RAM and the limited memory bandwidth that restricts the speed of today's reconstruction algorithms on CPU and GPU hardware.

In the past years, the *half* data type was discussed and finally described in the IEEE 754-2008 standard. For example, current GPUs support the *half* precision format and mainly profit from reduced memory requirements. Further on, AMD's streaming SIMD extensions 5 (SSE5) that supplement the advanced vector extensions (AVX) include instructions regarding the *half* precision format [2], [3]. Even more sophisticated support of this data type will be offered by the Intel Larrabee architecture as it allows to implicitly convert between *half* and *float* during data transfer between cache and register without loss of performance [4], [5]. *Half* floating–point values require only 16 bit for each datum while the *float* values use 32 bit. Since it is floating–point it combines a high dynamic range with sufficient significant digits to allow to reconstruct and represent high quality CT images. The authors have shown that the *half* data type is sufficient to represent convolved projection data [6] and to repesent the volume during backprojection [7]. This work shows that the image quality loss is negligible when using the *half* data type during standard iterative image reconstruction.

Institute of Medical Physics (IMP), University of Erlangen–Nürnberg, Henkestr. 91, 91052 Erlangen. Corresponding author: clemens.maass@imp.uni-erlangen.de

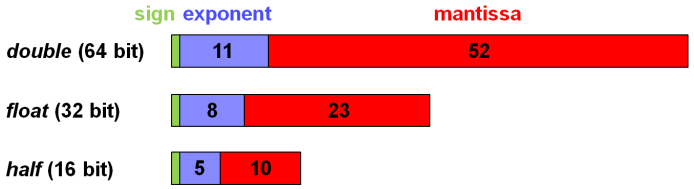


Fig. 1. Illustration of the allocation of the available bits to sign, exponent, and mantissa for selected data types.

II. Materials and Methods

A. IEEE floating–point standard

The IEEE 754 standard describes how real numbers are approximated by finite precision floating–point numbers and how this representation is done. Let $0 \leq S < 2^s$ be the $s = 1$–bit sign, $0 \leq E < 2^e$ be the e–bit exponent, and $0 \leq M < 2^m$ be the m–bit mantissa. A finite precision floating–point number X with $b = s + e + m$ bits can be represented as a function of S, E, and M as follows:

$$X = X(S, E, M) = (-)^S \cdot 2^{E+1-2^{e-1}} \cdot (1 + M \cdot 2^{-m})$$

Note the hidden–1 representation for the mantissa magnitude which effectively increases the number of bits by 1. Figure 1 depicts the allocation of the available bits of *half*, *float*, and *double* precision on s, e, and m. The special case $E = 0$, where IEEE 754 deviates from this normal representation and switches to the sub–normal representation [6] is not described here for the sake of simplicity.

B. Volume representation *in* half *precision*

In the range from -1000 HU up to 1000 HU there is a special accuracy required that should not be lower than 1 HU. A higher accuracy is not necessary, because most CT image visualization tools work with integer values and are not able to visualize a sub-HU scale. This is also in accordance with the accuracy of a well calibrated CT scanner.

The 10 bit mantissa yields an accuracy of 11 bit because there is a hidden–1 representation. This means, that for values beween 0 and 2047 an accuracy of at least 1 can be maintained. For the range between 2048 and 4095 at least an accuracy of 2 is available.

During the reconstruction process, the volume data are typically not stored in the Hounsfield scale (CT–values) but using μ-values. The Hounsfield scale is defined as

$$\text{CT–Value} = \left(1000 \cdot \frac{\mu}{\mu_{\text{Water}}} - 1000\right) \text{HU}. \qquad (1)$$

It is the nature of the floating–point format that the factor between the μ–space and the HU–space does not affect the data accuracy. However, the offset is relevant: -1000 HU correspond to $\mu = 0$. Therefore, the actual accuracy is at least 1 HU beween -1000 HU and 1047 HU and at least 2 HU in the range 1048 HU to 3095 HU. This fits exactly the needs of the physician: In ranges over 1000 HU the absolute CT value is not as important as below those values.

C. Backprojection

The simulations in this paper use a standard voxel–driven backprojection: Each voxel is projected on the detector and linear interpolation is used to determine the value that needs to be added to the given voxel. In reference [7] the authors have shown that special care is necessary when using the *half* data type during backprojection of filtered projection data. Filtered backprojection has high requirements on the dynamic range of the volume data type because each projection adds a very small contribution to potentially large values.

In iterative reconstructions there is typically not the full projection set backprojected at a time. In order to improve the performance, often only a subset of the projections is backprojected at a time (OS–EM, OSC, SART) [8], [9], [10]. Using subsets of the projection set relaxes the requirements on the volume data type accuracy.

However, as reference [7] clearly shows that large subset sizes (in combination with *half* data type) degrade the image quality in analytic image reconstruction significantly, the number of projections per subset N_{Sub} is an important parameter. The influence of N_{Sub} during iterative image reconstruction is analyzed in this work, given a fixed total number of projections N.

D. Forward projection

For the forward projection a Joseph–like algorithm is used [11]. Volume values are accumulated along rays from the source position to each detector pixel. The variable that accumulates the values needs a high dynamic range because a high number of voxels is accumulated. However, there is no use in saving memory for the accumulation variable as — depending on the implementation — this variable typically does not leave the CPU's cache or even not the CPU register. Therefore, the accumulator is represented in *float* precision. Once the line integral is completely calculated for a ray, the value is converted into *half* precision and stored into the RAM. This scenario shows that *half* precision does usually not constrain the forward projection step and the influence of the reduced data type accuracy is restricted to the sinogram representation.

E. Iterative reconstruction algorithm

In this paper, the simultaneous algebraic reconstruction technique (SART) is used. The SART update equation is

$$\boldsymbol{f}^{k+1} = \boldsymbol{f}^{k} + \text{BP}_s \left(\frac{\boldsymbol{p}_s - \text{FP}_s\left(\boldsymbol{f}^{k}\right)}{\text{FP}_s\left(\mathbf{1}\right)} \right), \qquad (2)$$

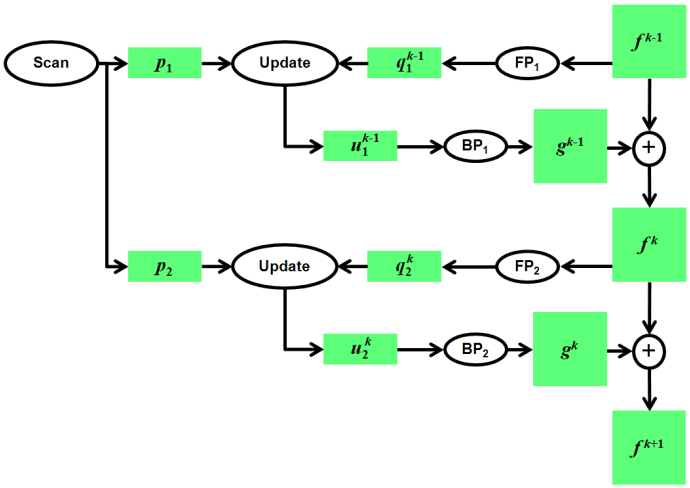


Fig. 2. Visualization of the first two volume updates during the SART reconstruction ($k = 1$). All occuring intermediate results are marked in green. These intermediate results (volumes or projection sets) may be represented in *half* precision.

where $\boldsymbol{f}^k$ is the reconstructed image after k, $1 \leq k < K$, updates, FP denotes the forward projection as described above, BP accordingly denotes the backprojection as described above, and $\boldsymbol{p}$ denotes the original rawdata. The index s is used to denote the subset number $0 \leq s < S$ of the current subiteration, where $S = \frac{N}{N_{\text{Sub}}}$ is the total number of subsets. Consequently, FP_s and BP_s work only on those projections that belong to the subset s and $\boldsymbol{p}_s$ contains only the according projection values. The parameter $K = I \cdot S$, with I being the number of iterations, defines how often the volume is updated. When comparing reconstructions of different subset sizes N_{Sub}, K is kept constant rather than the number of iterations as the results were found to be better comparable then.

Figure 2 illustrates the implementation of the SART image reconstruction. The forward projection step of the current volume $\boldsymbol{f}^k$ yields $\boldsymbol{q}_s^k$. After application of the update equation (2), the update rawdata set $\boldsymbol{u}_s^k$ needs to be backprojected to come up with an update image $\boldsymbol{g}^k$. All intermediate projection sets ($\boldsymbol{p}_s^k$, $\boldsymbol{q}_s^k$, $\boldsymbol{u}_s^k$), the update volume $\boldsymbol{g}_s^k$, and the result of update k, $\boldsymbol{f}^k$, are visualized as green blocks in figure 2. All those volumes and projections are represented in *half* precision for reconstruction of the *half* image $\boldsymbol{f}_{\text{h}}$ and in *float* precision for the standard *float* image $\boldsymbol{f}_{\text{f}}$.

F. Convergence speed

Since the used data type accuracy may influence the convergence speed of the iterative image reconstruction, the convergence speed is evaluated. To do so, after each update k of the reconstruction volume, the sum of the absolute difference between the image $\boldsymbol{f}^k$ and the *float* precision reconstruction result image $\boldsymbol{f}_{\text{f}}^K$ is calculated. This yields the normalized convergence

$$C_\nu(k) = \frac{|\boldsymbol{f}_\nu^k - \boldsymbol{f}_{\text{f}}^K|}{|\boldsymbol{f}_{\text{f}}^K|}, \qquad (3)$$

where $\nu = \text{h}$ denotes the convergence of the *half* image and $\nu = \text{f}$ denotes the convergence of the *float* image.

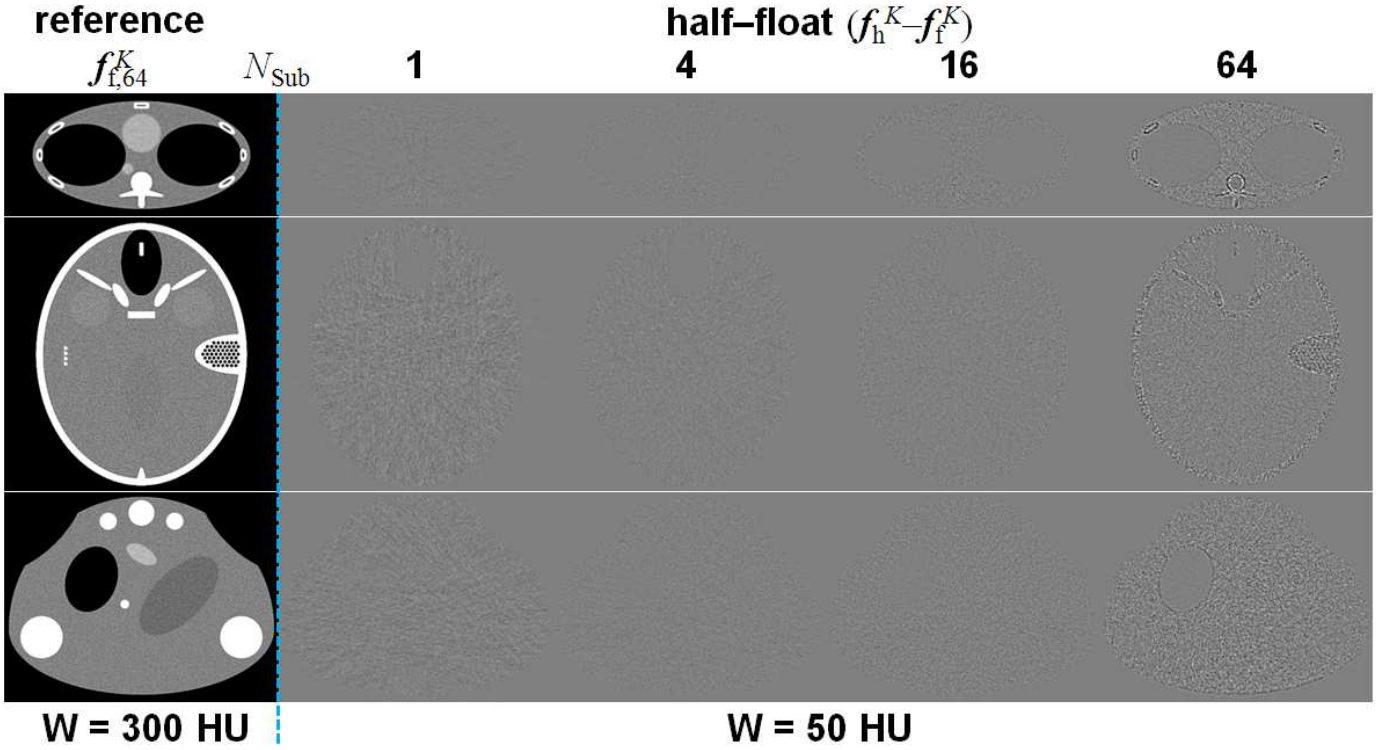


Fig. 3. SART reconstructions of selected N_{Sub} with noise added to projection data. Since all reconstructed images appear to be identical, only the difference images $\boldsymbol{f}_{\text{h},N_{\text{Sub}}}^K - \boldsymbol{f}_{\text{f},N_{\text{Sub}}}^K$ are shown next to the reference image $\boldsymbol{f}_{\text{f},64}^K$. The reconstruction error that is caused by the reduced precision of the data is insignificant for small subset sizes and becomes more significant when using larger subsets. Note that the presented difference images $\boldsymbol{f}_{\text{h}}^K - \boldsymbol{f}_{\text{f}}^K$ do not use the same number of iterations but the same number of updates $K = N = 768$, such that $I = N_{\text{Sub}}$.

G. Experiments

Simulations of different phantoms were performed to evaluate the abilities of iterative *half* precision reconstructions. All simulations were performed using the center slices of the Forbild thorax phantom, the Forbild head phantom, and the micro–CT mouse phantom [12]. Projection data were simulated by analytically calculating line integrals through the mathematically defined objects that are combined using constructive solid geometry. Noise–free projections and projections with Gaussian noise were used in all experiments.

Projection data of $N = 768$ projections that are equiangularly distributed on a 360° rotation were simulated in two dimensional parallel beam geometry with quarter detector offset. A detector with 512 pixel was simulated and reconstructions use an image matrix of 512×512 voxels. The size of the field of measurement (FOM) was adapted to fit the phantom into it by adaption of the detector pixel size and voxel size. This resulted in an FOM diameter of 500 mm for the thorax phantom, of 250 mm for the head phantom, and of 33 mm for the mouse phantom.

Using these simulations, the image quality is evaluated for SART reconstruction by comparing gold standard *float* precision reconstructions with *half* precision reconstructions for different values of $N_{\text{Sub}} \in \{1, 2, 4, 8, 16, 32, 64\}$. This is done visually by difference images, by voxel noise evaluation of a region of interest, and by plotting the convergence speed $C(k)$.

III. Results

A. Dependency on N_{Sub}

N_{Sub}	1	2	4	8	16	32	64
Thorax phantom							
$\boldsymbol{f}_{\text{f}}^K$	14.4	9.7	7.8	7.2	7.0	7.0	6.9
$\boldsymbol{f}_{\text{h}}^K$	14.5	9.8	7.9	7.4	7.4	7.5	8.4
$\boldsymbol{f}_{\text{f}}^K - \boldsymbol{f}_{\text{h}}^K$	0.1	0.1	0.1	0.2	0.4	0.5	1.5
Head phantom							
$\boldsymbol{f}_{\text{f}}^K$	16.0	10.3	8.2	7.5	7.3	7.2	7.2
$\boldsymbol{f}_{\text{h}}^K$	16.2	10.3	8.3	7.7	7.5	7.5	8.4
$\boldsymbol{f}_{\text{f}}^K - \boldsymbol{f}_{\text{h}}^K$	0.2	0.0	0.1	0.2	0.2	0.3	1.2
Mouse phantom							
$\boldsymbol{f}_{\text{f}}^K$	10.5	7.2	5.8	5.4	5.3	5.2	5.2
$\boldsymbol{f}_{\text{h}}^K$	10.7	7.3	6.1	5.7	5.8	6.1	6.8
$\boldsymbol{f}_{\text{f}}^K - \boldsymbol{f}_{\text{h}}^K$	0.2	0.1	0.3	0.3	0.5	0.9	1.6

TABLE I
EVALUATION OF THE STANDARD DEVIATION IN ROIS OF WATER–EQUIVALENT DENSITY AS A FUNCTION OF THE SUBSET SIZE N_{Sub} IN HU.

Figure 3 shows the reconstruction results using *float* and *half* data precision as a function of N_{Sub}. Thereby only selected numbers of N_{Sub} are shown. Table I lists the noise values measured in a selected ROI of water–equivalent material in *float* and *half* reconstructions of each phantom.

It is noticeable that the additional quantization noise that comes with the *half* precision is constant at an insignificantly low level up to a subset size of about 32. Then, image quality starts to suffer significantly from the reduced data accuracy.

This is also noticeable in the difference images of figure 3: These images show only low level noise for subset sizes between 1 and 32. For a subset size of 64, edges of the phantom appear in the difference image in addition to the noise. Consequently, *half* precision should not be used when reconstructions work with unusual large subsets.

B. Convergence

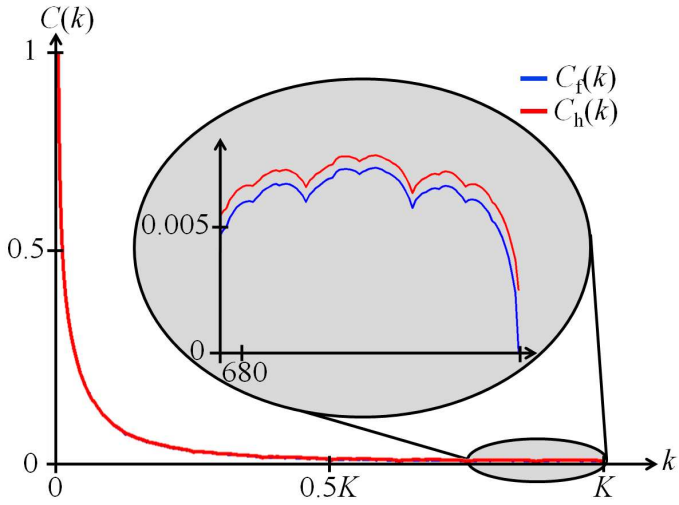


Fig. 4. Convergence of the Forbild head phantom during SART reconstruction with $K = N = 768$. The selected plot shows the convergence for a subset size of $N_{\text{Sub}} = 16$.

Figure 4 plots the convergence $C(k)$ of the *half* and the *float* reconstructions of the Forbild head phantom using $N_{\text{Sub}} = 16$. It is clearly visible that during the whole reconstruction process both plots are almost identical. This proves that the convergence of the image is not influenced by the reduced data precision.

By definition, the convergence $C_{\text{f}}(K) = 0$. This means, that $C_{\text{h}}(K) > 0$ describes the difference of the final images and represents the additional quantization noise and the different propagation of quantum noise in the *half* image.

IV. Discussion

The evaluation of the image noise shows that for large subset sizes the additional quantization noise is not acceptable. However, large subsets (i.e. more than 64 projections that are processed at a time to come up with one volume update) are unusual. Typical SART reconstruction uses small subsets and it has turned out that then the additional quantization noise that comes with the *half* precision is negligible. The additional quantization noise can be expected to be object–dependent but independent of the quantum noise. As a consequence, for higher quantum noise levels than those used in this study, the influence of the additional quantization noise is even lesser.

Since the *half* data type is a method to improve the usage of given hardware resources, the results must be seen in the context of a constant reconstruction time on a given hardware. There may be cases where the introduction of the *half* precision can actually double the reconstruction speed. This could be the case if the speed of the reconstruction process is strongly limited by the memory bandwidth or if the required data exceed the local RAM of a graphics board. One can read from table I that then doubling the subset size and simultaneously introducing *half* precision has the potential to yield reconstructions of less image noise for a fixed reconstruction time. Doubling the subset size of the SART reconstruction almost doubles the computational load what is compensated by the doubled reconstruction speed when using *half* precision data.

V. Conclusion

Half precision floating–point values have the potential to represent a CT image at sufficient accuracy and can thereby speed up image reconstruction on a given hardware. It has been shown that SART reconstruction where projection data and volume data are completely represented in *half* precision is possible with the penalty of slightly increased image noise. The use of *half* precision is not indicated if large subsets are used.

Acknowledgment

This work was supported by a grant from the Intel Corporation. Parts of the reconstruction software and simulation software were provided by RayConStruct GmbH, Nürnberg, Germany.

References

[1] K. Mueller and R. Yagel, “Rapid 3-D cone-beam reconstruction with the simultaneous algebraic reconstruction technique (SART) using 2-D texture mapping hardware,” *IEEE Transactions on Medical Imaging*, vol. 19, no. 12, pp. 1227–1237, December 2000.

[2] “128-bit SSE5 instruction set,” Advanced Micro Devices, Technical Report, October 2009, http://developer.amd.com/cpu/SSE5.

[3] “Intel AVX,” Intel Corporation, Technical Report, October 2009, http://softwareprojects.intel.com/avx/.

[4] L. Seiler, D. Carmean, E. Sprangle, T. Forsyth, M. Abrash, P. Dubey, S. Junkings, A. Lake, J. Sugerman, R. Cavin, R. Espase, E. Grochowski, T. Juan, and P. Hanrahan, “Larrabee: A many–core x86 architecture for visual computing,” *ACM Transactions on Graphics*, vol. 27, no. 3, pp. 18:1 – 18:15, Aug. 2008.

[5] M. Baer, C. Maaß, and M. Kachelrieß, “CT image reconstruction using Larrabee,” *IEEE Medical Imaging Conference Record*, pp. HP1–2, 2009.

[6] C. Maaß, S. Steckmann, M. Baer, L. Hillebrand, W. A. Kalender, and M. Kachelrieß, “CT image reconstruction with half precision floating point values,” *Proc. of the 2009 Int. Meeting on Fully 3D Image Reconstruction, Workshop on High Performance Image Reconstruction*, pp. 24–28, September 2009, www.fully3d.org.

[7] C. Maaß, M. Baer, and M. Kachelrieß, “CT image reconstruction on Intel Larrabee using half precision floating–point values,” *IEEE Medical Imaging Conference Record*, pp. HP1–3, 2009.

[8] H. M. Hudson and R. S. Larkin, “Accelerated image reconstruction using ordered subsets of projection data,” *IEEE Transactions on Medical Imaging*, vol. 13, no. 4, pp. 601–609, December 1994.

[9] S. H. Manglos, G. M. Gagne, F. D. Thomas, and R. Narayanaswamy, “Transmission maximum-likelihood reconstruction with ordered subsets for cone beam CT,” *Phys. Med. Biol.*, vol. 40, no. 7, pp. 1225–1241, 1995.

[10] C. Kamphuis and F. J. Beekman, “Accelerated iterative transmission CT reconstruction using an ordered subsets convex algorithm,” *IEEE Transactions on Medical Imaging*, vol. 17, no. 6, pp. 1101–1105, 1998.

[11] P. M. Joseph, “An improved algorithm for reprojecting rays through pixel images,” *IEEE Transactions on Medical Imaging*, vol. MI–2, no. 3, pp. 192–196, Nov. 1982.

[12] “Phantom group,” Institute of Medical Physics, Technical Report, 2009, www.imp.uni-erlangen.de/phantoms.

Novel Strategies to Reduce Patient Organ Dose in CT Without Reducing Tube Output

Di Zhang, Chris Cagnon, John DeMarco, Adam Turner, Michael McNitt-Gray
David Geffen School of Medicine at UCLA
Los Angeles, CA 90024 USA

Abstract- Radiation dose to individual organs is a proper dose metric to be used in CT because it more accurately reflects actual risk than do standard dosimetry indices such as CTDI. The feasibility of reducing organ dose by controlling tube start angle and table height was studied using Monte Carlo simulations. Simulations were performed on several voxelized patient models under a variety of conditions. The results showed that the control of tube start angle can reduce dose to small peripheral organs by 20% to 40%. The control of table height can reduce organ dose by 10% to 50%. The combination of controlling both amplifies the dose saving. This study introduces two new techniques to reduce radiation dose to selected organs in CT without changing tube output. They have minimal effect on image quality.

I. INTRODUCTION

Radiation dose from CT continues to be of interest to the medical imaging community. The risk involved with this level of radiation dose is not fully understood because of the lack of data at these low dose levels. However, the implications of CT radiation dose on public health may be significant. For example, a study suggested that 0.4% of all recent cancers in the United States may be attributed to CT radiation dose performed between 1991 and 1996. If organ specific cancer risk caused by ionization from CT was evaluated, this cancer rate can go up to 1.5% to 2% [1]. The radiation dose delivered to pediatric patients has come under scrutiny because of their high risk per unit dose [2] and the higher dose they receive unless the scanning protocol is adjusted for their smaller size. Various approaches have been developed to reduce radiation dose, including tube current modulation [3], adjusting mAs for patient size [4,5], and lowering kVp (especially for studies using iodinated contrast [6]).

The larger coverage afforded by wider z-axis beams in MDCT creates larger cone angles and greater beam divergence, which results in substantial surface dose variation for helical and contiguous axial scans. This phenomenon has the potential to be used in patient scans for dose reduction purposes. A previous study evaluated this variation of in both cylindrical and anthropomorphic phantoms [7]. Figure 1 shows the peripheral dose profiles on a CTDI body phantom at various tube start angles. Different tube start angles create a phase shift in the peripheral dose profile, resulting in dramatic variation in dose at a given z-axis location (e.g. location 0 in figure 1). In the light of these 'hot spots' or 'cold spots' created by the change of tube start angle, it may be feasible to exploit these variations to reduce dose to selected radiosensitive peripheral organs solely by varying the tube start angle in CT scans, especially if these organs have relative small extent in the longitudinal direction.

The table height during a CT scan has drawn some attention in recent years and it has implications for both radiation dose and image quality [8, 9]. Our recent measurements on a CTDI 32cm phantom indicated that as the table moves up, CTDI at the 12:00 position decreases. This phenomenon may be utilized for organ dose reduction.

Therefore, the purpose of this work was to extend the previous simulation work to investigate the effects of the change of tube start angle and table height on patient organ dose and explore their potential combined use for organ dose reduction

II. MATERIAL AND METHODS

A. CT scanner model and patient models

Monte Carlo method was used to simulate the spectrum and geometry of the CT source, the geometry of the patients, as well as the transport of photons. A previously developed CT source model was used in this study [10~12]. This model was validated and benchmarked using comparisons based on standard dosimetry (CTDI) measurements and corresponding simulations, which agreed to within 5%. In this study all the simulations were performed on a Siemens Sensation 64 MDCT scanner model. For patient models, the GSF (now: Helmholtz Zentrum München) phantom series was used. These are voxelized patient models with segmented individual organs [13]. In Monte Carlo simulations, the mean dose to an organ was estimated by averaging the dose to each voxel across all the voxels belonging to that organ, and the simulations were operated in photon transport mode with a low-energy cutoff of 1keV. Charged-particle equilibrium (CPE) was assumed so that all the secondary electrons deposit their energy at the photon interaction sites.

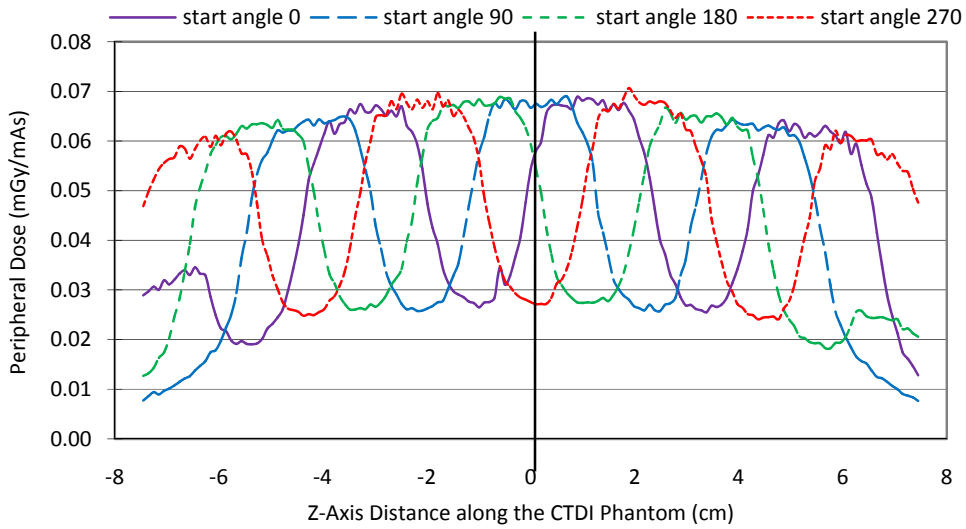


Figure 1. Peripheral dose profile on 32cm CTDI phantom for different start tube start angles using pitch 1.5 and collimation 24x1.2mm.

B. Simulation experiments

1. Dose reduction by varying tube start angle only

To demonstrate the organ dose reduction from the change of tube start angle, especially for small peripheral organs, radiation dose to several radiosensitive organs (including breast glandular tissue, thyroid, uterus, gonads, eye lens) resulting from helical scans. These were estimated using Monte Carlo simulation son voxelized patient models, including GSF's Baby, Child, and Irene, which were chosen to represent a range of patient size. Dose to fetus was also estimated using 4 pregnant female models. Whole body scans were simulated using 120kVp, 300mAs, both 28.8 mm and 40 mm nominal collimation, and pitch value of 1.5, under a wide range of start angles (0 to 340 degrees in 20 degree increments). The relationship between tube start angle and organ dose was examined for each organ and potential dose reduction was calculated.

2. Organ dose reduction by varying table height only

To demonstrate the organ dose reduction effect from the change of table height, Monte Carlo simulations for full body scans were performed at various table heights (from the table height where the center of the patient is 12cm below the isocenter of CT gantry to the table height where the center of the patient is 12cm above the isocenter of CT gantry, with 0.75cm increments) for GSF Irene phantom. All the simulated scans were performed at 120kVp, 200mAs, 28.8 mm collimation, and pitch of 1.5.

3. Organ dose reduction by varying both tube start angle and table height

To investigate the dose benefits to organs by controlling both tube start angle and table height in CT scans, simulations were performed to estimate radiation dose to all radiosensitive organs for four different GSF models which span a range of sizes (Baby, Irene, Helga, and Visible Human) under different combinations of tube start angles and table heights. The tube start angle ranges from 0 degree to 340 degree with 20 degree increments, and the table height ranges from 4cm below the isocenter to 12cm above the isocenter. Full body scans were performed for all four patient models, all combinations of tube start angle and table heights at 120kVp, 200mAs, 28.8mm collimation, and pitch of 1.5.

III. RESULTS

A. Dose reduction from changing tube start angle to small peripheral organs

Some organs exhibit strong dose variation depending on the tube start angle. For small peripheral organs, (e.g. the eye lenses of the Baby phantom at pitch 1.5 with 40 mm collimation), the minimum dose can be 41% lower than the maximum dose, depending on tube start angle. In general, larger dose reductions occur for small peripheral organs in small patients when wider collimation is used. The maximum dose reduction for pregnant patient models shows that the magnitude of dose reduction to the fetus for pitch 1.5 can be 18% for the early gestational age models (≤ 12 weeks) for 40 mm collimation setting. Results from the later gestational model (19 weeks) used in this study little dose reduction.

The angle reported is the angle of the source as it crosses the longitudinal center of the organ being investigated, which is called the Organ Crossing Tube Angle (OCTA). The OCTA resulting in the highest dose was used as the worst case reference and the dose reduction was calculated for each OCTA value. The organ dose reductions are shown in Figure 2 by plotting the % reduction of organ dose as a function of the OCTA for Child phantom for 28mm collimation and pitch of 1.5. The organ dose reduction as a function of OCTA can be observed from the figure. It was shown that for pitch 1.5 scans, the dose is usually lowest when the tube start angle is such that the x-ray tube is posterior to the patient when it passes the longitudinal location of the organ (OCTA = 180). For each organ on a specific patient, there is a deterministic relationship between the OCTA and tube start angle so a tube

start angle can be calculated to make OCTA of 180. These results demonstrate that radiation dose to small peripheral organs can be reduced by controlling tube start angle.

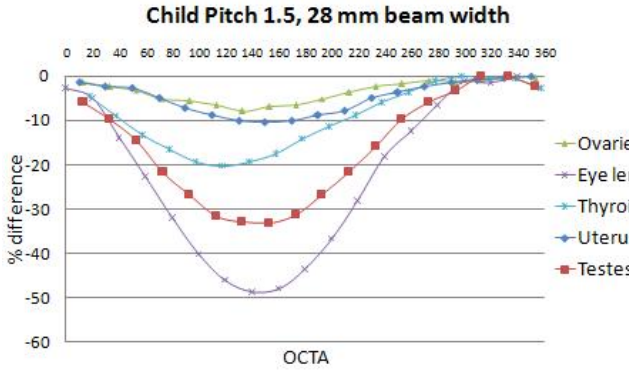


Figure 2. The organ dose variation curves for Child phantom from a simulated CT scan with pitch 1.5 and 28mm nominal collimation.

B. Dose reduction to all organs from changing table height

Figure 3 shows the results about radiation dose to several radiosensitive organs for GSF Irene model at various table heights. Dose to breast glandular tissue, kidney, eye lens and rectum are shown in this figure. These results show that the change of table height has considerable effect on the organ dose. For example, by moving the table 12cm up, dose saving to eye lens can be 48%, and dose saving to breast glandular tissue can be 29%, compared to the table position where the patient is center at CT isocenter.

The dose saving trend is not consistent among all the organs. For example, by moving the table up, dose to glandular breast tissue decreases, but dose to kidneys increases a little bit (around 10%) before it decreases again. By summarizing the dose saving behavior for all the organs, it was shown that for anterior organs (e.g. glandular breast tissue), dose decreases as the table moves up; for posterior organs (e.g. kidneys), dose decreases as the table move down. As the table is far away enough from the isocenter (e.g. more than 10cm), all the organ receive lower dose. The rationale of this trend will be discussed in section IV. These results demonstrate that radiation dose to selected organs can be reduced by controlling table height.

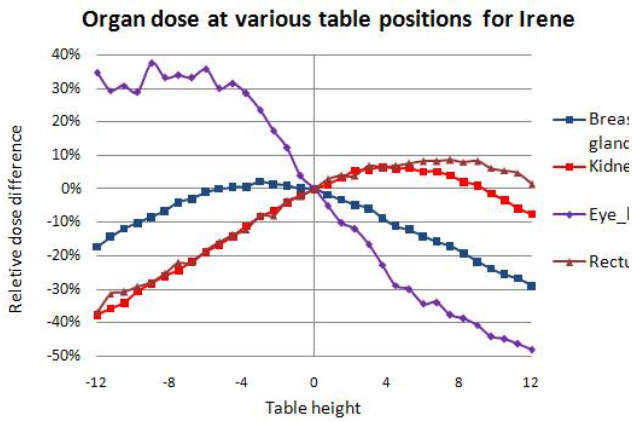


Figure 3. Radiation dose to several radiosensitive organs of GSF Irene at various table heights.

C. Organ Dose reduction on patients with different size by controlling both tube start angle and table height

For large organs (e.g. larger than the table increment per rotation), the dose saving from controlling tube start angle is very small, therefore most of the dose saving is from the control of table height. Figure 4a shows the dose to breast glandular tissue for Helga as a function of tube start angle at different table heights. Dose does not change much across tube start angle, but the dose saving could be as high as 20% when the table is raised by 12cm, compared to isocenter. Table I lists the percent dose reduction when table is raised by 12cm for several organs with highest weighting factors (radiosensitivity) for all four patient models, averaged over all the tube start angles (e.g. averaged over each curve in figure 4a). These organs happen to be fairly large or non-peripheral organs and therefore tube start angle has a minor effect.

Table I. Dose saving achieve by raising the table by 12cm, averaged over all the tube start angles.

	Red Bone Marrow	Colon	Lung	Stomach	Gonad
Baby	24%	30%	23%	32%	23%
Irene	16%	30%	13%	30%	16%
Helga	16%	23%	7%	24%	9%
Visible Human	12%	27%	10%	25%	34%

For smaller organs, dose variation is created by both table height and tube start angle. For example, figure 4b shows the dose to breast glandular tissue for Baby as a function of tube start angle at different table heights. This figure shows considerable dose saving from the combination of selection of tube start angle and table height. For example, there can be as much as 68% dose saving when the optimum OCTA is chosen at table height of 12cm comparing to the worst case scenario at table height of 0. In addition, dose reduction effects from tube start angle and table height interact with each other and the overall dose saving can be amplified. For example, when the patient is centered, the dose benefit from tube start angle dose is 25%; when the patient is located 12cm above isocenter, the dose benefit from tube start angle can be as high as 55%. This is because the dose saving from tube start angle is higher at more peripheral positions because of the more significant surface dose variations. And raising the table higher puts a lot of organs into more peripheral positions.

IV. DISCUSSION and CONCLUSION

All current dose reduction schemes in CT involve the change of the output of the x-ray tube, for example, reducing mAs or kVp. In this study, Monte Carlo method based simulations were performed to investigate the feasibility of two new dose reduction strategies in CT without change tube output by the controlling of tube start angle and table height. The results showed considerable dose reduction to selected organs by either using one of these two strategies or the combination of them.

The effectiveness of controlling tube start angle on dose reduction on small peripheral organs is due to the surface dose variation introduced by the helical source pathway and wide cone angle. This technique may be applied to a small organ on small patients, fetus on pregnant patients with early gestational stage, or eye lenses on adult patients.

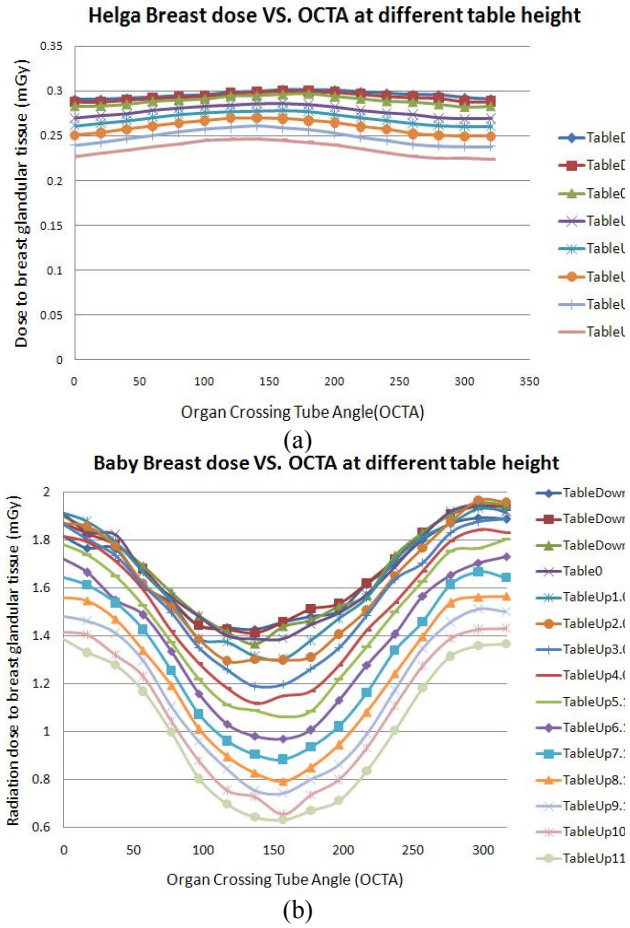


Figure 4. Radiation dose to breast glandular tissue as a function of OCTA at various table heights for: a) Helga; b) Baby.

The effectiveness of controlling table height on dose saving can be attributed to the bowtie filter. Because of the shape of the bowtie filter, if an object is scanned close to the isocenter in the CT gantry, it always receives lowest filtration from the bowtie filter from each projection and therefore receives higher dose, while if it is far away from the isocenter, it receives high filtration caused by the bowtie filter in many projections and hence has lower dose. When this principle is applied to a patient, dose to organs at different locations will behave differently. Specifically, dose to an organ would decrease if the table is moved such that this organ is further away from isocenter. And it would increase if the table is moved such that this organ is closer to isocenter. However, when the table is far enough from isocenter all the organs would receive lower dose because they are essentially all moved away from isocenter. This was illustrated in figure 3.

In summary, this study introduces two new techniques to reduce radiation dose to selected organs in CT. These two techniques should have minimum effect on image quality. Tube start angle has no effect on image quality because it is not controlled by the user and is either a completely random variable or a limited random variable in current commercial CT scanners. Table height does have some effect on image quality (for example, increase the noise for the part of the image where dose is lower [9]), however, this can be avoided in clinical applications by identifying the anatomy of interest and the dose reduction site. For example, the table can be raised to reduce breast dose for kidney examinations.

REFERENCES

[1] Brenner, D.J. & Hall, E.J. Computed tomography--an increasing source of radiation exposure. N Engl J Med 357, 2277-2284, 2007.

[2] ICRP. RECOMMENDATIONS OF THE ICRP. ICRP publication 103, 2008.

[3] C. H. McCollough, M. R. Bruesewitz and J. M. Kofler, Jr., "CT Dose Reduction and Dose Management Tools: Overview of Available Options," Radiographics 26, 503-512, 2006.

[4] J.-F. Paul and H. Abada, "Strategies for reduction of radiation dose in cardiac multislice CT," European Radiology 17, 2028-2037, 2007.

[5] U. K. Udayasankar, K. Braithwaite, M. Arvaniti, D. Tudorascu, W. C. Small, S. Little and S. Palasis, "Low-Dose Nonenhanced Head CT Protocol for Follow-Up Evaluation of Children with Ventriculoperitoneal Shunt: Reduction of Radiation and Effect on Image Quality," AJNR Am J Neuroradiol 29, 802-806, 2008.

[6] Z. Szucs-Farkas, L. Kurmann, T. Strautz, M. Patak, P. Vock and S. Schindera, "Patient Exposure and Image Quality of Low-Dose Pulmonary Computed Tomography Angiography: Comparison of 100- and 80-kVp Protocols," Investigative Radiology 43, 871-876, 2008.

[7] D. Zhang, A. S. Savandi, J. J. Demarco, C. H. Cagnon, E. Angel, A. C. Turner, D. D. Cody, D. M. Stevens, A. N. Primak, C. H. McCollough and M. F. McNitt-Gray, "Variability of surface and center position radiation dose in MDCT: Monte Carlo simulations using CTDI and anthropomorphic phantoms," Medical Physics 36, 1025-1038, 2009.

[8] Matsubara, K., et al. Misoperation of CT automatic tube current modulation systems with inappropriate patient centering: phantom studies. AJR Am J Roentgenol 192, 862-865, 2009.

[9] Toth, T., Ge, Z. & Daly, M.P. The influence of patient centering on CT dose and image noise. Med Phys 34, 3093-3101, 2007.

[10] G. Jarry, J. J. DeMarco, U. Beifuss, C. H. Cagnon and M. F. McNitt-Gray, "A Monte Carlo-based method to estimate radiation dose from spiral CT: from phantom testing to patient-specific models," Physics in Medicine and Biology, 2645, 2003.

[11] J. J. DeMarco, C. H. Cagnon, D. D. Cody, D. M. Stevens, C. H. McCollough, J. O. Daniel and M. F. McNitt-Gray, "A Monte Carlo based method to estimate radiation dose from multidetector CT (MDCT): cylindrical and anthropomorphic phantoms," Physics in Medicine and Biology, 3989, 2005.

[12] J. J. DeMarco, C. H. Cagnon, D. D. Cody, D. M. Stevens, C. H. McCollough, M. Zankl, E. Angel and M. F. McNitt-Gray, "Estimating radiation doses from multidetector CT using Monte Carlo simulations: effects of different size voxelized patient models on magnitudes of organ and effective dose," Physics in Medicine and Biology, 2583, 2007.

[13] N. Petoussi-Henss, M. Zankl, U. Fill and D. Regulla, "The GSF family of voxel phantoms," Physics in Medicine and Biology, 89, 2002.

A proposed approach for validating Monte Carlo Computed Tomography dosimetry simulations

Adam C. Turner[1] and Michael McNitt-Gray[1]
[1]Departments of Biomedical Physics and Radiology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90024 USA

Abstract- The use of Monte Carlo radiation transport codes have become a popular method to estimate the radiation dose to radiosensitive organs in patient models from scans performed with modern Multidetector row Computed Tomography (MDCT) scanners. The purpose of this abstract is to review common approaches to the development and utilization of Monte Carlo MDCT dosimetry packages and to propose a standardized series of benchmarking experiments that can be used to validate different components of these codes. The ultimate goal of this work is to develop a standardized test suite that is universal to any Monte Carlo CT dosimetry code whose results can be used to directly compare different simulation packages.

I. INTRODUCTION

Recent studies report that from 1993 to 2006 the number of computed tomography (CT) imaging procedures increased at an annual rate of over 10% in the United States leading to a considerable increase in the collective radiation dose from CT.[1] Specifically, CT exams now constitute 15% of the total number of radiological imaging procedures and contribute over half of the population's medical radiation exposure.[1] It has been suggested that the most appropriate quantity for assessing the risk due to diagnostic imaging procedures is the radiation dose to individual organs.

The use of Monte Carlo radiation transport codes that simulate the delivery of radiation from CT scanners to patient models has become a popular method of investigating organ dose. Typically, these codes take into account scanner-specific characteristics such as x-ray energy spectra, filtration designs, beam collimation, fan-angle, and pitch to model. Conventional Monte Carlo radiation transport techniques are used to track the path of simulated photons through a computational anthropomorphic phantom and tally the dose deposited in regions of interest.

The Monte Carlo approach has a number of advantages over measuring dose to organs in anthropomorphic phantoms using point dosimeters. Physical measurements require detailed calibration of dosimeters and assume organ doses can be approximated by point dose measurements. Also, Zhang et al. have shown, measurements made on the surface of cylindrical and anthropomorphic phantoms may have large variations (up to 50%) for helical scans depending on pitch and start angle.[2]

A number of different techniques have been employed by different research groups in order to develop Monte Carlo CT dosimetry packages that model specific multidetector CT (MDCT) scanners.[3-6] The disparities between the different packages range from fundamental radiation transport techniques to advanced aspects of modeling MDCT scanners. For example, it is common to base simulation packages on well-validated, general purpose radiation transport codes such as the Monte Carlo N-Particle (MCNP) code from Los Alamos National Laboratory; however, some groups have created radiation transport code from scratch. Also, the methods used to model the delivery of radiation from CT scanners can be quite different. On an even higher level, the data sets used to simulate a specific scanner, such as x-ray energy spectrum or filtration design, can vary across different codes.

Most Monte Carlo modeling publications include descriptions of benchmark experiments carried out to validate the code. Commonly, simple phantom measurements (such as CTDI) are compared to analogous simulations. The question arises as to whether typical validation experiments sufficiently provide the confidence necessary to support the detailed results and conclusions being published in the literature. Is it sufficient to match a measured value for a single axial scan measurement (such as CTDI) to validate a study that ultimately reports dose from helical scans? Do the assumed scanner-specific input parameters (such as energy spectrum and filtration) adequately represent the physical characteristics of that scanner with respect to two- and three-dimensional dose profiles in a realistic heterogeneous environment? Is the degree of detail when modeling patient anatomy and tissue suitable to calculate and report specific organ dose values?

In order to address these questions, the purpose of this work is to propose a set of generalized benchmark techniques meant to more completely validate various pieces of a simulation package. These validation experiments will be proposed in such a manner so that a centralized database of physically measured results can be made available along with all relevant simulation inputs except for the portion of the simulation package being validated. In other words, all variable inputs necessary for the simulation will be provided except for the input being tested. Ideally, the final product will be a standardized benchmark suite that can be universally applied for validating any Monte Carlo CT dosimetry code and that will allow direct comparisons of different simulation packages.

II. OVERVIEW OF MDCT MONTE CARLO SIMULATIONS

A. Radiation transport algorithm

The most elemental component of a dosimetry simulation is the Monte Carlo radiation transport code used to track particles and tally energy deposition. These codes use

statistical methods to randomly sample the potential physical interactions of elementary particles (photons and electrons) with matter while keeping track of the resulting transferred energy. This can be done using either experimental or theoretically derived cross-section data. Simulating a large number of particles and computing a running average should result in computed energy depositions that approach those observed in nature.

Many MDCT dosimetry packages are built upon general purpose transport codes like MCNP.[2-5] These "canned" transport codes have undergone rigorous benchmarking and the MDCT simulation packages that utilize them can be considered sufficiently accurate in regards to their transport capabilities. However, some packages utilize transport codes written in-house[6]. The intricate details of the techniques employed in these transport codes are generally not described in publications. A first step in developing a convincing simulation package would therefore be to validate the accuracy of the underlying transport methodology in order to gain confidence in the sampling techniques and cross-section data being employed.

B. Source motion and x-ray trajectories

Various techniques have been used to simulate the paths traversed by the x-ray source of MDC scanners. In some packages the initial position of each photon is randomly sampled from continuous functions that describe the source path with respect to the patient geometry[3]. Source motion functions can be formulated for both axial and helical scans and typically are dependent on scanner- and protocol-specific characteristics such as the source-to-isocenter distance, the nominal collimation setting, and pitch. Alternatively, some packages individually specify a discrete number of source positions to describe the source path[4,5]. For example, one group simulated a single axial scan with 72 stationary line sources, parallel to the axis of rotation, each with a length equal to the longitudinal length of the bowtie filter and located on a circle with a radius equal to the focal spot to isocenter distance. A bowtie filter was then placed in front of each source for the simulation[4].

Based on the starting location of the photon, it is necessary to select an initial direction. The techniques used for sampling the direction space also differ among the published simulation packages. Some groups use isotropic x-ray emitters and collimate based on the source position while others explicitly define the potential trajectories for a given source position and randomly sample from them. In most cases this step takes scanner-specific fan-angles and measured beam widths are into account. The particular choice of methods can alter the necessary modeling of actual beam widths as well as $1/r^2$ intensity effects.

C. Scanner-specific source model

In order to perform Monte Carlo simulations of a specific commercially available MDCT scanner it is necessary to acquire a detailed description of the scanner's x-ray energy spectrum, the bowtie and inherent filtration design, and the geometry of the scanner (e.g. focal spot to isocenter distance, fan angle, z-axis collimation, cone angle settings, etc.). While it is usually possible to ascertain the geometry of a scanner from documentation, descriptions of filtration and spectra are generally proprietary. To circumvent this limitation some groups have worked with scanner manufacturers to obtain spectra and filtration designs (usually through a non-disclosure agreement)[4]. Others have used generic tungsten anode x-ray energy spectra[3] that are based on experimentally measurements or theoretically derivations. This approach still requires information about filtration, which is usually acquired from the manufacturer. Alternatively, studies have been performed to examine the utility of obtaining spectrum and filtration schemes based on measurement. Turner, et al. published a method to generate scanner-specific "equivalent" spectrum and filtration based on a series of measurements[7]. Gu et al. described an approach in which a number of parameters are optimized in order for simulated 32 cm CTDI values at center and periphery to match measured values[5]. Their parameters included anode angle (used to choose an energy spectrum), flat filter thickness, bowtie filter shape, and number of source points.

D. Patient modeling techniques

Most simulation packages use voxelized anthropomorphic computational models that feature detailed organ contours, typically generated directly from patient images, either by manual segmentation or by threshold algorithms based on CT numbers. Despite the improvement over stylized and mathematical phantoms in terms of anatomical accuracy, these packages still make assumptions and simplifications when assigning material composition and density values to the individual voxels. For example, a number of codes assume scanned objects consist only of air, water/soft tissue, and bone. This type of approximation is often used when organs are not individually identified prior to simulation. On the other hand, if contoured organs are identified it would possible to assign material composition and density values to voxels based on their corresponding organ assignment. Appropriate elemental compositions and density values can be found in the literature, for example in the ICRU-44 composition of body tables. Also, the data sets that these models are built from do not always include the entire patient anatomy. Therefore, with these models it impossible to tally dose in regions of the body adjacent to the scan volume or account for potential scatter from absent anatomy. It is currently unknown the magnitude of errors these type of assumptions introduce for dose calculations.

III. PROPOSED VALIDATION METHODS

In this section a proposed series of experiments are described that could be used to rigorously validate the accuracy of a Monte Carlo MDCT dosimetry package. The goal of these experiments will be to benchmark each part of the simulation

code, starting with the most fundamental pieces and moving toward the more detailed. Specifically, we recommend experiments to validate the radiation transport code, the generalized source motion/collimation model, scanner-specific source model inputs (x-ray energy spectrum, filtration, geometry), and the anthropomorphic patient models.

The approach taken in these proposals all involve publishing data sets for Monte Carlo simulation results to be benchmarked against. Additionally, all inputs and parameters necessary to perform the required simulation, except for those being tested, will be provided. For example, if longitudinal dose profile measurements for a helical scan are obtained with a specific scanner the energy spectrum and filtration design necessary to simulate that scanner will be published. It would then be possible to test the methods used to model the source motion for a helical scan. This systematic approach was chosen because it allows individual components of a MDCT simulation package to be tested and compared to other packages.

A. Radiation transport algorithm

As described in section II.A, simulation packages that utilize general purpose transport codes like MCNP are typically considered validated with respect to their transport capabilities. However, it is important for in-house transport codes to be thoroughly tested. A series of simple simulations were proposed by Nikolopoulos et al. that were performed using both MCNP and various application specific codes for simulations in the diagnostic energy range.[9] Each of the proposed simulations involves simulating a narrow beam of photons perpendicularly incident on a water phantom with infinite surface area from a distance of 1 cm. The tests include tallying:

- the number of interactions per simulated photon in water phantoms of three different thicknesses (10, 15, and 20 cm) at various energies (20-100 keV)
- the depth of energy deposition in a 10 cm thick water phantom (energy deposited in 1 cm thick slabs, moving away from the source) for monoenergetic x-ray beams of 15, 20, and 30 keV
- the lateral spread of energy deposition inside of an 8 cm thick water phantom (energy deposited in 1 cm wide slabs 4 cm deep, emanating out from the phantom center) for monoenergetic x-ray beams of 50 keV.

The results presented by Nikolopoulos et al. suggest that relative differences between the codes should be less than approximately 10% for the majority of these tests (a few differences were 15-20%) in order to sufficiently validate a transport code. The data presented in this paper may be used as an initial benchmark set that researchers can use to validate their own codes. It also may be beneficial to extend the range of energies up to 140 keV, which is the typical upper limit used by MDCT scanners.

B. Source motion and collimation

Section II.B outlined the various techniques used to model source motion and photon trajectory. Benchmarking is normally performed by comparing analogous simulated and measured longitudinal dose profiles obtained with a series of point dosimeters placed the surface of a phantom. This approach can be problematic because, as described by Zhang et al., the shape of a surface dose profile depends on both pitch (for a helical scan) and the gantry angle at which the tube turned on[2]. Typically, the "start angle" cannot be controlled when obtaining contiguous axial or helical scans for most scanners (especially in clinical scan modes). As a result, it is often necessary to artificially shift the simulated dose profile so its phase matches that of the measured profile, which potentially negates the comparison. Additionally, Zhang et al.'s work indicates that the surface dose effects for very low pitches may necessitate nearly continuous measurements instead of a series of discrete points.

In order to more carefully and thoroughly assess the source motion for axial and helical simulations we propose that longitudinal dose profiles be obtained using a method similar to that described by Deak et al.[8] Approximately continuous dose profile measurements will be made in the center and 12 o'clock positions of the 16 and 32 cm CTDI phantoms. This can be done using custom made TLD inserts (as described by Deak et al.) or using OSL strips (as described in Turner, et al[7]). These dose profiles will be obtained for a single axial scan as well as 10 cm helical scans with pitches of 0.5, 1, and 1.5. Each measurement will be performed using both the narrowest and widest collimations using a single MDCT scanner (i.e. Sensation 64) at one tube voltage. For each scan, the start angle of the scan will be obtained from the raw data file of the scanner. The single axial scans will determine the accuracy of the collimation model while the helical scans will validate the simulated source motion. The provided inputs for the simulations will include the scanner-specific equivalent x-ray energy spectrum, the bowtie filter description, scanner geometry, start angle for each scan, and phantom model. This information can then be used with the user's simulation package to simulate single axial and helical scans analogous to the described measurements and compare the results.

C. Scanner-specific source model

In order to publish results that are intended to be specific to a particular MDCT scanner it is necessary to thoroughly validate the source model parameters described in section II.C. This is often done using CTDI validation techniques, which confirms that the source models can match measurements in a simple, homogenous cylindrical phantom. We propose a more meticulous procedure that validates the energy spectrum and bowtie filter with a variety of benchmarks. Measurements include:

- $CTDI_{100}$ at center and periphery for 16 cm and 32 cm phantoms for each protocol condition
- exposure in air samples from isocenter to top of bed's range, with the tube parked, normalized to isocenter value (to sample across bowtie filter), as shown in figure 1

- dose at several points in multiple slices of a heterogeneous, non-uniform phantom (i.e. Rando-Alderson phantom) using point dose dosimeters (i.e. TLDs or small ion chambers) for each protocol condition.

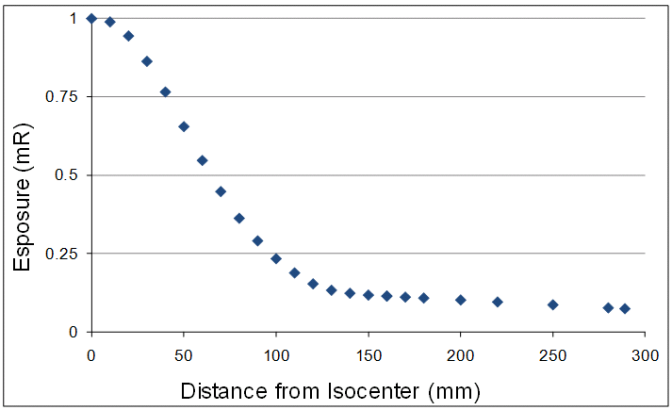


Fig. 1. Exposure as a function of distance from isocenter using a fixed tube position.

A data base of these measurements will be published for all MDCT scanners and available protocols (all bowtie filters, collimations, tube voltages, etc). Additionally, voxelized and/or geometric models of the phantoms will also be published. Analogous simulations can then be performed with the input spectrum and filtration data in order to validate these models.

D. *Patient modeling techniques*

In regards to constructing patient models, the current literature does not address the magnitude of errors introduced by making the approximations described in section II.D. It is almost impossible to obtain a measured value to serve as a benchmark gold standard in order to validate patient modeling. In theory, a point dosimeter could be inserted into actual patients or cadavers, however, the dose variation issues described by Zhang et al.[2] would make the measurements difficult to interpret. Instead, it would be informative to perform a sensitivity study in which the level of detail that goes into patient modeling is varied and the results are compared. Such studies will be performed with regards to the number of materials used in the model (a full set of ICRU-44 anatomical tissues vs. air/soft tissue/bone) and the extent of the anatomy (full body vs. partial body models). The results will be used to determine the magnitude of the approximations that can made while still having confidence in the results when modeling a patient for dosimetry simulations.

IV. DISCUSSION

The methods presented in this work were developed in order to address the limitations of common Monte Carlo MDCT simulation code validation techniques. A series of benchmark experiments were proposed that each address a particular component common to most simulation packages. Each experiment consists of a simulation description, all inputs necessary to perform the simulation except for that being tested, and the benchmark measurements or data that the simulation results should be compared with. The benchmark measurements consist of physically measured data wherever possible, however, some simulation data obtained with well-validated codes are also used. The experiments were presented in a specific order so that the fundamental aspects of the code are validated first, followed by more detailed components. This approach to provides a standardized approach to validation that will allow for high confidence in individual simulation packages as well as the ability to directly compare different packages with each other in order to determine the optimal Monte Carlo MDCT simulation techniques.

ACKOWLEDGMENTS

This work was supported by NIBIB Grant R01-EB004898.

REFERENCES

[1] F.A. Mettler Jr., B.R. Thomadsen, M. Bhargavan, D.B. Gilley, J.E. Gray, J.A. Lipoti, J. McCrohan, T.T. Yoshizumi, and M. Mahesh, "Medical radiation exposure in the U.S. in 2006: preliminary results," Health Phys. 95(5), 502-507 (2008).

[2] D. Zhang, A. S. Savandi, J.J. DeMarco, C.H. Cagnon, E.A. Angel, A.C. Turner, D.D. Cody, D.M. Stevens, A.N. Primak, C.H. McCollough, and M.F. McNitt-Gray, "Variability of surface and center position in MDCT: Monte Carlo simulations using CTDI and anthropomorphic phantoms," Med. Phys. 36(3), 1025-1038 (2009).

[3] J.J. DeMarco, C.H. Cagnon, D.D. Cody, D.M. Stevens, C.H. McCollough, J. O'Daniel, and M.F. McNitt-Gray, "A Monte Carlo based method to estimate radiation dose from multidetector CT (MDCT): cylindrical and anthropomorphic phantoms," Phys. Med. Biol. 50 3989–4004 (2005).

[4] A. Tzedakis, J. Damilakis, K. Perisinakis, J. Stratakis, and N. Gourtsoyiannis, "The effect of z overscanning on patient effective dose calculated for CT examinations," Med. Phys. 32(6), 1621-1629 (2005).

[5] J. Gu, B. Bednarz, P.F. Caracappa, and X.G. Xu, "The development, validation and application of a multi-detector CT (MDCT) scanner model for assessing organ doses to the pregnant patient and the fetus using Monte Carlo simulations," Phys. Med. Biol. 54, 2699-2717 (2009).

[6] B. Schmidt and W.A. Kalendar, "A fast voxel-based Monte Carlo method for scanner- and patient-specific doe calculations in computed tomography," Physica Medica XVIII. 43-53 (2002).

[7] A. Turner, D. Zhang, H.J. Kim, J.J. DeMarco, C.H. Cagnon, E. Angel, D.D. Cody, D.M. Stevens, A.N. Primak, C.H. McCollough, and M.F. McNitt-Gray. "A method to generate equivalent energy spectra and filtration models based on measurement for multidetector CT Monte Carlo dosimetry simulations," Med. Phys. 36(6), 2154-2164 (2009).

[8] P. Deak, M. van Straten, P.C. Shrimpton, M. Zankl, and W.A. Kalendar, "Validation of a Monte Carlo tool for patient-specific dose simulations in multi-slice computed tomography," Eur. Radiol. 18, 759–772 (2008).

[9] D. Nikolopoulos, D. Linardatos, I. Valais, C. Michail, S. David, P. Gonias, N. Bertsekas, D. Cavouras, A. Louizi, and I. Kandarakis, "Monte Carlo validation in the diagnostic radiology range," Nucl. Instr. and Meth. **571** (1-2), 267-269 (2007).

Beam Shaper with Optimized Dose Utility for Helical Cone-Beam CT

Thomas Köhler[1], Bernhard Brendel[1], and Roland Proksa[1]

Abstract— **A beam shaper is proposed that modulates the intensity of the x-ray beam as a function of the detector aperture. Such a beam shaper can increase the SNR by 12% if analytical reconstruction is used. In combination with iterative reconstruction, the beam shaper creates an insensitivity to patient motion without introducing any heuristic data weighting.**

I. Introduction

Computerized tomography (CT) uses ionizing radiation that harms people. Although the individual risk to cause cancer with a single CT examination is small, the excessive use of CT in diagnostic imaging may lead to a significant total number of induced cancer cases. Recent studies and estimates claim that the current use of CT in the US of about 70 million examinations per year may cause several thousands of cancer incidences [1, 2].

In all practical cases, the data acquired during a helical scan are at least partially redundant. Proper incorporation of redundant data in the framework of analytic reconstruction methods was one of the hot topics in CT reconstruction research in the beginning of this century. A number of heuristic approximate methods [3–7] as well as heuristic extensions of mathematically exact methods [8–10] have been proposed.

All these analytic cone-beam reconstruction algorithms perform weighting of redundant data, which is not optimal from a statistical point of view. But there is a need to trade-off dose utilization, level of remaining cone-beam artifacts, and the sensitivity to patient motion. These aspects are discussed in some of the aforementioned papers [4,8,9].

Motion artifacts as well as other artifacts due to inconsistencies in the data can also show up in iterative reconstruction methods. One way to mitigate these artifacts is to introduce heuristically an aperture weighting function that fades out the contribution of data near the edge of the detector [11, 12]. This weighting increases the robustness of the algorithm, but it necessarily increases noise since the contribution of data are no longer in accordance with their statistical significance.

The basic idea presented in this paper is to modulate the incoming intensity of the x-ray beam such that the weighting that is used during reconstruction fits to the statistical significance of the measured data, which leads to an improved dose utilization. This can be achieved using a proper beam shaper. The standard beam shaper used in current medical CT scanner (aka bowtie filter) modulates the intensity of the x-ray beam as a function of the fan-angle [13], see Fig. 1. The goal of this beam shaper is to compensate roughly for the different path length of the x-rays through the patients body. The proposed new beam shaper introduces an additional modulation of the beam intensity as function of the cone-angle.

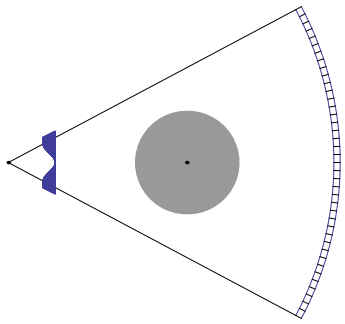


Fig. 1. Illustration of a beam shaper located between focal spot and patient.

II. Method

A. Analytic Reconstruction Algorithms

Consider the aperture weighted wedge reconstruction as described in [8, 14] as a prototype for analytic reconstructions. For each object point to be reconstructed and each projection value that contributes to this object point, all redundant data are identified and a brute-force normalization of the corresponding aperture weights is performed. This procedure ensures that all available data are used during reconstruction.

The statistically optimal way to average redundant data d_i with variances $\mathrm{var}(d_i)$ is to use relative weights proportional to $1/\mathrm{var}(d_i)$. Since the weights used in analytic reconstruction are strongly influenced by the need to avoid cone-beam and motion artifacts, the goal is here to modulate the incoming intensity of the x-ray beam such that the noise variance of the measured line-integrals matches the statistical optimum.

The line integrals d_i are determined from the measured intensity y_i and the non-attenuated intensities, aka the blank scan, b_i according to

$$d_i = -\log(y_i/b_i)\,. \tag{1}$$

Using Gaussian error propagation, the relation $\mathrm{var}(y_i) = y_i$ for Poisson distributed noise, and the assumption that the variance of the blank scan b_i can be neglected leads to the relation

$$\mathrm{var}(d_i) = 1/\mathrm{var}(y_i)\,, \tag{2}$$

and the statistically best weights w_i for the data are

$$w_i \propto 1/\mathrm{var}(d_i) = \mathrm{var}(y_i)\,. \tag{3}$$

[1] Philips Technologie GmbH Forschungslaboratorien

Since the data d_i, which are averaged during reconstruction, are considered to be redundant data, it is reasonable to assume that all these data are approximately the same. In other words, approximately the same fraction f of photons is absorbed along the corresponding lines, which leads to the relation

$$w_i \propto \mathrm{var}(y_i) = y_i \approx f b_i \,. \tag{4}$$

Finally, note that up to here only relative weights were used in the argumentation. Normalization is performed during reconstruction and thus, the constant factor f can be dropped. In conclusion, if a reconstruction algorithm uses weights w_i for a set of redundant data, then it is desired that the non-attenuated beam intensity for the corresponding detector pixels accords with

$$b_i \propto w_i \,. \tag{5}$$

For aperture weighted wedge and most other helical cone-beam CT reconstruction algorithms, the normalization is performed for each voxel independently. Thus, the final normalized weight for a particular detector sample varies from object point to object point. However, the non-normalized aperture weighting function has the general feature of putting a larger weight to the central part of the detector. This leads also to a larger mean normalized weight for detector samples in the central part. Fig. 2 illustrates these mean normalized weights which are finally used for the detector values for an example of a helical scan with a pitch factor of 1 and a non-normalized aperture weighting function that is trapezoidal with a plateau width of 50%.

The desired spatial distribution of the non-attenuated beam intensity according to Eq. 5 for the mean weights can be easily achieved using a beam shaper. For an ideal x-ray source that irradiates isotropically in all directions, a beam shaper as illustrated in Fig. 3 results. The values on the z-axis in Fig. 3 are the desired line integral values of the beam shaper over the detector panel. Considering Teflon as a material for the beam shaper, the required thickness to achieve the maximum line integral value of approximately 2.5 is about 6 cm. Note that this beam shaper comes in addition to the conventional beam shaper that modulates the beam in fan direction.

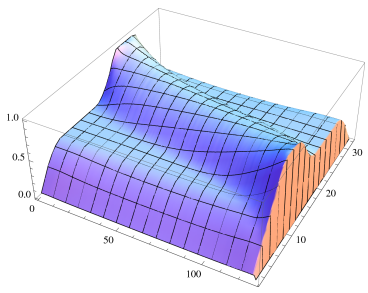


Fig. 2. Mean normalized weights for each pixel on the detector used during aperture weighted wedge reconstruction. The x-axis going to the right corresponds to the fan-angle direction, the y-axis going to the back corresponds to the cone-angle direction.

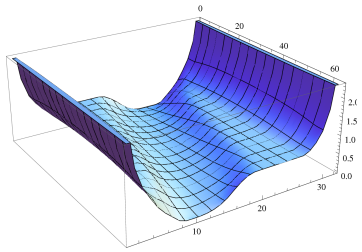


Fig. 3. Illustration of the thickness of a beam shaper that creates the desired photon statistics. The z height of the surface indicates the desired line-integral value of the beam shaper. The conventional angular modulation of the beam shaper that will be maintained is not shown.

B. Iterative Reconstruction Algorithms

As a prototype for iterative reconstruction a penalized maximum likelihood (ML) is considered. Specifically, the objective function

$$\Psi(x) = -L(x) + \beta R(x) \tag{6}$$

is minimized, where x denotes the image, $-L(x)$ the negative log-likelihood of the image and the roughness penalty $R(x)$ has the form

$$R(x) = \frac{1}{2} \sum_{j=1}^{p} \sum_{k \in \mathcal{N}_j} w_{jk} \psi(\Delta_{jk}) \,, \tag{7}$$

where $\mathcal{N}_j$ represents the neighborhood of the voxel x_j and ψ is a symmetric and convex function. Furthermore, $\Delta_{jk} = x_j - x_k$. The additional weights w_{jk} are used to reduce the penalty with increasing distance.

For the maximization of the objective function $\Psi(x)$, the ordered subset (OS) version of the separable paraboloid surrogate (SPS) algorithm [15, 16] is considered. Let M denote the number of subsets, $\hat{x}_j^{(n)}$ the image after n sub-iterations, and $\hat{l}_i = \mathrm{FP}(\hat{x}_j^{(n)})$ the forward projection of the image $\hat{x}_j^{(n)}$. Furthermore, let $\mathrm{BP}_{S_m}(\cdot)$ denote the back-projection of all projection values within a subset S_m and $\omega_\psi(t) = \dot{\psi}(t)/t$. Then the OS-SPS update step can be written as

$$\hat{x}_j^{(n+1)} = \hat{x}_j^{(n)} + \frac{M \mathrm{BP}_{S_m}(b_i e^{-l_i}) - \beta \sum_{k \in \mathcal{N}_j} w_{jk} \dot{\psi}(\Delta_{jk})}{\mathrm{BP}(y_i \mathrm{FP}(1)) + 2\beta \sum_{k \in \mathcal{N}_j} w_{jk} \omega_\psi(\Delta_{jk})}$$

In [11] and [12] a heuristic aperture weighting was introduced to iterative reconstruction in order to mitigate artifacts that arise if iterative reconstruction is applied to clinical data. Along the same line, the OS-SPS algorithm is modified here to

$$\hat{x}_j^{(n+1)} = \hat{x}_j^{(n)} + \frac{M \mathrm{BP}_{S_m}(a_i b_i e^{-l_i}) - \beta \sum_{k \in \mathcal{N}_j} w_{jk} \dot{\psi}(\Delta_{jk})}{\mathrm{BP}(a_i y_i \mathrm{FP}(1)) + 2\beta \sum_{k \in \mathcal{N}_j} w_{jk} \omega_\psi(\Delta_{jk})}$$

where the additional weights a_i are weights which are 1 in the center of the detector and smoothly approach 0 near the upper and lower edge of the detector. This aperture weighted algorithm is denoted as AW-OS-SPS.

Looking at the way these aperture weights enter the update formula, it is clear that the same effect of down-weighting certain detector values can be achieved using the OS-SPS algorithm, if the blank scan is modulated accordingly.

III. Results

We simulated helical scans of a 32 row scanner with a pitch factor of 1 and 1160 views per rotation. For the noise analysis, a 330 mm water phantom was used. For the analysis of the sensitivity to motion, the forbild head phantom was used, where an additional moving high contrast object was inserted. Poisson distributed noise was added to the projection data. For each beam shaper, i.e., the standard beam shaper and the proposed beam shaper, the total number of photons in the beam after passing the beam shaper was the same ($1.5 \cdot 10^{10}$ photons per projection). Taking into account that the additional beam shaper does not introduce a modulation in the fan-angle direction (i. e., the mean number of photons is actually kept constant for each detector column), it can be concluded that the dose is the same for both beam shapers. As a consequence, the photon flux towards the central part of the detector is higher if the proposed beam shaper is used, which will lead to the nice additional effect that the SNR becomes more homogeneous. The same beam shaper (as illustrated in Fig. 3) is used for the analysis of analytic and iterative reconstructions although for iterative reconstruction a simpler shape would be sufficient since no optimization process is needed to obtain the shape.

A. Noise Analysis

The noise level is determined spatially resolved by using an ensemble of 10000 statistically independent noise realizations reconstructed with aperture weighted wedge. Results for the noise level (square root of the noise variance) are shown in Fig. 5. The right image in Fig. 5 indicates the increase of the SNR by the proposed method. Evidently, the increase has a strong spatial dependence. Actually, it appears that the increase is biggest in the areas where the original method has the worst SNR. This can be understood by the different illumination ranges of voxels in a helical acquisition: The illumination range for the image plane is shown in Fig. 4. The illumination range for voxels in the left half of the image is shorter than on the right half, resulting in a higher noise level in the left half. These voxels are illuminated only little more 180 degrees. In other words, most of the time, these voxels are projected onto the PI-window. The PI-window is – as visible in Fig. 2 – the detector area where the mean data weights are largest and where correspondingly the proposed method increases the SNR on the projection. Therefore, the SNR gain is in particular high for these voxels.

B. Motion Artifacts

Fig. 6 and 7 illustrate the effect of aperture weighting on image quality in the presence of inconsistencies in the projection data due to motion for the wedge method (Fig. 6) and for the penalized ML reconstruction (Fig. 7). The Huber penalty with $\delta = 0.2$ HU is used as a penalty. The left images in Fig. 6 and 7 show reconstruction results obtained by using the standard beam shaper and SNR optimized data weighting (i.e., equal weighting in the wedge reconstruction and statistically correct weighting for the ML reconstruction). Evidently, both reconstructions show motion artifacts. The motion artifacts are less severe in the ML reconstruction because motion artifacts are quickly changing in z-direction and are therefore suppressed by the roughness penalty. Additionally, the analytic reconstruction shows severe cone-beam artifacts. All these artifacts are suppressed using aperture weighting (middle images in Fig. 6 and 7). The artifacts are suppressed equally well if the proposed beam shaper is used, which increases SNR in analytic reconstruction as shown in the previous subsection, and increases also the SNR in iterative reconstruction since all data are weighted statistically correct using the OS-SPS algorithm. However, we are reluctant to make a quantitative statement based on a single noise realization due to the complicated noise properties of iterative reconstruction with edge-preserving regularization [17].

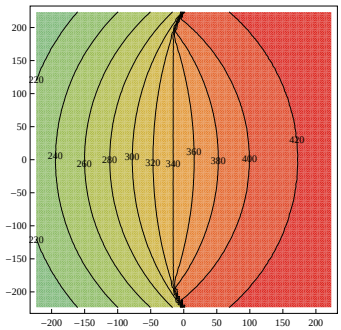


Fig. 4. Illumination map for the analyzed slice. Contour lines indicate the illumination range in degrees. Regions of a short illumination interval correspond to regions with higher noise level in the standard acquisition (Fig. 5 left).

IV. Summary

We proposed the general idea of beam shaping in accordance with the mean weight with which the corresponding detector pixel will contribute to the reconstructed images. For a particular reconstruction method, namely the aperture weighted wedge reconstruction, the proposed method provides an SNR gain of 12% at the iso center and also an average SNR gain of 12% over the full phantom extend (i. e., within the central area of 330 mm diameter). In consequence, the SNR of the conventional case can be preserved using 25% less dose. A second positive effect is that the SNR becomes more symmetric.

The method can be readily applied to any other cone-beam reconstruction method, e. g., the methods described in [4–7, 9, 10]. It must be acknowledged that the optimal shape depends on the helical pitch since the mean data weighting depends on the pitch. However, we found in further experiments, which are not shown here, that the beam shaper that was optimized for a pitch of 1 leads to a substantial SNR gain for a large range of pitches. Furthermore, it should be noted that the possible dose savings depend on the weighting scheme used in practice.

Additionally, the proposed beam shaper increases the robustness of iterative reconstruction algorithms with respect to motion without the introduction of heuristic data weighting schemes.

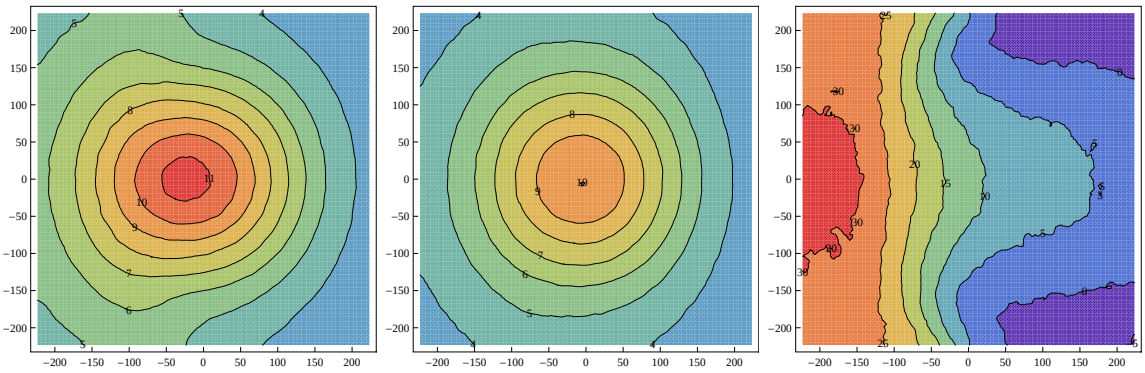


Fig. 5. Noise level in a 450 mm field of view obtained by an acquisition with the standard beam shaper (left) and the optimal beam shaper (middle). The right image shows the SNR gain in %.

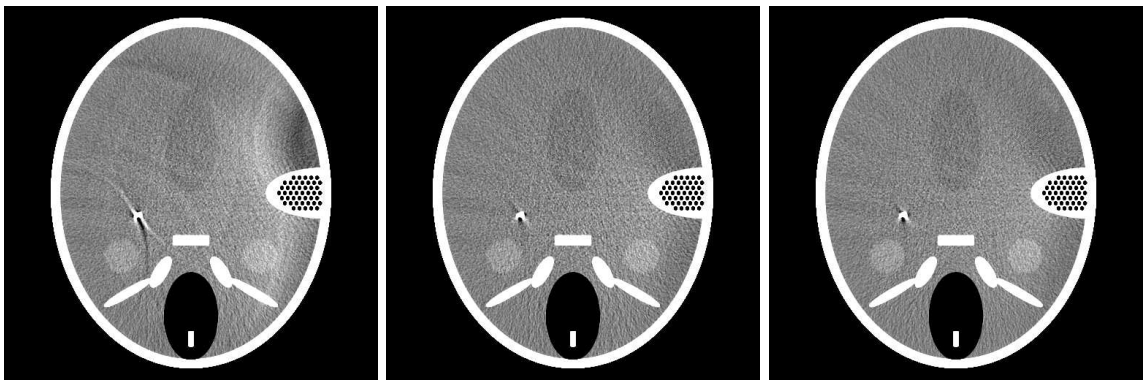


Fig. 6. Illustration of the effect of the proposed beam shaper on analytic reconstruction results in the presence of motion: Left: Acquisition using a standard beam shaper and reconstruction with SNR optimal weighting. Middle: Acquisition using a standard beam shaper and reconstruction with heuristic aperture weighting. Right: Acquisition using the same heuristic aperture weighting and a matched beam shaper.

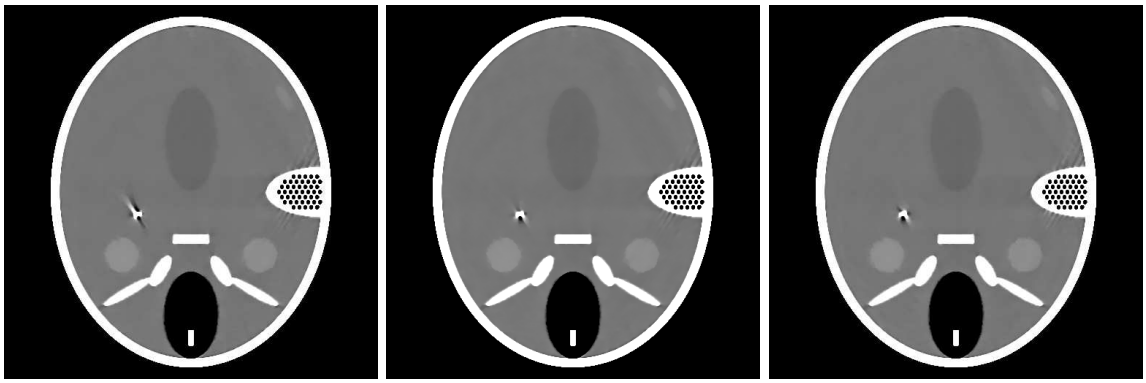


Fig. 7. Illustration of the effect of the proposed beam shaper on iterative reconstruction results in the presence of motion: Left: Acquisition using a standard beam shaper and reconstruction with statistically correct data weighting (OS-SPS). Middle: Acquisition using a standard beam shaper and reconstruction with heuristic aperture weighting (AW-OS-SPS). Right: Acquisition using the proposed beam shaper and reconstruction with statistically correct data weighting (OS-SPS).

References

[1] D. J. Brenner and E. J. Hall. *N Engl J Med*, 357(22) (2007) 2277.

[2] A. B. de Gonzales et al. *Arch. Intern. Med.*, 169(22) (2009) 2071.

[3] H. Kudo et al. In *Nuclear Science Symposium Conference Record*, 3211 – 3215. IEEE, 2003.

[4] K. Stierstorfer et al. *Phys. Med. Biol.*, 49 (2004) 2209.

[5] K. Taguchi et al. *Phys. Med. Biol.*, 49(11) (2004) 2351.

[6] M. Kachelrieß et al. *Med. Phys.*, 31(6) (2004) 1623.

[7] G. Shechter et al. *Med. Phys.*, 31(8) (2004) 2230.

[8] D. Heuscher et al. *Phys. Med. Biol.*, 49(11) (2004) 2219.

[9] T. Köhler et al. *IEEE Trans. Med. Imag.*, 25(7) (2006) 882.

[10] A. Katsevich et al. *IEEE Trans. Med. Imag.*, 28(7) (2009) 982.

[11] A. Ziegler et al. Method and device for the iterative reconstruction of cardiac images, 2005. US patent 7596204B2.

[12] K. Zeng et al. In *Proceedings of the Fully 3D*, 242 – 245. 2009.

[13] J. Hsieh. *Computed Tomography - Principles, Design, Artifacts,and Recent Advances*. SPIE, Bellingham, Washington, 2003.

[14] P. Koken and M. Grass. *Phys. Med. Biol.*, 51 (2006) 3433.

[15] J. A. Fessler. In J. M. Fritzpatrick and M. Sonka (eds.), *Handbook of Medical Imaging, Vol. 2*, chap. 1, 1 – 70. SPIE Press, Bellingham, WA, 2000.

[16] H. Erdogan and J. A. Fessler. *IEEE Trans. Med. Imag.*, 18(9) (1999) 801.

[17] T. Köhler and R. Proksa. In *Proceedings of the Fully 3D*, 263 – 266. 2009.

Iterative correction of beam hardening artifacts in CT

Katrien Van Slambrouck, Gert Van Gompel, Michel Defrise, K. Joost Batenburg,
Jan Sijbers and Johan Nuyts

We present a method to reduce beam hardening artifacts in computed tomography. We started from the linearisation pre-correction method of Van Gompel et al. [1] and modified this model to a least squares reconstruction method, MGRA. This modification makes it more easy to interpret the physical meaning of the reconstructed values. MGRA, however, is computationally expensive and therefore, we developed an acceleration step based on a FBP density update, MFBP. All three methods have been evaluated with simulations and phantom measurements.

I. Introduction

The polychromatic nature of the X-rays in computed tomography (CT) causes beam hardening, which results in cupping and streak artifacts. Several approaches have been applied to reduce these artifacts with good results. In a first approach one uses statistical reconstruction with a polychromatic model [2], [3], [4]. The main drawback of such a model is the high computation time. Another possibility is to use a linearisation method. These methods are effective but often there are restrictions. Some methods can only be applied on a limited class of objects [5] or for limited geometries [6]. Knowledge about material properties is often required [7], [8] and also a calibration scan can be necessary [9].

Recently, Van Gompel et al. [1] and Krumm et al. [10] proposed linearisation methods for which neither spectrum nor material properties need to be known. The methods modify the measured sinogram by adding the differences of a monochromatic and a polychromatic simulation. In [1], the polychromatic sinogram is simulated using a parameterized polychromatic model that is fitted to the measured data. For the monochromatic simulation a set of monochromatic attenuation coefficients needs to be fitted based on the simulated polychromatic sinogram.

We propose a modified version of the algorithm of [1], which treats the problem as a reconstruction task rather than as a sinogram pre-correction task. For that purpose, the density in each voxel is introduced as an extra parameter set. These densities and the material characteristics are estimated from the data.

In section II the method of [1] is briefly explained and the modified model is presented. Section III describes the phantom we used to evaluate the model, the results are given in section IV. Finally in section V the results are discussed and the different algorithms are compared.

K. Van Slambrouck and J. Nuyts are with the Dept. of Nuclear Medicine, K.U.Leuven, Leuven, Belgium. G. Van Gompel is with the Dept. of Radiology, UZ Brussel, Brussels, Belgium. M. Defrise is with the Dept. of Nuclear Medicine, V.U.Brussel, Brussel, Belgium. K.J. Batenburg and J. Sijbers are with the VisionLab, U. Antwerp, Antwerp, Belgium

This work was financially supported by the SBO-project QUANTIVIAM (060819) of the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen).

II. Models

The original model and the newly proposed algorithms are all based on the hypothesis that there are N different materials present in the object with $n = 1, ..., N$ and that each voxel can only contain one material defined by a binary variable $s_{n,j}$ taking the value $s_{n,j} = 1$ if voxel j contains n and $s_{n,j} = 0$ otherwise.

A. Sinogram pre-correction method (GVG)

In the model of [1] the measured intensity I_p^{meas} is approximated by

$$I_p^{\text{sim}}(i) = I_0 \sum_{e=1}^{E} I_e^F \exp\left(- \sum_{n=1}^{N} \mu_{n,e} \sum_j l_{ij} s_{n,j} \right) \tag{1}$$

where i denotes the projection ray, E is the number of energy bins that will be used to simulate the spectrum, l_{ij} is the intersection length of ray i with pixel j and $I_e^F = I_e / I_0$ is the fraction of the total spectrum corresponding to energy bin e, with $I_0 = \sum_e I_e$, the intensity in absence of the object. The polychromatic attenuation, estimated from the parameters $\{\boldsymbol{\mu}, \boldsymbol{I}^F, \boldsymbol{s}\}$

for ray i is then given by $A_p^{\mathrm{sim}}(i) = \ln\left(\frac{I_0(i)}{I_p^{\mathrm{sim}}(i)}\right)$. The corresponding estimate from the measurement is $A_p^{\mathrm{meas}}(i) = \ln\left(\frac{I_0(i)}{I_p^{\mathrm{meas}}(i)}\right)$.

The measurement is corrected by iteratively adding the difference between the simulated monochromatic and polychromatic attenuation:

$$A_m^{\mathrm{cor},w}(i) = A_p^{\mathrm{meas}}(i) + \left(A_m^{\mathrm{sim},w}(i) - A_p^{\mathrm{sim},w}(i)\right) \quad (2)$$

where w is the iteration number and $A_m^{\mathrm{cor},0} = A_p^{\mathrm{meas}}$. The monochromatic simulation is given by

$$A_m^{\mathrm{sim}}(i) = \sum_{n=1}^{N} \mu_n^{\mathrm{ref}} \sum_j l_{ij} s_{n,j}$$

with μ_n^{ref}, the reference attenuation coefficient for material n. In each iteration, $A_m^{\mathrm{cor},w}$ is reconstructed with filtered backprojection (FBP), and segmented with thresholding to produce a new estimate of $\boldsymbol{s}$. Then $\boldsymbol{\mu}$, $\boldsymbol{I}^F$ and $\boldsymbol{\mu}^{\mathrm{ref}}$ are estimated by minimizing (3) and (4).

$$\Phi(\boldsymbol{\mu}, \boldsymbol{I}^F, \boldsymbol{s}) = \sum_i \left(\ln\left(\frac{I_p^{\mathrm{meas}}(i)}{I_0(i)}\right) - \ln\left(\sum_{e=1}^{E} I_e^F \exp\left(-\sum_{n=1}^{N} \mu_{n,e} \sum_j l_{ij} s_{n,j}\right)\right)\right)^2 \quad (3)$$

$$\begin{aligned}\Psi(\boldsymbol{\mu}, \boldsymbol{s}) &= \sum_i \left(A_m^{\mathrm{sim}}(i) - A_p^{\mathrm{sim}}(i)\right)^2 \\ &= \sum_i \left(\sum_{n=1}^{N} \mu_n^{\mathrm{ref}} \sum_j l_{ij} s_{n,j} \right. \\ &\quad \left. - \ln\left(\sum_{e=1}^{E} I_e^F \exp\left(-\sum_{n=1}^{N} \mu_{n,e} \sum_j l_{ij} s_{n,j}\right)\right)\right)^2\end{aligned} \quad (4)$$

In [1] this method was applied on piecewise uniform objects with three materials. It was found that the use of three energy bins is sufficient to make an appropriate reconstruction. We will refer to this model as the GVG model.

B. Modified model (MFBP and MGRA)

The major difference between the modified model and GVG is that a density d_j is assigned to each voxel in addition to the material index $s_{n,j}$ so that the cost function (3) becomes:

$$\tilde{\Phi}(\boldsymbol{\mu}, \boldsymbol{I}^F, \boldsymbol{d}, \boldsymbol{s}) = \sum_i \left(\ln\left(\frac{I_p^{\mathrm{meas}}(i)}{I_0(i)}\right) - \ln\left(\sum_{e=1}^{E} I_e^F \exp\left(-\sum_{n=1}^{N} \mu_{n,e} \sum_j l_{ij} d_j s_{n,j}\right)\right)\right)^2 \quad (5)$$

where d_j denotes the density in pixel j. To minimize this cost function we need to estimate the attenuation coefficients $\boldsymbol{\mu}$, the fractional energies $\boldsymbol{I}^F$, the density $\boldsymbol{d}$ and the segmentation $\boldsymbol{s}$. Based on the estimated parameters we produce the following image for visual inspection:

$$R_w = d^w \sum_n \mu_n^{\mathrm{mono},w} s_n^w \quad (6)$$

where $\boldsymbol{\mu}^{\mathrm{mono}}$ are monochromatic attenuation coefficients which can be estimated by $\boldsymbol{\mu}^{\mathrm{ref}}$ as in GVG, but also the mean attenuation can be chosen. For an easy comparison we choose $\boldsymbol{\mu}^{\mathrm{ref}}$. All parameters are iteratively updated.

Algorithm

Each iteration of the modified algorithms consists of the following steps:

1) Segmentation: The current image estimate R_w is segmented to update the material indices $s_{n,j}$. This is accomplished by a threshold method. The thresholds are adapted in such a way that the new thresholds lower the cost function (5). This is done by a gradient descent algorithm.

2) Density: As a second step we update the density. Two approaches for the density step are proposed. The first one uses a gradient descent step ensuring that $\tilde{\Phi}(\boldsymbol{\mu}, \boldsymbol{I}^F, \boldsymbol{d}, \boldsymbol{s})$ does not increase. The second one is inspired by the iterative filtered-backprojection method. A new image is obtained as $R_w' = R_w + \mathrm{FBP}(A_p^{\mathrm{meas}} - A_p^{\mathrm{sim}})$ and the updated densities are then directly obtained by setting

$$R_w' = d^{w+1} \sum_{n=1}^{N} \mu_n^{\mathrm{mono},w} s_n^{w+1}.$$

In the following text the modified model with FBP density update will be called the MFBP model. In case a gradient descent update is used we will call it the MGRA model. The gradient descent update contains a projection and backprojection, and can be accelerated with ordered subsets.

3) Attenuation and fractional energy: Updated attenuation coefficients and fractional energies are calculated by a gradient based minimization of the cost function (5) with constant density and constant segmentation.

4) Visualization: After estimation of the parameters, we calculate the monochromatic attenuation coefficients and update the visualization (6).

In the first iteration we segment the uncorrected FBP reconstruction, $\mathrm{FBP}(A_p^{\mathrm{meas}})$, and we assume the density to be equal to one everywhere. As was done in

[1], we chose to use three energy bins and we accelerated the method by using a downsampled sinogram together with a Gaussian filter till the following stop condition is met:

$$\frac{\tilde{\Phi}^{w-1} + \tilde{\Phi}^{w}}{\tilde{\Phi}^{w-3} + \tilde{\Phi}^{w-2}} > 0.9. \qquad (7)$$

Secondly, we iterate again till stop condition (7) but without a filter. Finally, we upsample the result and do two full size iterations.

III. Materials

We tested this new model on simulations and on a phantom measurement. Here, only the phantom measurements will be presented. The *bean phantom* (3cm x 1cm) is depicted in figure 1(a). It consists of four materials: air, plexiglass, white spirit, and water. The density of water is only slightly higher than the density of plexiglass, the main material, whereas the density of white spirit is substantially lower. The algorithms were applied with $N = 3$, consequently, the water part will not be segmented separately for this phantom.

A Skyscan 1172 μCT at 60 kV was used to obtain the data. Hardware filtering and the beam hardening correction option were turned off. The central slice of the cone beam geometry was rebinned to a parallel beam geometry. The parallel beam sinogram consists of 300 views, equally spaced with an angular range $[0, \pi[$ and 1000 radial samples. We used downsampling factor 4 for the radial samples and 2 for the views. The Gaussian filter had a standard deviation of 1.5 pixels. FBP reconstruction is done using a Hamming filter with cutoff frequency 0.5.

IV. Results

Three algorithms are compared: the original GVG pre-correction model, and the two proposed reconstruction algorithms: the modified model with FBP density update (MFBP) and the modified model with the gradient density update (MGRA).

Figure 1 shows the uncorrected FBP reconstruction of the bean phantom and the results obtained with GVG, MFBP and MGRA. The evolution of the cost function $\tilde{\Phi}(\boldsymbol{\mu}, \boldsymbol{I}^F, \boldsymbol{d}, \boldsymbol{s})$ (5) as a function of the iteration number is shown in figure 2.

V. Discussion

The GVG pre-correction method is fast and effective. However, it is difficult to assign a physical meaning to the reconstructed values. Therefore,

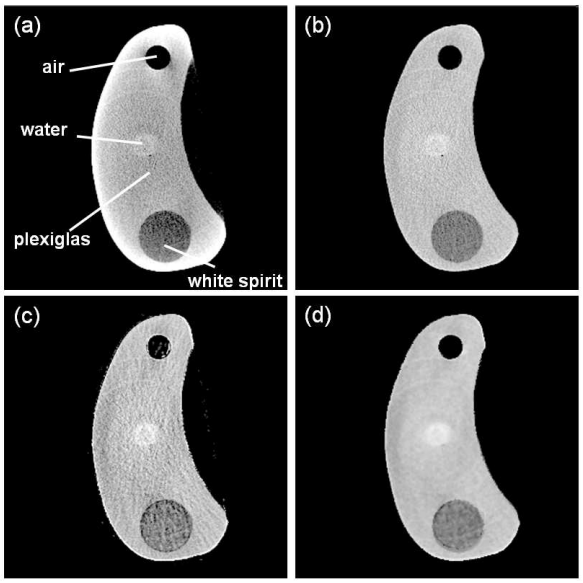


Fig. 1. **Bean phantom**. Uncorrected FBP reconstruction (a), correction with GVG (b), correction with MFBP (c) and correction with MGRA (d).

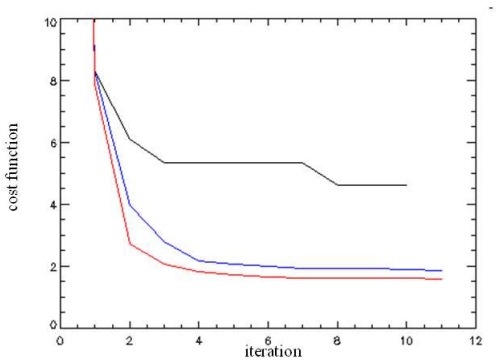


Fig. 2. Evolution of the cost function $\tilde{\Phi}(\boldsymbol{\mu}, \boldsymbol{I}^F, \boldsymbol{d}, \boldsymbol{s})$ as a function of the iteration number. GVG: black line, MFBP: blue line and MGRA: red line.

we proposed two related iterative reconstruction algorithms. The algorithms are based on an extension of the original cost function $\Phi(\boldsymbol{\mu}, \boldsymbol{I}^F, \boldsymbol{s})$ (4) to $\tilde{\Phi}(\boldsymbol{\mu}, \boldsymbol{I}^F, \boldsymbol{d}, \boldsymbol{s})$ (5). The main difference between both is the introduction of a material density, which explicitly models the deviation from the segmented image. In the original method, this deviation is incorporated in an implicit way, by reconstructing the sum of the measured sinogram and a correction term deduced from the segmentation. In the MGRA method, each operation in the iteration attempts to decrease the cost function, and is guaranteed not to increase it. A

monotonic decrease with effective convergence was observed in all simulations. Unfortunately, the cost function is not convex, and convergence to the global maximum is not guaranteed.

For GVG and MFBP, the cost function does not decrease systematically at every iteration, but effective convergence was still observed. The value of the cost function obtained with the reconstruction methods (MGRA and MFBP) is usually lower than the one obtained with the GVG method. It is difficult to compare the corresponding reconstructions for two reasons. First, the meaning of the reconstructed values is not the same in the pre-correction and the reconstruction methods. Second, the different algorithms seem to converge to a local minimum as intended, but the local minimum may be different. The difference in image quality is then due to the anecdotal (image dependent) convergence trajectory and not to fundamental differences between the methods. As illustrated in figure 1, all three methods give similar results.

MGRA has the best theoretical foundation but it is significantly slower than GVG and MFBP and hence cannot be used in daily practice. Especially the density update step takes a lot of computation time. Therefore, we introduced the faster FBP update step in MFBP. This change results in a strong acceleration. Monotonic decrease of the cost function $\tilde{\Phi}(\boldsymbol{\mu}, \boldsymbol{I}^F, \boldsymbol{d}, \boldsymbol{s})$ is no longer guaranteed, but good results were obtained in simulations and the first experiment on real data (figure 2). However, the MFBP correction method is more sensitive to noise compared to GVG and MGRA.

Originally, Otsu [11] thresholding was used for the segmentation in the GVG method. However, this approach did not always produce a decrease of the cost function. Here, the thresholds were determined by minimizing the cost function, which improved the convergence behavior. The disadvantage of this method is the need to calculate several projections while updating the segmentation. This is a time consuming job and often the changes to the segmentation are small. This step could be strongly accelerated, by using a projector that only projects the differences between the old and new segmentations, as was done in [12].

VI. Conclusion

The advantages of the modified algorithm are:

- The problem is defined as an explicit optimization problem, which makes it easier to interpret and analyze the reconstructed image. The reconstructed values correspond to the parameters of an (approximate) model, whereas in the original method, it is more difficult to assign a physical meaning to the reconstructed image values.
- The introduction of the density offers the possibility to use this method for objects with materials which have small density differences. This way, this important feature of the method of [1] could be preserved.

The limitations of the proposed algorithm are

- The modified model MGRA needs more computation time, while the results are not very different from those of the original algorithm.
- The MGRA method is too slow for routine use. However, it is useful as a reference method when evaluating the performance of the two related approximation methods.

References

[1] Van Gompel G., PhD thesis: *Towards accurate image reconstruction from truncated X-ray CT projections*, Antwerp 2009.

[2] B. De Man, J. Nuyts, P. Dupont, G. Marchal, P. Suetens."An iterative maximum-likelihood polychromatic algorithm for CT." *IEEE Trans Med Imaging*, 2001; 20 (10): 999-1008.

[3] I. Elbakri and A. Fessler. "Statistical Image Reconstruction for Polyenergetic X-Ray Computed Tomography." *IEEE Transactions on Medical Imaging*; 2002, 21 (2):89-99.

[4] N. Menvielle, Y. Goussard, D. Orban, G. Soulez. "Reduction of beam-hardening artifacts in X-ray CT." *Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference Shanghai*, China, September 1-4, 2005

[5] H. Goa, L. Zhang, Z. Chen, Y. Xing and L.Shuanglei. "Beam hardening correction for middle-energy industrial computerized tomography." *IEEE Transactions on Nuclear Science*; 2006, 53 (5):2796-2807.

[6] X. Mou, S. Tang and H. Yu. "Comparison on beam hardening correction of CT based on HL consistency and normal water phantom experiment." *SPIE medical Imaging*; 2006: 6318V

[7] V. Vedula and P. Munshi. "An improved algorithm for beam hardening corrections in experimantal X-ray tomography." *NDT&E International*; 2008, 41: 25-31.

[8] Y. Censor, T. Elfving and G.T. Herman. "A method of iterative data refinement and its applications." *Math. Meth. in the Appl. Sci.*; 1985, 7:108-123.

[9] R. Grimmer, C. Maass, M. Kachelreiss. "A new method for cupping and scatter precorrection for flat detector CT." *IEEE Nuclear Science SYmp. Conf. Rec.*; 2009: 3517-3522.

[10] M. Krumm , S. Kasperk and M. Franz. "Reducing non-linear artifacts of multi-material objects in industrialised 3D computed tomography." *Nondestructive Testing and Evaluation*; 2008, 41: 242-281.

[11] N. Otsu. "A threshold selection mehtod from gray-level histograms." *Trans. on Systems, Man, and Cybernetics*; 1979, 5 (9): 62-66.

[12] K. J. Batenburg and J. Sijbers. "Optimal threshold selection for tomogram segmentation by projection distance minimization." *IEEE Trans. Medical Imaging*; 2009, 28 (5):676-686.

A Reconstruction Technique for Low Dose CT using Equally-Sloped Tomography

Benjamin P. Fahimian†, Yunzhe Zhao, Russell Fung, Michael F. McNitt-Gray, John J. DeMarco, Gregory Chu, Maryam Khatonabadi, and Jianwei Miao†

***Abstract*—We present an iterative Fourier based reconstruction technique, which through the use of physical constraints, mathematical regularization, and the oversampling method, allows for accurate tomographic reconstruction from low flux and undersampled projection data. The method, termed Equally-Sloped Tomography, has previously been applied to parallel beam modalities such as electron and x-ray phase contrast tomography for the purpose of limited angle tomography. Here, we develop a generalization to fan beam geometry, and incorporate advanced regularization constraints, such as the Non-Local Means Total Variational model. Experimentally, we implement the method on the Siemens Sensation 64 scanner. Using a series of image quality phantoms, we evaluate the resulting image quality as a function of source flux; quantitative comparisons to current scanner reconstructions are made through the measurements of resolution, signal to noise ratio, and contrast to noise ratio. As the algorithm guides the reconstruction solution to the less noisy state in a manner strictly consistent with measured projections and constraints, we explore the possibility of radiation dose reduction through the reduction of flux. The experimental phantom studies indicate that comparable reconstructions, relative to current scanner reconstructions, may be achievable with 65-80% lower x-ray flux.**

I. Introduction

Due to the tomographic requirement of multiple projections from multiple directions, transmission tomographic imaging techniques utilizing ionizing radiation, such as medical x-ray CT, are inherently high dose imaging techniques. Two general methods for reducing the radiation dose in tomographic modalities include undersampling the number of projections and or reducing the source flux per projection. However under such conditions, the accuracy of the reconstruction is compromised due to the violation of the Nyquist sampling requirement when the projections are undersampled, and the presence of heavy noise when the source flux is significantly reduced [1, 2]. The development of more sophisticated reconstruction algorithms capable of solving for a portion of the data and guiding the solution to a less noisy state constitute one method for dose reduction [3, 4]. Recent simulation and experimental results in electron tomography and synchrotron x-ray CT have demonstrated that the technique of Equally-Sloped Tomography (EST) , an exact iterative Fourier based reconstruction algorithm, significantly outperforms Filtered Back Projection (FBP) and algebraic reconstruction technique, and is capable of producing comparable reconstructions with 60-70% fewer projections [5-8]. Here, we generalize and implement the method on a medical CT scanner, and subsequently study the possibility of radiation dose reduction through the reduction of flux in particular.

B. P. Fahimian, Prof. M. F. McNitt-Gray, Prof. J. J. DeMarco, G. Chu, and M. Khatonabadi are with the Biomedical Physics Interdepartmental Program, David Geffen School of Medicine at UCLA, Los Angeles, CA 90095-1770 USA.

Prof. J. Miao, Y. Zhao, and Dr. R. Fung are with Department of Physics and Astronomy at UCLA and the California NanoSystems Institute, Los Angeles, CA 90095-1770 USA.

This work was partially supported by UC Discovery / TomoSoft Technologies Grant # ITL07-10166 (PI: J. Miao).

†Correspondence should be addressed to J. Miao (miao@physics.ucla.edu) and B. P. Fahimian (bfahimian@mednet.ucla.edu).

II. Methods

A. Fourier Formulation

The Fourier slice theorem, which equates the 1D Fourier transform of a projection to a slice of the 2D Fourier transform of the object at the corresponding angle, forms the foundation of tomographic imaging. While the theorem is conceptually elegant, its direct computerized application is problematic as the aggregation of projection data forms a polar data set in frequency space, while conventional FFT algorithms operate on a Cartesian point set, leading to degrading Fourier space interpolation for direct implementation [1, 2]. Although it is believed that no direct exact fast Fourier transform algorithm can be constructed between the polar and Cartesian grids, recently the existence of a direct exact fast Fourier transform algorithm and its inverse between an oversampled pseudopolar grid and the Cartesian grid, termed as the Pseudopolar Fast Fourier Transform (PPFFT), has been proved [9, 10]. As depicted in Fig. 1., for a $N \times N$ Cartesian grid, the corresponding pseudopolar grid is defined by a set of $2N$ lines, each line consisting of $2N$ grid points mapped out on N concentric squares, with each successive line changing by an equal slope increment of as opposed to a fixed equal angled increment as in the polar grid. Unlike the polar grid, the distance between sampling points on the individual $2N$ lines of the pseudopolar grid vary from line to line. The fractional Fourier transform (FrFT) can be used to vary the output

sampling distance of the Fourier transform. The 1D FrFT is defined as, $\mathcal{F}_\alpha f = F_l = \sum_{k=-n}^{n-1} f_k \exp(-i\alpha kl\pi / n), \quad -n \le l \le n,$ which is equivalent to the standard discrete Fourier transform but with a factor of α in the exponent [11]; the FrFT reduces to the forward Fourier transform when $\alpha = 1$ and to the inverse Fourier transform when $\alpha = -1$. Through the use of the FrFT, an analog of the Fourier Slice Theorem can be reformulated for the pseudoplar grid to be: the 1D FrFT of a projection at an angle θ is equivalent to a slice of the 2D pseudopolar Fourier transform of the image, $f(\mathbf{r})$, at the same angle θ in the frequency domain.

B. Iterative Reconstruction Algorithm

Since the objective is to limit the flux and or the number of projections for the purposes of limiting the radiation dose, an iterative algorithm is necessary to recover the resulting missing projection data and guide the solution to a less noisy state through regularization. In a manner similar to the oversampling Fourier algorithm that was used to solve the phase problem in diffraction imaging [12, 13], the reconstruction algorithm we introduce consists of iterations between object and Fourier space, with constraints enforced in each iteration. The parallel beam formulation is described first, as the fan beam formulation in section II.C is based upon it. For parallel modalities, a characteristic of the EST methodology is that, if projections are acquired at equal slope intervals, they can be directly mapped onto the pseudopolar grid, thereby eliminating all interpolations in tomographic reconstruction as first suggested by [7].

The iterative algorithm is initialized by padding the projections, followed by the appropriate FrFFT to map the projection on the corresponding line on the oversampled pseudopolar grid [7]; the oversampling results in the sample in the object domain to be surrounded by a region of mathematical zeros, termed the support, which provide additional constraints that aid guide the iterative process. The algorithm is initiated by placing the measured projection data onto the Fourier domain using the FrFT. The *j*th iteration of the algorithm can then be considered in four steps as depicted in Fig. 1b. First, the PPFFT^{-1} is applied to the Fourier space data to obtain an object space image $f_j'(\mathbf{r})$. Second, a new object $f_j(\mathbf{r})$ is obtained through the enforcement of three constraints which include support, positivity, and regularization of the object. In this study, we utilized the Non-Local Means Total Variational model, which represents a relatively recent improvement in edge preserving TV optimization [14-16]. Third, the forward PPFFT is applied to the modified image to obtain set of calculated slices in the Fourier domain. Fourth, the Fourier space data is updated with the measured slices. The iterations are monitored by an error function, and the algorithm is automatically terminated if the error does not reach a minimum after a set number of iterations. The computation time for each iteration is comparable to a FBP reconstruction. With respect to this, we note that recently a gradient descent version of the EST algorithm, which will be implemented in future studies, has been developed that reduces the computation time for each iteration by up to 70% [17].

C. Experimental Implementation

The method was experimentally implemented on the Siemens Somatom Sensation 64 scanner. As the scanner employs fan beam geometry, a rebinning step was performed prior to initiating the algorithm in section II.B in order to transform the fan beam projections to a set of parallel projections along equally-sloped lines of the pseudopolar gird [18-20]. Since the scanner utilizes a flying focal spot technology to increase detector sampling, the raw projections were interlaced and corrected prior to rebinning [21]. All scans were performed under axial mode with the tube current modulation off and the voltage set to 120 kVp. Image quality was determined using the ACR CT Accreditation Phantom and Siemens EMMA phantom as a function of photon flux by systematically lowering the mAs from a maximum of 583 mAs. All scanner (FBP) reconstructions were performed with Spline interpolation for the back projection process; comparison to the multitude of other filters for FBP is beyond the scope of this paper, and accordingly the standard uncropped ramp filter in order to not degrade the resolution [20].

III. Results and Discussion

The FBP and EST reconstructions for material and resolution modules of the ACR phantom are presented in Fig. 2. Starting at the maximum flux setting of 583 mAs, the signal to noise ratio (SNR) of the FBP in the body of the phantom is observed to degrade from a value of 61.6 in Fig. 2a to a value of 18.8 in Fig 2b, when the flux is reduced to 50 mAs; the EST reconstruction (terminated after 21 iterations) at 50 mAs, as shown in Fig 2c, results in a SNR of 62.3 in the same region. To assess the effect of the method on resolution, the resolution module of the ACR Phantom was reconstructed for FBP at 583, FBP at 50 mAs and EST at 50 mAs, as shown in Fig. 3c-e. The results indicate no degradation of high contrast resolution which may be a concern when implementing regularized algorithms.

To evaluate the effect of the method on objects of varying contrast and size, reconstructions were performed on the EMMA phantom which contains a set of cylindrical inserts with diameters .3 cm, .5 cm, .7 cm, 1 cm, 2 cm, and with varying normalized electron density relative to solid water of 0.001, 1.48, 1.09, 1.17, which we refer to as Regions I-IV, respectively. Table 1 summarizes the SNR for the each region and respective reconstruction method, while Table 2 summarizes the Contrast to Noise Ratio (CNR) relative to the body. The results indicate that the SNR of the 50 mAs EST reconstruction (terminated after 41 iterations) exceeds both the 50 mAs and 583 mAs FBP by a mean factor of 3.1 and 1.3, respectively, and the CNR of the low EST reconstruction exceeds that of the FBP reconstruction by a factor of 2.8 and

1.2, respectively. However, it is noed that the visibility lowest contrast rod with the smallest diameter of .3 cm is arguable for the low dose reconstructions.

The analysis indicates that the technique results in comparable or higher values in a variety of image quality parameters relative to the FBP reconstruction with greater than a factor of 10 times the dose. These results are relative to the scanner reconstruction at maximum allowable flux of 583 mAs which is greater than the typical scan protocol. Due to complexities of quantifying image quality, especially when considering clinically relevance, it is difficult to determine a dose reduction factor. With that said, if one assumes that a typical scan is between 150-250 mAs, then comparable reconstructions may be possible with 65-80% lower dose using the methodology.

REFERENCES

[1] A. C. Kak, and M. Slaney, *Principles of Computerized Tomographic Imaging*: SIAM, 2001.

[2] J. Hsieh, *Computed Tomography: Principles, Design, Artifacts, and Recent Advances*, Bellingham, Washington: SPIE Press Monograph 2003.

[3] J. B. Thibault, K. D. Sauer, C. A. Bouman *et al.*, "A three-dimensional statistical approach to improved image quality for multislice helical CT," *Med Phys,* vol. 34, no. 11, pp. 4526-44, Nov, 2007.

[4] I. A. Elbakri, and J. A. Fessler, "Statistical image reconstruction for polyenergetic X-ray computed tomography," *IEEE Trans Med Imaging,* vol. 21, no. 2, pp. 89-99, Feb, 2002.

[5] E. Lee, B. P. Fahimian, C. V. Iancu *et al.*, "Radiation dose reduction and image enhancement in biological imaging through equally-sloped tomography," *J Struct Biol*, Aug 15, 2008.

[6] B. P. Fahimian, E. Lee, P. Cloetens *et al.*, "Development of Low Dose X-ray Phase Contrast CT." Submitted.

[7] J. Miao, F. Förster, and O. Levi, "Equally sloped tomography with oversampling reconstruction," *Phys. Rev. B*, no. 72, 052103, 2005.

[8] B. P. Fahimian, A. Chatziioannou, J. J. DeMarco *et al.*, "Dose Reduction in CT Using a Novel Fourier-Based Iterative Reconstruction Method," *Med. Phys.,* vol. 35, no. 6, 2008.

[9] A. Averbuch, R. R. Coifman, D. L. Donoho *et al.*, "A Framework for Discrete Integral Transformations I—The Pseudopolar Fourier Transform," *SIAM Journal on Scientific Computing,* vol. 30, no. 2, 2008.

[10] A. Averbuch, R. R. Coifman, D. L. Donoho *et al.*, "A Framework for Discrete Integral Transformations II—The 2D Discrete Radon Transform," *SIAM Journal on Scientific Computing,* vol. 30, no. 2, 2008.

[11] D. H. Bailey, and P. N. Swarztrauber, "The fractional Fourier transform and applications," *SIAM Review,* vol. 33, no. 3, 1991.

[12] K. S. Raines, S. Salha, R. L. Sandberg *et al.*, "Three-dimensional structure determination from a single view," *Nature,* vol. 463, no. 7278, pp. 214-217, 2010.

[13] J. Miao, P. Charalambous, J. Kirz *et al.*, "Extending the methodology of X-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens," *Nature,* vol. 400, pp. 342-344, 1999.

[14] A. Buades, B. Coll, and J. M. Morel, "A Review of Image Denoising Algorithms, with a New One," *Multiscale Modeling & Simulation,* vol. 4, no. 2, pp. 490-530, 2005.

[15] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, no. 60, pp. 259-268, 1992.

[16] G. Gilboa, and S. Osher, "Nonlocal operators with applications to image processing.," *Multiscale Modeling and Simulation,* vol. 7, no. 3, pp. 1005-1028, 2008.

[17] Y. Mao, B. P. Fahimian, S. Osher *et al.*, "Development and Optimization of Tomographic Reconstruction Algorithms Utilizing Equally-Sloped Tomography " *IEEE Trans Image Process.* Accepted, 2009.

[18] F. Noo, M. Defrise, and R. Clackdoyle, "Single-slice rebinning method for helical cone-beam CT," *Phys Med Biol,* vol. 44, no. 2, pp. 561-70, Feb, 1999.

[19] J. Hsieh, and X. Tang, "Tilted cone-beam reconstruction with row-wise fan-to-parallel rebinning," *Phys Med Biol,* vol. 51, no. 20, pp. 5259-76, Oct 21, 2006.

[20] M. Kachelriess, S. Schaller, and W. A. Kalender, "Advanced single-slice rebinning in cone-beam spiral CT," *Med Phys,* vol. 27, no. 4, pp. 754-72, Apr, 2000.

[21] Y. Kyriakou, M. Kachelriess, M. Knaup *et al.*, "Impact of the z-flying focal spot on resolution and artifact behavior for a 64-slice spiral CT scanner," *Eur Radiol,* vol. 16, no. 6, pp. 1206-15, Jun, 2006.

Table 1 Signal to Noise Ratio (Arb. Units)

	Region I	Region II	Region III	Region IV
FBP 583 mAs	1.13	115.16	62.05	66.20
FBP 50 mAs	0.66	58.23	34.92	35.40
EST 50 mAs	2.39	156.66	112.35	115.19

Table 2 Contrast to Noise Ratio (Arb. Units)

	Region I	Region II	Region III	Region IV
FBP 583 mAs	58.61	53.17	5.37	8.05
FBP 50 mAs	30.86	24.57	2.50	4.02
EST 50 mAs	90.18	68.41	7.63	11.89

Figures:

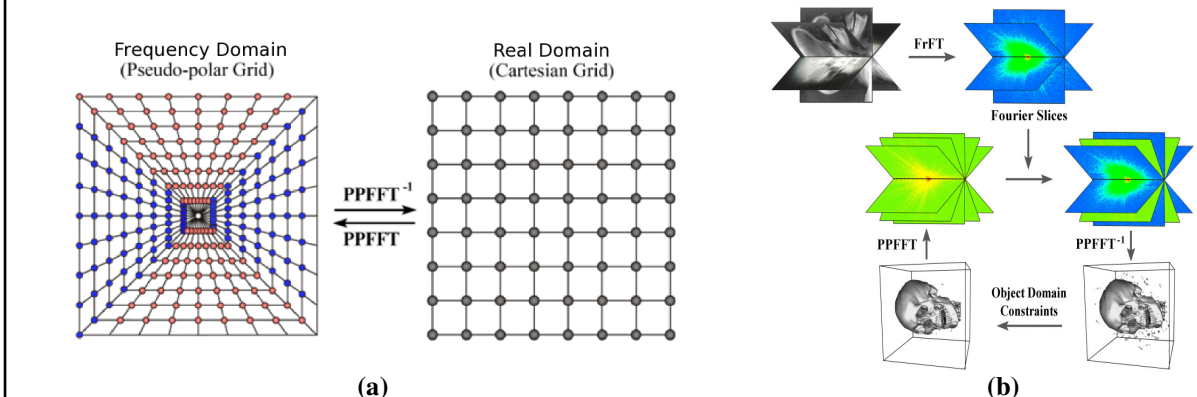


Fig. 1 (a) Graphical relationship between the oversampled pseudo-polar and Cartesian grids. (b) Schematic of the iterative algorithm.

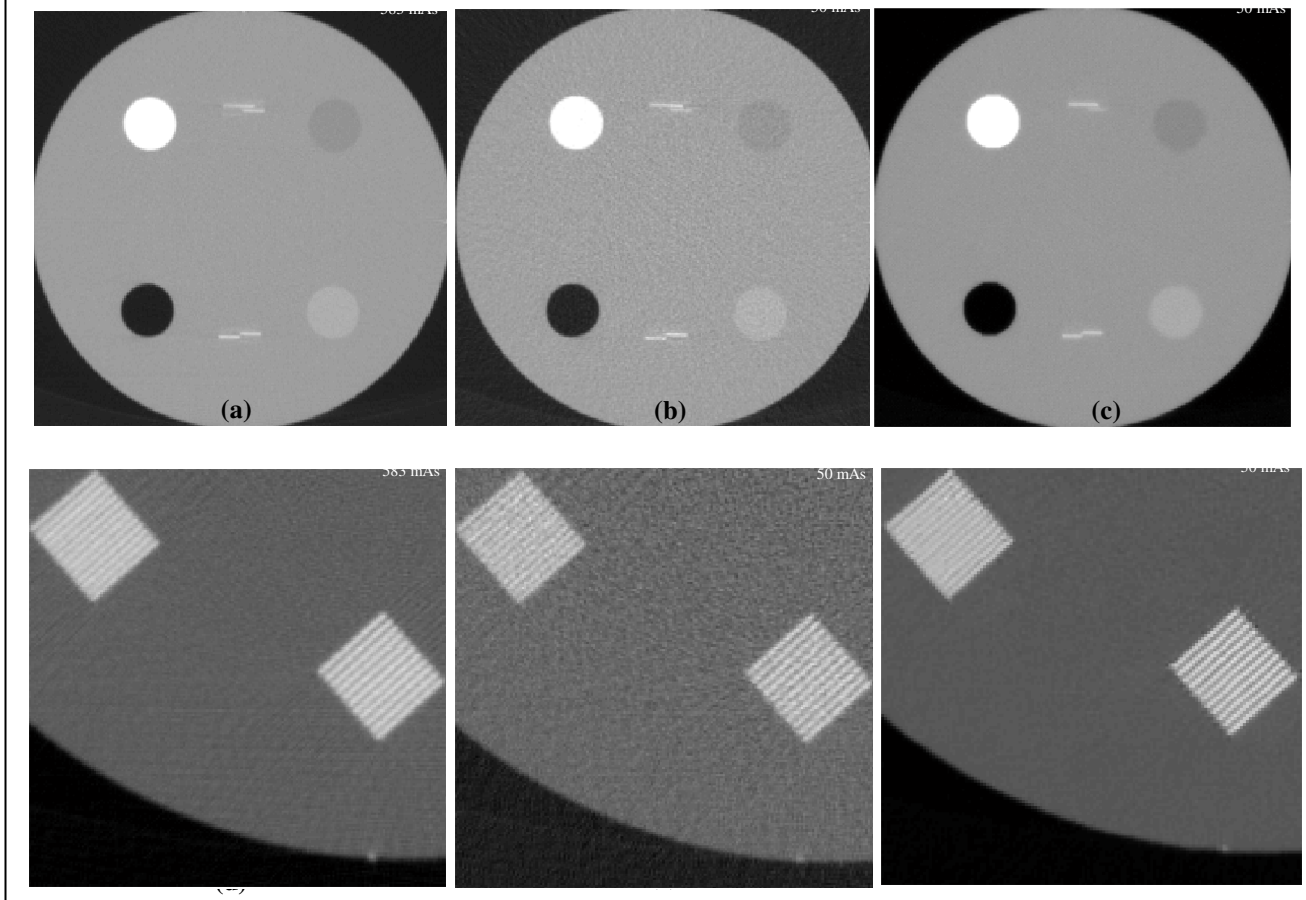


Fig. 2 (a) FBP reconstructions at 583 mAs (b) FBP reconstructions at 50 mAs (c) EST reconstruction at 50 mAs (d) FBP reconstructions at 583 mAs (e) FBP reconstructions at 50 mAs (f) EST reconstruction at 50 mAs

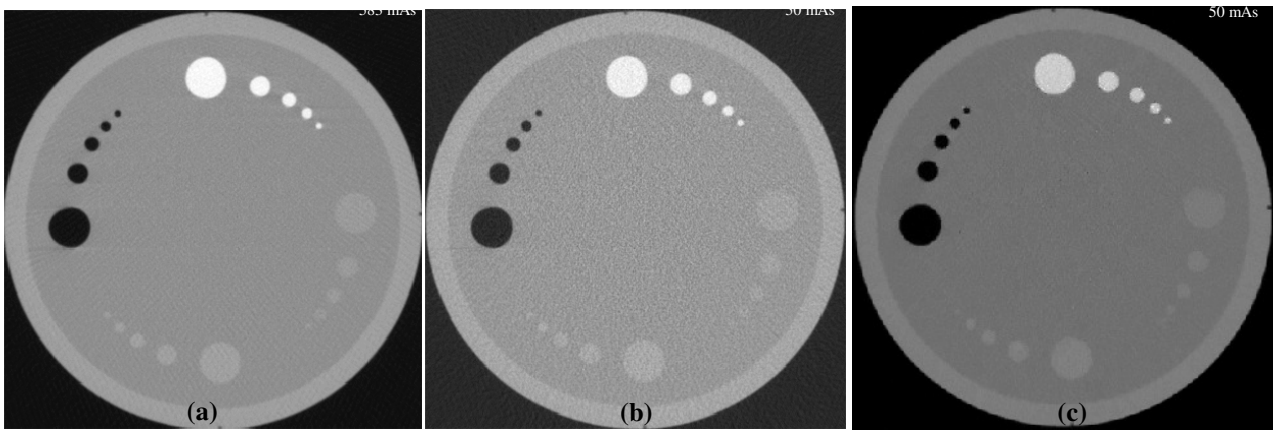


Fig. 3 (a) FBP reconstructions at 583 mAs (b) FBP reconstructions at 50 mAs (c) EST reconstruction at 50 mAs

2D X-Ray CT Reconstruction Based On TV Minimization And Blob Representation

Han WANG * Laurent DESBAT† Samuel LEGOUPIL‡

***Abstract*—The rotationally symmetric function (blob) is an ideal alternative to pixel representation of image. In this paper we present a 2D strip-integral projector based on blob which is independent to grid and blob's radius. This projector can be implemented on parallel architecture like NVIDIA GPU. A reconstruction algorithm by TV minimization is also developed for blob represented image. Some preliminary numerical experiences of this algorithm on fixed regular grid are presented.**

I. Introduction

The X-Ray generation and detection system, as in many inverse problems, can be modeled as an application from the "continuous object space" to the "discrete data space". In fact, the physical object illuminated by X-Ray is a function supported on a continuous compact spatial region. Due to detector's limited resolution, the sinogram data is discretized in form of a finite dimensional vector which contains consequently aliasing introduced in angular direction (for 2D). Therefore the discretization of sinogram inherently limits the resolution of the final reconstructed image. For an object $f(\cdot) : \mathbb{R}^2 \to \mathbb{R}$ supported on a continuous and compact spatial region of $\mathbb{R}^2$, an efficient way to handle this is to represent the real object as an interpolation with a family of translated local basis :

$$f(x) = \sum_j f_j \Phi(x - x_j), \text{ for } j = 1 \dots N \qquad (1)$$

where $\Phi(\cdot) : \mathbb{R}^2 \to \mathbb{R}$ is compactly supported basis function and $f_j \in \mathbb{R}$ is the interpolation coefficient. The X-Ray transform on f taken at direction $\theta \in \mathcal{S}^1$ is defined as :

$$\mathcal{P}_\theta f(y) = \mathcal{P} f(y, \theta) \triangleq \int_{\mathbb{R}} f(y + t\theta) dt, \quad \text{for } y \in \theta^\perp \quad (2)$$

and by linearity :

$$\mathcal{P}_\theta f(y) = \sum_j f_j \mathcal{P}_\theta \Phi(y - (x_j - x_j^\theta)) \qquad (3)$$

* Email : han.wang@cea.fr
† TIMC-IMAG, Medical Faculty, University Joseph Fourier, 38706 La Tronche, France. Email : laurent.desbat@imag.fr
‡ Email : samuel.legoupil@cea.fr
* ‡ CEA Saclay, Bat.516, PC 72, 91191 Gif-Sur-Yvette, France.
Manuscript received May 21, 2010; revised May 21, 2010.

where $x_j^\theta = \langle x_j, \theta \rangle \theta$ is the projection of x_j onto θ ($\langle , \rangle$ denotes the scalar product). We use in this paper the definition of Fourier transform : $\hat{f}(\omega) \triangleq \int_R f(x) \exp(-2\pi i x \omega) dx$. Taking Fourier transform on $\mathcal{P}_\theta f(\cdot)$ and following the projection-slice theorem $\hat{f}(\omega) = \widehat{\mathcal{P}_\theta f}(\omega)$ for $\omega \in \theta^\perp$, we obtain :

$$\hat{f}(\omega) = \hat{\Phi}(\omega) \sum_j f_j \exp(-2\pi i \langle \omega, x_j - x_j^\theta \rangle) \qquad (4)$$

The aliasing contained in the sum at right hand side of (4) is attenuated if Φ has a fast decreasing rate in frequential space. From this point of view, the pixel basis, which is the *de facto* standard of representation basis used by most of the state-of-art algorithms, is not an optimal choice for image representation due to the slow decreasing rate of Sinc function.

A. Blob function

The generalized Kaisser window function(also called Lewitt-blob) [3] has many advantages like, beside its rotation invariance, the fast decreasing rate in frequential plan and the analytical expression of X-Ray transform. These make it an ideal alternative to the pixel basis. The major default of Lewitt-blob function is that it's not positive-definite which prevents it to perfectly represent an homogeneous image [1]. However we can minimize the error by carefully choosing the parameters of blob controlling its shape [6]. In general the reconstruction based on blob is more smooth than that of pixel due to the low-pass filter character of blob function. In this paper, we address the issue of edge-preserving reconstruction based on blob representation.

We note in the following A the X-Ray projection operator. Depending on the projection geometry, A could be (in 2D) a parallel-beam or fan-beam projection. As shown in (3), the sinogram Af is determined by the contribution of each $A\Phi(\cdot - x_j)$, and followed by a sampling processes to yield finally the finite dimensional sinogram vector $g \in \mathbb{R}^M$. This indicates us to write the forward problem in a matrix-vector form :

$$g_i = \sum_{j=1}^{N} A_{ij} f_j, \text{ for } i = 1 \dots M \qquad (5)$$

with A_{ij} the value of $A\Phi(\cdot - x_j)(y)$ taken at i-th sampling point. We consider A as a linear map operating from $\mathbb{R}^N$ to $\mathbb{R}^M$, and the back-projector as transpose matrix $A^\top$.

B. Edge-preserving reconstruction

The algorithms of reconstruction by edge-preserving priors, particularly that of Total Variation(TV) minimization, have been shown to be very efficient on images with homogeneous regions. The total variation of a Bounded Variation function f is calculated as :

$$TV(f) = \int_\Omega \|\nabla f(x)\| dx \tag{6}$$

where Ω is the support region of f. It corresponds effectively to the length of all level set contour of image f. Obviously, a piecewise constant image's TV norm is small since it concentrates only on the contour of image. In most of numerical implementations, a discrete version of TV based on the differences between pixels is used. For an image represented by blob basis, the evaluation of (6) can be done in a non discrete manner, as presented in section III of this paper.

The paper is organized as follows : in section II we present a projector and back-projector based on blob representation of image and can be easily implemented on parallel architecture like NVIDIA GPU. Some reconstruction results by minimization of TV norm based on this new projector is given in section III.

II. Blob based strip-integral projector and back-projector

We can associate a scalar profile function $\phi(\cdot)$ to spherical symmetric basis function $\Phi(\cdot)$ by $\Phi(x) = \phi(\|x\|)$. The X-Ray transform is then reduced to the Abel transform :

$$\mathcal{P}_\theta \Phi(y) = 2\int_{\|y\|}^{+\infty} \phi(t) \frac{t}{\sqrt{t^2 - \|y\|^2}} dt \triangleq \mathcal{A}\phi(\|y\|) \tag{7}$$

and by (3):

$$\mathcal{P}_\theta f(y) = \sum_j f_j \mathcal{A}\phi(\|y - (x_j - x_j^\theta)\|) \tag{8}$$

As shown in Fig. 1, a simple implementation consists in tracing a line SD and evaluate (8) at corresponding $\vec{\theta} = SD/\|SD\|$ and $y = \langle Sx_j, \vec{\theta}\rangle \vec{\theta} - Sx_j$. Taking $D = (D_{i-1} + D_i)/2$, this becomes the so called Siddon ray-driven projector and equals to sampling $\mathcal{P}_\theta f(y)$ at frequency $1/\Delta_d$ in parallel beam geometry, with Δ_d the size of detector pixel.

A more realistic model is to calculate the strip-integral on a curve $\Gamma \subset \mathbb{R}^2$ formed by detector pixel area between two pixel borders D_{i-1} and D_i. This strip-integral acts as a low-pass filter and reduces the aliasing artefacts introduced by the sampling procedure in Siddon projector. Parameterizing Γ by $D : \mathbb{R} \supset I \to \Gamma$, the strip-integral reads, by definition of curve integration :

$$\int_\Gamma \mathcal{P}_\theta f(y) |d\gamma| = \int_I \mathcal{P}_\theta f(y)\ \|D'(l)\| dl \tag{9}$$

where $|d\gamma|$ denotes the length of integration element $d\gamma$, and θ, y depend on $D(l)$. For a 1D linear detector we have $D(l) = D_0 + l\vec{d}$ with $\vec{d}$ the unitary vector from one extremity of detector to the other, and $\|D'(l)\| = 1$. In case of parallel beam geometry, the strip-integral (9) is simplified to :

$$\int_{(i-1)\Delta_d}^{i\Delta_d} \mathcal{P}_\theta f(y) dl = \sum_j f_j \int_{B_{ij}} \mathcal{A}\phi(|l|) dl \tag{10}$$

where B_{ij} denotes the interval defined by the distance from the blob center x_j to parallel rays $\{SD, D \in \Gamma\}$: $B_{ij} = [(i-1)\Delta_d, i\Delta_d] - \langle S_0 x_j, \vec{d}\rangle$. See Fig 1. The integral in the right hand of (10) can be retrieved by looking up the one dimensional pre-calculated table $T(t) = \int_{-\infty}^{t} \mathcal{A}\phi(|s|) ds$ for upper and lower borders of B_{ij} then make the difference.

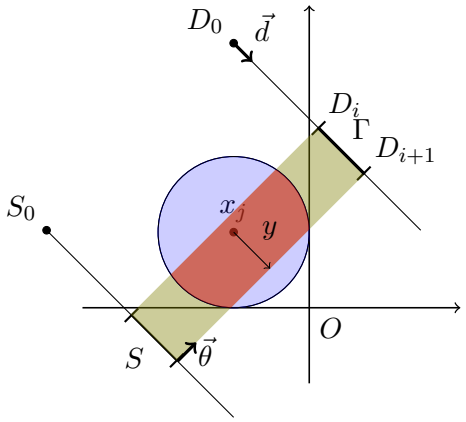


Fig. 1. Contribution of blob centered at x_j to detector $[D_i, D_{i+1}]$ in parallel beam is the strip-integral of the red region and depends only on the distance from x_j to SD_i and SD_{i+1}. This makes it easily evaluated by looking up pre-calculated one dimensional table $T(t)$.

The situation for fan beam geometry is shown in Fig. 2. One has a similar expression as (10) but with the integrand of strip integral depending on the detector pixel area Γ. One solution (already proposed in [4]) is to calculate the same integral as (10), which is equivalent to integrate the blob function on the red region and independent to detector curve. This allows us to use the same pre-integrated table $T(t)$ as in parallel beam case and unify the form of 2D projector for both parallel and fan beam.

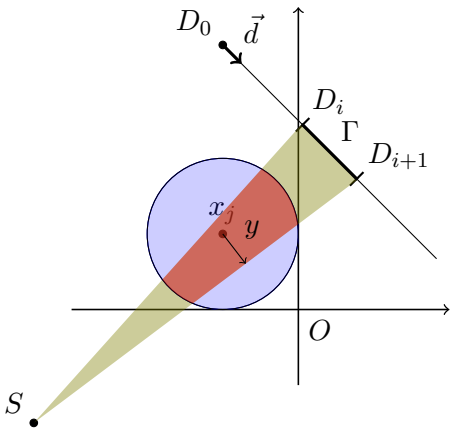


Fig. 2. In fan beam, the strip-integral is calculated as the integration of blob function on red region. As in parallel beam case, this depends only on the distance from x_j to SD_i and SD_{i+1}, and the same look-up table technique applies.

The back-projector is defined by :

$$(A^\top g)_j = \sum_i g_i \int_{B_{ij}} \mathcal{A}\phi(|l|)dl \qquad (11)$$

from where we see that the technique presented above applies identically for on the back-projector.

A. Dilated blob and irregular grid

We can control the radius of blob function Φ by a dilating parameter. Let $\Phi_\lambda(x) \triangleq \lambda^{-1}\Phi(x/\lambda), \lambda > 0$ be the dilated blob and $\phi_\lambda(\|x\|) \triangleq \lambda^{-1}\phi(\|x\|/\lambda)$ its profile function, then $\mathcal{A}\phi_\lambda(\|y\|) = \mathcal{A}\phi(\|y\|/\lambda)$ and :

$$\int_{B_{ij}} \mathcal{A}\phi_\lambda(|l|)dl = \lambda \int_{B_{ij}/\lambda} \mathcal{A}\phi(|l|)dl \qquad (12)$$

Therefore the same pre-integrated table $T(t)$ can be used to calculate the strip-integral of blobs with variant radius. This is useful in the projection of a image representation model as :

$$f(x) = \sum_j f_j \Phi_{\lambda_j}(x - x_j) \qquad (13)$$

where $\Phi_{\lambda_j}(x-x_j)$ is a blob centered at x_j and dilated by λ_j. Finally we remark that the looking-up table method proposed here is independant to the choice of x_j, in other words, the same projector can be applied on any kind of grid.

B. Implementation on parallel architecture

The implementation of projector A can take the form of either an on-the-fly operator or a pre-calculated sparse matrix. The storage of sparse matrix usually becomes prohibited for high dimension data, while the on-the-fly operator is generally very slow if implemented in a sequential way.

It's very easy to parallelize the projection method discussed above. One can allocate to each thread either a blob function $\Phi(\cdot - x_j)$ and evaluate the strip-integrals weighted by f_j on influenced detector pixels, or a detector pixel area $\Gamma = [D_i, D_{i+1}]$ and evaluate the expression (10) on it. Since we don't suppose any structure on the blob grid here, for the ray-driven projector one needs to gather first the index of blobs contributing to a given ray, which could be a lengthy operation. While for the blob-driven projector, one must manage the threads concurrency of writing on the same memory address. Our first experiences show that the blob-driven projector is conceptually easier than ray-driven projector, and that the threads concurrency can be handled correctly by atomic operations. Table I shows the acceleration achieved on a NVIDIA GeForce 8600M card for blob-driven projector (left) and back-projector (right) of fan beam geometry at 3 image size (number of blobs) and 3 projection sets : 64, 96, 128 projections. The reference CPU is Intel Core Duo T7500 2.20GHz. Although the GPU platform that we used supports only 32bit float operations, the numerical difference between CPU and GPU implementations is very small, about 10^{-5} for projector and 10^{-6} for back-projector.

	128^2	256^2	512^2	128^2	256^2	512^2
64	20.8	20.3	18.2	35.8	35.5	33.0
96	24.4	24.2	20.9	33.4	32.5	30.5
128	26.1	25.9	21.7	36.1	35.5	32.4

TABLE I
ACCELERATION OF GPU AGAINST CPU

III. TOTAL VARIATION MINIMIZATION

Consider now f sampled on a regular cartesian grid $G = \Delta\mathbb{Z}^2$, with sampling step length $\Delta > 0$. When $1/\Delta$ is sufficiently high, the total variation of f is well approximated by :

$$TV(f) \simeq \sum_{y_k \in G} \|\nabla f(y_k)\|_2 \Delta^2 \qquad (14)$$

The question here is how to efficiently calculate the gradient $\nabla f(x) = (\partial_1 f(x), \partial_2 f(x))^\top$ on G. We can use the following relation of Fourier transform :

$$\widehat{\partial_1 f}(\omega) = (-2\pi i\omega_1)\hat{f}(\omega), \text{ and} \qquad (15)$$

$$\widehat{\partial_2 f}(\omega) = (-2\pi i\omega_2)\hat{f}(\omega) \qquad (16)$$

where $\hat{f}(\omega) = \hat{\Phi}(\omega) \sum_j f_j \exp(-2\pi i \langle x_j, \omega \rangle)$. We define the dual cartesian grid $G' = \Delta^{-1}([-1/2, 1/2)^2 \cap \mathbb{Z}^2/K)$, with integer $K > 0$ and evaluate (15) on G'. Depending on the grid of blob representation, the sum involved in $\hat{f}(\omega)$ may be a non uniform Fourier transform and it can be treated efficiently by the fast transform technique like [2], while the Fourier transform of blob $\hat{\Phi}(\omega)$ can be precalculated on G'. Therefore one obtains the frequency value $\widehat{\partial_1 f}(\omega)$ on G' with low complexity operations. Now we only need to take the inverse FFT of $(K\Delta^2)^{-1}\widehat{\partial_1 f}(\omega)$ evaluated on G' to get $\partial_1 f(x)$ on G. The same operations apply for $\partial_2 f(x)$.

We note the operations presented above by $\mathcal{J}_G : \mathbb{R}^N \to \mathbb{R}^{2\times|G|}$: it's the operator yielding the value (with a precision dertermined by K) of $\nabla f(x)$ on G. It is linear on the vector $\vec{f}$ formed by $\{f_j\}$, and $(\mathcal{J}_G \vec{f})_k \in \mathbb{R}^2$ for $k = 1 \ldots |G|$. We can formulate the TV minimization problem as :

$$\min_{\vec{f}} \sum_{k=1}^{|G|} \|(\mathcal{J}_G \vec{f})_k\| \quad \text{s.t. } A\vec{f} = b \tag{17}$$

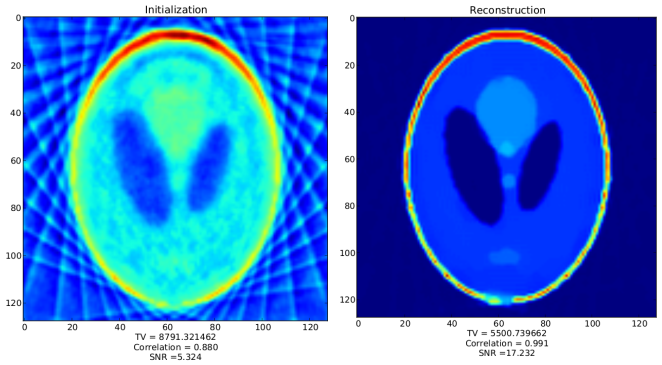
with A here the projector proposed in section II and b is the noiseless sinogram. The algorithm TVAL3 [5] is used in this paper to solve the problem (17). The final image represented by blobs is then obtained from $\vec{f}$ by (1).

A. Numerical experiences

We present now some reconstruction results based on (17). A standard 128×128 Shepp-logan phantom's pixel values are taken as the blobs coefficients (therefore the true image is the one represented by these blobs) and the blobs with standard radius are placed on a regular cartesian grid. The object is fitted into a square of size 25×25 cm centered at origin, and the distances from rotation center to source and detector are 80 cm and 40 cm respectively. We've simulated a fan-beam noiseless acquisition of 16 projections equally distributed on $[0, \pi)$ by applying the same blob-driven projector on the blob coefficients. The gain of TV reconstruction by (17) compared to the least square reconstruction is shown in Fig. 3. The horizontal central profile is shown in Fig. 4.

IV. Discussion

We have presented a parallelizable strip-integral projector based on blob functions and a related TV minimization algorithm. The independence of this projector to the grid and blob radius brings the possibility of modification of blob's position and radius during the iterations. In fact, the choice of blob position x_j, radius λ_j and weight f_j can be done automatically by the reconstruction algorithm : one needs to fix the number of blobs and minimize (17) with respect to all these parameters. We will investigate the possibilities in another publication.



(a) (b)

Fig. 3. Least square and TV reconstructions from 16 projections

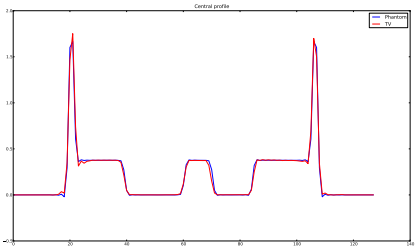


Fig. 4. Central profile of TV reconstruction from 16 projections

References

[1] M. D. Buhmann. Radial basis functions. *Acta numerica*, 9:1–38, 2001.

[2] J. Keiner, S. Kunis, and D. Potts. Using NFFT 3—A software library for various nonequispaced fast fourier transforms. *ACM Transactions on Mathematical Software (TOMS)*, 36(4):19, 2009.

[3] R. M. Lewitt. Multidimensional digital image representations using generalized Kaiser-Bessel window functions. *Journal of the Optical Society of America A*, 7(10):1834–1846, 1990.

[4] R. M. Lewitt. Alternatives to voxels for image representation in iterative reconstruction algorithms. *Physics in Medicine and Biology*, 37:705–716, 1992.

[5] Chengbo Li, Wotao Yin, and Yin Zhang. TV minimization by augmented lagrangian and ALternating direction ALgorithms, 2009.

[6] S. Matej and R.M. Lewitt. Practical considerations for 3-D image reconstruction using spherically symmetric volume elements. *Medical Imaging, IEEE Transactions on*, 15(1):68–78, 1996.

Effect of Modeling Anode Size and Angulation in Fully Iterative CT Reconstruction

Kevin J. Little* and Patrick J. La Rivière

***Abstract*—The small beveled angle in the rotating anode of an X-ray tube may potentially result in the loss of peripheral resolution in computed tomography (CT) due to the differences in the projection of the focal spot at the central and peripheral detectors. Many recent studies of iterative reconstruction for CT work with logged data and a lineaerized imaging model. In the logged domain of line integrals, we model the finite detector size and anode angle using sourcelets, detectorlets, and numerical integration based on Kaiser-Bessel basis functions. Utilizing a novel fully iterative reconstruction algorithm, we use a phantom containing an array of point impulses in order to compare the effect of modeling a finite source and anode angulation in the logged data domain. Provisionally, we find that there is not a significant difference in image resolution when reconstructing with a finite source model versus a point source model.**

***Index Terms*—Computed tomography (CT), iterative reconstruction**

I. Introduction

A rotating anode made of a high-atomic-number material such as tungsten is usually a core component of modern X-ray tubes. Incident electrons, which are accelerated across a vacuum by a potential difference after they are boiled off the tube's cathode, impinge on the anode and result in X-rays due to bremsstrahlung radiative losses and the emission of characteristic X-rays. Nonradiative (heat) energy is also generated in large amounts in these interactions, and the rotation of the anode allows for the dissipation of this energy over a larger area [1], [2].

In order to dissipate this nonradiative energy over an even greater area, the edge of the anode known as the focal track is generally a beveled edge with a small angle (typically $5°-7°$ in a computed tomography (CT) tube). While a comparatively larger area of the focal track is exposed to electrons, the small size of the effective projected focal spot is maintained due to the *line-focus principle*, which is illustrated in Fig. 1 [1], [2].

The line-focus principle works well for the central detector channels, but not all detectors see the same projection of the rectangular, angled focal spot. Due to the anode angulation, a more peripheral detector sees a larger effective focal spot size than a central detector [2]. This anode angle effect has been shown to degrade peripheral resolution in some cases, and sinogram-domain methods for compensating for the resulting blurring have been proposed [3]. In most analytic and iterative CT reconstruction algorithms, the anode is modeled as a point source. It is not known whether modeling a finite, angled anode as opposed to a point source will have a significant effect on image resolution and noise properties.

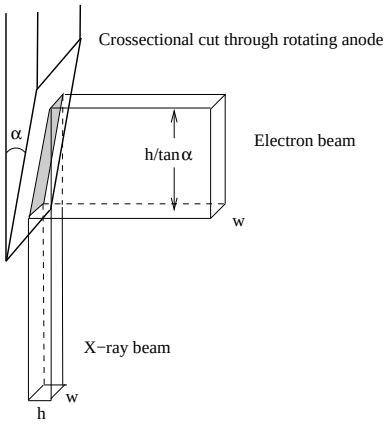


Figure 1. Illustration of the line-focus principle employed in X-ray anodes. Due to the beveled edge of the focal track, a relatively large area of the anode can be exposed by the electron beam (traveling in from the right in this figure) while the projected cross section of the X-ray beam (traveling downward in this figure) remains relatively small.

Many recent investigations of iterative reconstruction for CT work with logged data and a linearized imaging model [4], [5]. Along with an assumption of a Gaussian noise model in the log domain, this allows for the use of a quadratic objective function and avoids evaluating exponentials. However, any linear modeling of finite source and detector effects in the log domain is necessarily approximate, since these finite apertures lead to linear averaging in the transmitted intensity domain, not in the logged domain of line integrals.

Our goal in this paper is to evaluate this approximation, i.e. to assess whether linear modeling of the anode size and angulation in the logged data domain can indeed lead to improved image quality. Focusing on the single-slice circular fanbeam case for the purpose of simpler calculations, we perform numerical integration using Kaiser-Bessel spherically symmetric basis functions in order to compute a system matrix that takes the finite, angled anode into account for a linearized imaging system. Using the Radonis CT simulation package (Philips Research and Development, Hamburg, Germany) in order to generate realistic, exponentially attenuated forward projection data for a finite, angled anode, we then perform fully iterative image reconstruction using a modified version of the image space reconstruction algorithm (ISRA) that has been revised to include a roughness penalty (smoothing). A provisional conclusion about the effects of anode modeling on fully iterative image reconstruction is made.

This work was supported by NIH Grant R01CA134680. *Asterisk indicates corresponding author.*

K. J. Little is with the Department of Radiology, The University of Chicago, Chicago, IL 60637 USA (e-mail: little@uchicago.edu).

P. J. La Rivière is with the Department of Radiology, The University of Chicago, Chicago, IL 60637 USA.

II. METHODS

A. System model

In order to take the finite detector size into account, we divide the finite anode into a grid of sourcelets and each detector into a grid of detectorlets [6]. A diagram showing fanbeam CT measurement geometry, including sourcelets and detectorlets, is shown in Fig. 2. We want to efficiently calculate the average line integral from each of the sourcelets to each of the detectorlets, so we use the spherical Kaiser-Bessel basis function described in Ref. [7] and implemented for CT reconstruction in Ref. [4]. This basis is useful because the line integral through each volume element "blob" is based only on the distance of the line integral from the center of the volume element. While the implementation of the "blob" model in Ref. [4] assumed finite detectors, it assumed an ideal point source. We modify the system model in Ref. [4] to include a finite source, and we also do not use an interpolated method of integration.

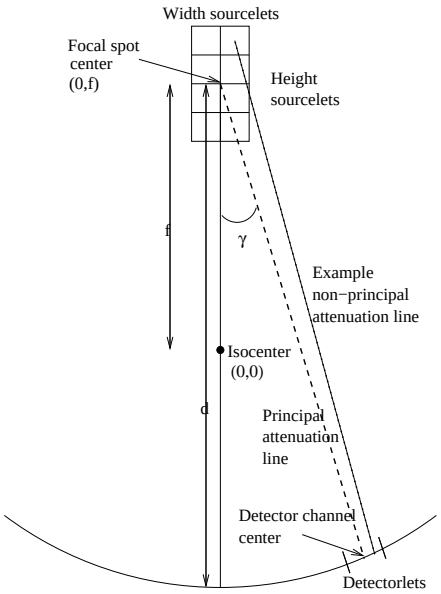


Figure 2. Geometry, including sourcelets and detectorlets, of a fanbeam CT measurement for a single projection view.

We represent the continuous distribution, $f(\vec{x})$, of the absorption coefficient at spatial position $\vec{x}$ as a sum of basis functions $\tilde{f}(\vec{x})$ such that

$$\tilde{f}(\vec{x}) = \sum_{j=1}^{N} c_j g b(\vec{x} - \vec{x}_j), \tag{1}$$

where g is the grid increment between each of the N volume elements, c_j is the weight given to each volume element, and $b(\vec{r})$ represents the Kaiser-Bessel basis function at each volume element. The Kaiser-Bessel function is given by [7]

$$b_{m,a,\alpha}(r) = \begin{cases} \frac{\left(\sqrt{1-(r/a)^2}\right)^m I_m\left(\alpha\sqrt{1-(r/a)^2}\right)}{I_m(\alpha)}, & 0 \le r \le a \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

where I_m represents the modified Bessel function of order m, a is the radius of the basis function, α is a parameter modifying the blob shape, and r is the radial distance from center of the volume element. A single line integral $L[\vec{x}_d, \vec{x}_s, \vec{x}_j]$ through a single volume element will be given by

$$L[\vec{x}_d, \vec{x}_s, \vec{x}_j] = \int_0^1 b[\vec{x}_s + l(\vec{x}_d - \vec{x}_s) - \vec{x}_j]\, dl. \tag{3}$$

where $\vec{x}_d$ is a given detectorlet location, $\vec{x}_s$ is a given sourcelet location, and $\vec{x}_j$ is the center location of a volume element. The line integral in terms of the perpendicular distance w from the center of the volume element to the line integral can be written as [4]

$$\begin{aligned} &L_{m,a,\alpha}[\vec{x}_d, \vec{x}_s, \vec{x}_j] = \\ &L_{m,a,\alpha}(w_{dsj}) = L_{m,a,\alpha}(w) = \\ &2\int_0^{\left(a^2-w^2\right)^{1/2}} b_{m,a,\alpha}\left[\left(w^2+t^2\right)^{1/2}\right] dt \quad |w| \le a \end{aligned} \tag{4}$$

where

$$w_{dsj} = w = \sqrt{\|\vec{x}_j - \vec{x}_s\|^2 - \{(\vec{x}_j - \vec{x}_s)\cdot[\vec{x}_d - \vec{x}_s]/\|\vec{x}_d - \vec{x}_s\|\}^2}.$$

The integral above is known exactly and is solved in Ref. [7]. Eq. (4) becomes

$$\begin{aligned} &L_{m,a,\alpha}(w) = \\ &\frac{a}{I_m(\alpha)}\left(\frac{2\pi}{\alpha}\right)^{\frac{1}{2}}\left[\sqrt{1-(w/a)^2}\right]^{m+\frac{1}{2}} \cdot \\ &\times I_{m+\frac{1}{2}}\left[\alpha\sqrt{1-(w/a)^2}\right] \end{aligned} \tag{5}$$

For a given volume element j and detector i, we want to compute the element a_{ij} for system matrix A that represents the contribution of the given volume element to the recorded output of the given detector such that

$$\vec{y} = A\vec{c}, \tag{6}$$

where $\vec{y}$ is the data vector comprised of the M measurements of all the detectors at all rotation angles and and $\vec{c}$ is the array of N weights c_j in Eq. (1). We may find the average line integral over all detectorlets and sourcelets (an effective numerical integration) contributing to a system matrix element a_{ij} by calculating

$$a_{ij} = \frac{1}{DS}\sum_{d=1}^{D}\sum_{s=1}^{S} L(w_{dsj}), \tag{7}$$

where D is the total number of detectorlets and S is the total number of sourcelets. By making the statement in Eq. (7), we assume that we are working with a linearized imaging equation and linearized data.

B. Iterative reconstruction algorithm

We wish to solve Eq. (6) for the $N \times 1$ vector of "blob" weights $\vec{c}$ using a fully iterative reconstruction method. In order to achieve this, we seek to minimize the penalized least squares (PLS) objective function given by

$$\Phi^{PLS}(\vec{c};\vec{y}) \equiv \mathcal{L}^{PLS}\left(\vec{c};\vec{y}\right) + \beta R\left(\vec{c}\right), \tag{8}$$

where

$$\mathcal{L}^{PLS}\left(\vec{c};\vec{y}\right) = \sum_{i=1}^{M}\left[y_i - \sum_{j=1}^{N} a_{ij}\tilde{c}_j\right]^2, \tag{9}$$

$R(\vec{c})$ is a roughness penalty, and β is the smoothing parameter which controls the relative influence of the roughness penalty. The tildes above c_j and $\vec{c}$ represent an estimation of the blob weights. The roughness penalty $R\left(\vec{\tilde{c}}\right)$ can be expressed generally as

$$R\left(\vec{\tilde{c}}\right) = \sum_{k=1}^{K}\psi_k\left(\left[T\vec{\tilde{c}}\right]_k\right), \tag{10}$$

given by Fessler [8], where T is a matrix and ψ_k is a potential function that assigns a cost to the K combinations of blob coefficient values represented by the matrix product $T\vec{\tilde{c}}$. In this work, we use a penalty applied to difference of each estimated blob weight $\tilde{c}_j$ with its horizontal and vertical neighbors.

In order to perform fully iterative image reconstruction, we seek an algorithm that readily guarantees the enforcement of positivity constraints and of monotonic decreases in the objective function. The derivation of the update equation used in this paper closely follows the derivation in Appendix C of Ref. [9] and has the result:

$$\tilde{c}_j^{(n+1)} = \tilde{c}_j^{(n)}\left[1 + \frac{\tilde{n_j} - \beta\left(4\tilde{c}_j^{(n)} - \sum_{k=1}^{K} t_{kj}t_{ki}\tilde{c}_j\right)}{2\sum_{i=1}^{M} a_{ij}\overline{\tilde{y}}_i^{(n)} + 8\beta\tilde{c}_j^{(n)}}\right], \tag{11}$$

where

$$\tilde{n}_j \equiv 2\sum_{i=1}^{M} a_{ij}\left(y_i - \overline{\tilde{y}}_i^{(n)}\right) \tag{12}$$

and

$$\overline{\tilde{y}}_i^{(n)} \equiv \sum_{j=1}^{N} a_{ij}\tilde{c}_j^{(n)}. \tag{13}$$

When $\beta = 0$ in Eq. (11) above, the algorithm takes the form of the image space reconstruction algorithm (ISRA). So, we think of the above update as ISRA with an included roughness penalty.

Due to the increased computational load introduced by multiple sourcelets and detectorlets, it is not always computationally judicious to compute the system matrix for all sinogram projection angles using Eq. (7). We instead calculate the system matrix elements a_{ij} for one projection angle and utilize rotations with linear interpolation for the remaining projection angles. For forward projections into sinogram space, we rotate the image space information to the angles complementary to the rotation of the source and detectors at each projection angle and then make the forward projection. For backprojections into image space, we rotate the result of the backprojection to the projection angles and sum the results over all angles.

While Eq. (11) is guaranteed to produce monotonic decreases in the objective function, it is sometimes necessary to speed up the algorithm for large system matrices. We implement ordered subsets (OS) in order to speed up convergence. That is, each iterative update takes into account only a subset of all projection angles. By doing so, we decrease the time to complete each iteration by approximately O^{-1}, where O is the number of ordered subsets.

III. Results

A. Data simulation model

In order to generate independent data that account for finite focal spot size, for finite detector size, and for an independent model of attenuation, we used the Radonis CT simulation package (Philips Research and Development). The simulation package allows the specification of the geometry of the source and detectors, including anode angulation. We simulated a focal length of 570.0 mm and a detector-to-isocenter distance of 470.0 mm. A single-row detector was simulated with 672 channels, each with a size of 1.4 mm in the transverse direction and 1.0 mm in the longitudinal direction, measured at the detector. We simulated 1160 projections, spaced equally over 360°. We used 3 × 3 detectorlets when generating simulation data using Radonis. The output of the system was a linearized sinogram.

We simulated two types of sinograms with the simulation software. The first "realistic sinogram" was simulated using an anode angle of 5°, with a projected focal spot size of 0.9 mm in height and 0.7 mm in width. For this simulation, 5 × 5 sourcelets were used. Our second sinogram was meant to represent an idealized case. This "ideal sinogram" modeled a point source by simulating only one sourcelet.

B. Iterative reconstruction model

The iterative reconstruction model used the same focal length, detector-to-isocenter distance, detector size, and number of projections as the simulation model. The iterative reconstruction method based on Eq. (11) was used. The ordered subsets and rotation-based methods discussed above were employed. For the current study, 200 iterations of 20 ordered subsets were completed. The blob volume elements were placed on a Cartesian, two-dimensional (2D) grid with spacing $g = 1.296$ mm. The "standard" parameters for 2D blob-based reconstruction were used {m=2, a/g=2.00, and α=10.826} [4]. After iterating, Eq. (1) was used to convert the blob weights to discrete Cartesian pixels.

In modeling the source during reconstruction, two models were used. The first, the "realistic reconstruction," generated a finite source of the same dimensions as were used in the simulation model. For calculating the numerical integrations using this finite source, the source was divided into 5 × 5 sourcelets, equally spaced along the angled edge of the source. The detectors were divided into 5 × 5 detectorlets, equally spaced in the transverse and longitudinal directions. The second source model, the "ideal reconstruction," computed the numerical integrations using a point source (only one sourcelet). In this second model, the detectors were divided into 11 × 11 detectorlets, in an attempt to partially compensate for any possible effects of computing fewer line integrals for each numerical integration.

C. Reconstructed images of impulse phantom

We used the impulse phantom developed in Ref. [3] in order to investigate the resolution properties of the reconstructed images. The impulse phantom, as can be see in the reconstructed image of Fig. 3, was an elliptical phantom with an array of point impulses of size 0.1 mm.

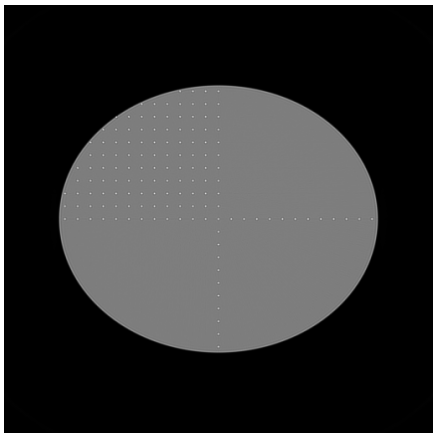


Figure 3. Reconstructed impulse phantom using realistic sinogram, realistic reconstruction, and $\beta = 1$.

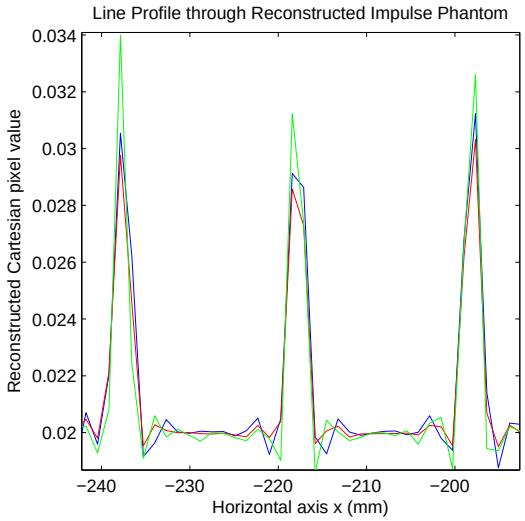


Figure 4. Horizontal line profiles through middle of reconstructed impulse phantom, zoomed to the three most peripheral impulses, for the realistic sinogram reconstructed with the realistic model (blue), the realistic sinogram reconstructed with the ideal model (red), and the ideal sinogram reconstructed with the ideal model (green). Due to slight normalization differences, the two reconstructions using the ideal model have been shifted up by 0.0006 so that all three reconstructions have similar background values. In the above data, $\beta = 1$.

Horizontal line profiles through the center of three sinogram and reconstruction combinations are shown in Fig. 4. The plot is zoomed on the three most peripheral impulses in the reconstructed images. The purpose of this plot is to illustrate the resolution differences between the three combinations in the region where differences are most expected. The impulse peaks in the realistic sinogram reconstructed with the realistic model do not show large differences from the impulses in the realistic sinogram reconstructed with the ideal model. The reconstruction of the ideal sinogram with the ideal model–a hypothetical best case scenario–shows considerably higher and narrower peaks than the other two sinogram-model combinations.

IV. Discussion and Conclusion

We have implemented a fully iterative image reconstruction algorithm that models a finite source and finite detectors by using a numerical integration model based on sourcelets, detectorlets, and volume elements modeled with the Kaiser-Bessel basis function. Our iterative reconstruction algorithm updates using a modified image space reconstruction algorithm which guarantees the positivity constraint will be met. We generated independent simulation data using the Radonis software package, and reconstructed an impulse phantom for comparison of various simulation and reconstruction models.

By comparing profiles through the image reconstructions of the realistic impulse phantom sinogram data for both the realistic reconstruction model and the ideal reconstruction model (Fig. 4), we provisionally conclude that modeling the finite spot size in a linearized iterative reconstruction method did not have a large effect on the resolution of our reconstructed image. However, it is important to note that we are still optimizing a few factors that would allow us to claim there was negligible effect with greater confidence. These include the effect that the Kaiser-Bessel blobs may have on resolution and the effect of varying the smoothing parameter β.

In order to achieve conclusive results, we will compare the present results with those obtained when the anode size and angulation effects are modeled in the transmitted intensity domain. This is what we did using a sinogram restoration strategy in Ref. [3], and we plan to implement a fully iterative algorithm based on a similar imaging equation for comparison with the current linearized model results.

References

[1] J. T. Bushberg, J. A. Seibert, E. M. Leidholt, Jr., and J. M. Boone, *The Essential Physics of Medical Imaging*. Philadelphia, PA: Lippincott Williams Wilkins, 2002.

[2] J. Hsieh, *Computed Tomography: Principles, Design, Artifacts, and Recent Advances*. Bellingham: SPIE Press, 2003.

[3] P. J. La Rivière and P. Vargas, "Correction for Resolution Nonuniformities Caused by Anode Angulation in Computed Tomography," *IEEE Trans. Med. Imag.*, vol. 27, no. 9, pp. 1333-1341, Sep. 2008.

[4] A. Ziegler, T. Köhler, T. Nielsen, and R. Proksa, "Efficient projection and backprojection scheme for spherically symmetric basis functions in divergent beam geometry," *Med. Phys.*, vol. 33, pp. 4653-4663, 2006.

[5] J. B. Thibault, K. D. Sauer, C. A. Bouman, and J. Hsieh, "A three-dimensional statistical approach to improved image quality for multislice helical CT," *Med. Phys.*, vol. 34, pp. 4526-4544, 2007.

[6] C. R. Crawford and K. F. King, "Computed tomography scanning with simultaneous patient translation," *Med. Phys.*, vol. 17, pp. 967-982, 1990.

[7] R. M. Lewitt, "Multidimensional digital image representations using generalized Kaiser-Bessel window functions," *J. Opt. Soc. Am. A*, vol. 7, pp. 1834-1846, 1990.

[8] J. A. Fessler, "Statistical image reconstruction methods for transmission tomography," in *Handbook of Medical Imaging*, J. M. Fitzpatrick and M. Sonka, Eds. Bellingham, WA: SPIE Press, 2000, vol. 2, pp. 1-70.

[9] P. J. La Rivière, J. Bian and P. Vargas, "Penalized-likelihood sinogram restoration for computed tomography," *IEEE Trans. Med. Imag.*, vol. 25, no. 8, pp. 1022-1036, Aug. 2006.

Few-view Image Reconstruction of Coronary Arteries: A Preliminary Study Using Simulation and Real Data

Xiao Han*, Junguo Bian*, Diane R. Eaker†, Emil Y. Sidky*, Erik L. Ritman†, and Xiaochuan Pan*
*Department of Radiology, The University of Chicago, 5841 S. Maryland Avenue, Chicago, IL 60637
E-mail: xiaohan@uchicago.edu and xpan@uchicago.edu
†Department of Physiology and Biomedical Engineering, Mayo Clinic College of Medicine, Rochester, MN, 55905

***Abstract*—Tomographic images offer accurate three-dimensional (3D) structural information about coronary arteries. For a number of practical considerations such as imaging time and radiation dose, it is often desirable to obtain 3D coronary-artery images from data acquired at a few views, *i.e.*, less than 10 views. In this case, analytic algorithms which generally require data collected at a large number of projection views would yield reconstructions with significant aliasing artifacts, while effective optimization-based (*i.e.*, iterative) algorithms may be designed that exploit the sparsity of coronary-artery structures for improving image quality. In this work, we have designed and investigated optimization-based algorithms for image reconstruction from few-view data and applied them to reconstructing coronary artery images from both few-view simulation data and few-view real data. In particular, the core algorithms are based upon constrained minimization of image's total-variation (TV) in combination of an iterative hard-thresholding technique. Quantitative assessment of image quality has been performed, and results of the study appear to suggest that carefully designed optimization-based, sparsity-exploiting algorithms can yield images with higher fidelity than some of the existing, well-known algorithms. A potentially significant implication of the work is that coronary-artery images of practical utility may be obtained from highly sparse data, thus minimizing radiation dose to the imaged subject and also shortening imaging time.**

I. Introduction

Computed tomography (CT) can be used for obtaining accurate three-dimensional (3D) coronary-artery structural information that can significantly benefit a large number of preclinical and clinical applications [1]. In a standard CT setting, a large number (300-1000) of projections are typically acquired from which images can be reconstructed by use of existing algorithms. However, the collection of data at a large number of views would not only prolong considerably imaging time but also involve a substantial amount of radiation dose that can induce potentially significant damage to the imaged subject. It is therefore of high practical interest in obtaining accurate reconstruction of coronary arteries from a very small number of projections.

Conventional filtered backprojection (FBP)-based reconstruction algorithms assume a continuous imaging model, and therefore work well for data densely sampled at a large number of projection views. When applied to few-view projections, however, the algorithms yield images with obscuring aliasing artifacts, largely due to insufficient angular sampling. On the other hand, optimization-based algorithms that are based upon a discrete imaging model can, in general, reconstruct images with fewer aliasing artifacts than the FBP-based algorithms. In particular, the optimization-based algorithms allow the incorporation of appropriate prior knowledge about the imaged subject for constraining properties of the reconstructed images. For example, optimization-based algorithms have been explored in which the sparsity of coronary-artery structures has been used as prior knowledge in reconstruction of coronary-artery images from few-view projections [2], [3], [4].

In recent years, a great deal of work on optimization-based algorithms has been reported in which the sparsity of images or their transforms has been demonstrated to be potentially useful for determining an underlying discrete function from its sparse samples [5]. Inspired by the work, we have developed optimization-based algorithms for reconstructing CT images from cone-beam data collected at a small number of projection views, which minimizes the image's total variation (TV) subject to data and other constraints [6], [7]. The TV algorithms have been demonstrated to yield images of potential utility in some CT applications involving real data acquired at a number of views substantially lower than what are used in current CT imaging procedures.

In this work, we design and apply optimization-based algorithms specifically tailored to reconstructing coronary-artery images from few-view projections by exploiting the image sparsity. Specifically, we investigate and exploit the contrained TV-minimization [5] and iterative hard thresholding (IHT) [8], [9] techniques for efficient reconstruction of sparse images, and evaluate the impact of some key parameters in the algorithms on image quality. In our study, we consider the case where the chest wall and soft-tissues are removed, and the underlying image is comprised of vasculature structure only, which therefore is itself a sparse spatial distribution. We employ a circular cone-beam micro-CT geometry for generating simulated data in the

Figure 1. Volume-rendered gold-standard images. Left: The FORBILD numerical phantom. Right: The coronary artery specimen image reconstructed by use of FDK from 721 real projections.

numerical phantom study and for acquiring real projections in the real-data study. We also carry out a preliminary evaluation study of the reconstruction results. For example, in addition to visual assessment of image quality, we specifically designed and computed a number of quantitative metrics to evaluate and compare the quality of the reconstructed images.

II. Materials and Methods

A. Projection data

We generated simulated coronary-artery projection data from the FORBILD phantom [10], [2] using a circular cone-beam geometry, in which the source-to-detector distance was 61.0 cm, and the source-to-rotation-axis distance was 59.7 cm. Five projections were obtained at angles uniformly distributed over the 2π scanning range. The phantom consists of a $256 \times 256 \times 256$ binary image array, and each cubic voxel represents a physical size of 82 μm. The detector was modeled as a 512×512-bin flat panel, with each bin measuring 41×41 μm^2. We display the phantom image in Fig. 1(a), which served as the ground truth for our simulation studies.

Real cone-beam projection data were acquired with a custom-made micro-CT scanner [11] sharing the same source-to-detector and source-to-rotation-axis distances. The key components of the scanner included a spectroscopic x-ray source emitting quasi-monoenergetic beam at approximately 18keV, and a fluorescing thin crystal plate, which was imaged with a lens onto a 2.5×2.5cm, 1024×1024-pixel CCD array operated at 16-bit gray-scale resolution (Princeton Instruments, NJ). A human coronary artery cast specimen was positioned close to the crystal and was rotated in 721 equiangular steps around 360^o between each x-ray exposure. We display a schematic and a photograph of the scanner in Fig. 2, and show the FDK reconstruction of the specimen from all 721 projections in Fig. 1(b). This full-view FDK reconstruction was used as the gold-standard image for our real-data studies.

B. Optimization-based image reconstruction

We have developed optimization-based algorithms for reconstructing the coronary artery images by exploiting the image sparsity. The first algorithm is the IHT-POCS algorithm, which

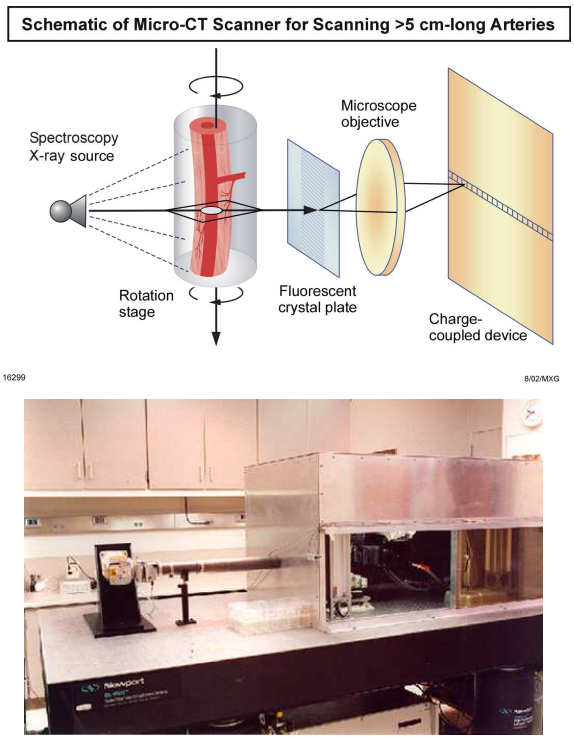


Figure 2. The custom-made micro-CT system for data acquisition. Top: a schematic of the system and its components; Bottom: a photograph of the actual system.

employs POCS to minimize the data distance, while setting a hard threshold at each iteration to enforce s-sparsity to the image, *i.e.* keeping only the largest s voxels and setting the rest to zero [9]. The sparsity parameter s is obtained from prior knowledge about the underlying image. When s is chosen as a large number, for instance, the total number of voxels in the image array, the hard-thresholding has no effect and the algorithm reduces to standard POCS. When we reduce the value of s to an appropriate range, the hard-thresholding step helps the algorithm to seek among all images in agreement of data an image with only s non-zero voxels. However, s cannot be chosen unrealistically small, as the algorithm will fail to effectively minimize the data distance [9].

The second algorithm is the TV-minimization algorithm, which exploits the sparsity of images in their gradient magnitude representation. The algorithm minimizes the image TV by steepest-descent while enforcing the data distance constraint by POCS, and therefore seeks among all image candidates in agreement with projection data the one with minimum TV [7]. Since real data are always contaminated by inconsistency from physical factors such as noise, a data-error tolerance parameter ϵ is introduced in the algorithm to relax the constraint.

The third algorithm combines IHT with TV-minimization, and is thus named the IHT-TV algorithm. The algorithm seeks an image with minimum TV among all images satisfying the data-distance constraint and the s-sparsity requirement [9]. Both

parameters ϵ and s need to be selected appropriately based upon prior knowledge about the image and the data, and the selection of these parameters significantly impact the the final reconstructed image.

III. Results

We first applied the IHT-POCS, TV-minimization, and IHT-TV algorithms to reconstructing the coronary-artery phantom image from simulated projection data. In Fig. 5 we display the middle transverse slices of images reconstructed from 5 projections by running these algorithms for 200 iterations. Also displayed are the ground-truth image as well as images reconstructed by use of the standard FDK and POCS algorithms. While little information about the coronary-artery structure could be derived from the FDK reconstructed image, it appears that the optimization-based algorithms are able to improve the image quality by reducing some artifacts. In Fig. 4 we show the reconstructed images obtained after running the algorithms for 2000 iterations. It appears that the image quality is further improved by some of the optimization-based algorithms. In addition, quantitative assessment of the reconstructed image was carried out, in which we calculated a number of image-based metrics relative to the ground truth. We list in Tab. I the results of two of these metrics, the Pearson correlation coefficient and the root mean square error (RMSE), which suggest that the IHT-TV algorithm yields an image closest to the original phantom. Additional results involving other quantitative metrics will be presented at the conference.

From the real data of the coronary artery specimen, we formed two few-view data sets consisting of 5 and 9 projections, respectively, uniformly distributed over the 2π range. We then reconstructed images from these two data sets by applying the FDK and different optimization-based algorithms. In Figs. 5 and 6 we show the reconstructed images obtained after 200 iterations. Based upon image visualization and quantitative assessement results in Tab. II, similar observations as in the numerical phantom study can be made. Additional extensive results of reconstructed images and quantitative evaluations will be presented at the conference.

Table I
QUANTITATIVE COMPARIONS OF IMAGES RECONSTRUCTED BY DIFFERENT ALGORITHMS FROM NUMERICAL PHANTOM DATA

Metrics	# Iter.	FDK	POCS	IHT-POCS	TV	IHT-TV
RMSE	200	32.7	5.5	5.6	5.0	4.8
(10^{-2}cm^{-1})	2000		4.0	4.0	2.9	1.6
Corr. Coeff.	200	0.31	0.89	0.88	0.91	0.91
	2000		0.94	0.94	0.97	0.99

IV. Discussion

We have investigated and developed optimization-based algorithms for image reconstruction from few-view projections by exploiting image sparsity, and we applied these algorithms to

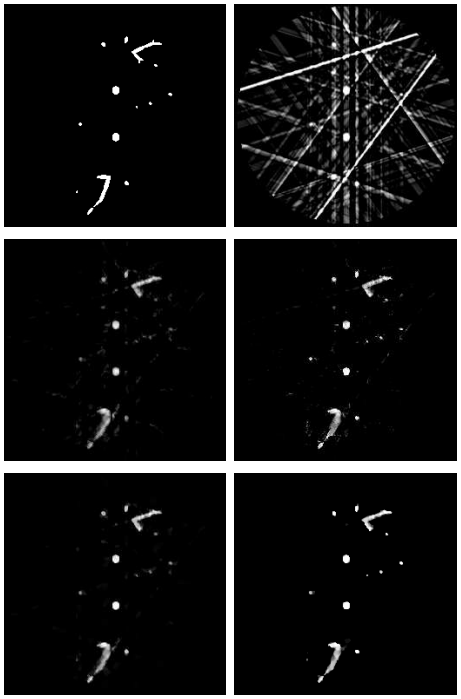


Figure 3. Images at the middle transverse slice of the numerical phantom reconstructed from 5 projections by use of different algorithms. From left to right, top row: ground truth, FDK. Middle row: POCS, IHT-POCS. Bottom row: TV-minimization, IHT-TV. Iterative results were obtained after 200 iterations. Gray-scale window: [0,1]cm^{-1}.

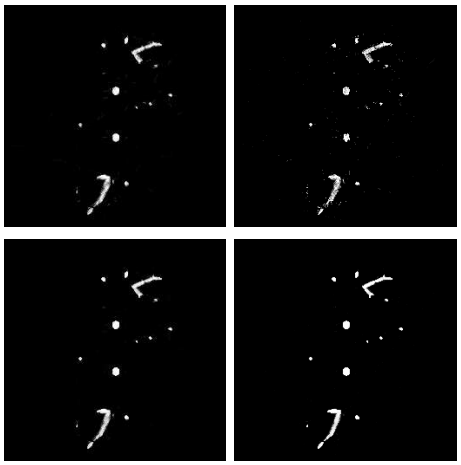


Figure 4. Images at the middle transverse slice of the numerical phantom reconstructed from 5 projections by use of different algorithms. From left to right, top row: POCS, IHT-POCS. Bottom row: TV-minimization, IHT-TV. All images were obtained after 2000 iterations. Gray-scale window: [0,1]cm^{-1}.

Table II
QUANTITATIVE COMPARIONS OF IMAGES RECONSTRUCTED BY DIFFERENT ALGORITHMS FROM REAL SPECIMEN DATA

Metrics	# Proj.	FDK	POCS	IHT-POCS	TV	IHT-TV
RMSE	5	18.5	3.4	2.9	3.6	2.2
(10^{-2}cm^{-1})	9	14.1	2.7	3.5	2.4	1.6
Corr. Coeff.	5	0.25	0.80	0.79	0.86	0.93
	9	0.36	0.88	0.85	0.91	0.96

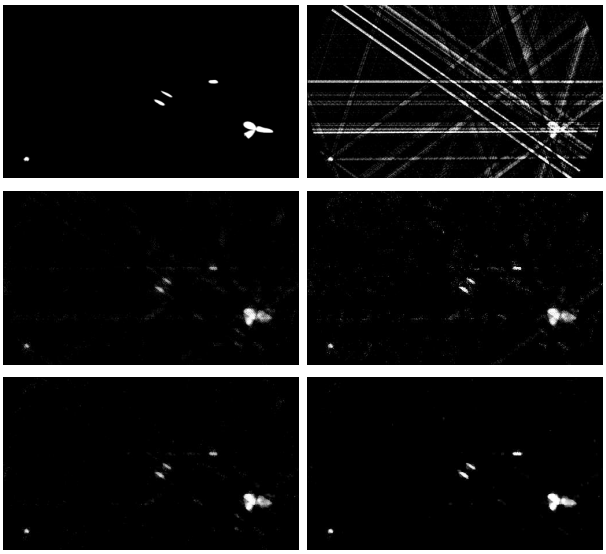


Figure 5. Images at the middle transverse slice of the real coronary artery specimen reconstructed from 5 projections by different algorithms. From left to right, top row: ground truth, FDK. Middle row: POCS, IHT-POCS. Bottom row: TV-minimization, IHT-TV. Gray-scale window: [0.05,0.6]cm^{-1}.

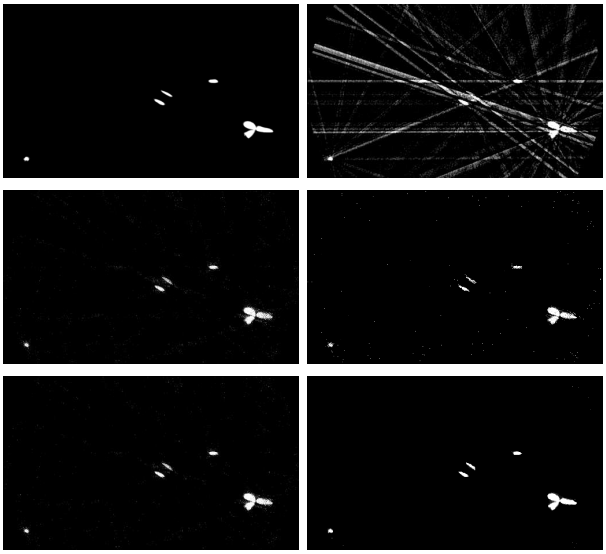


Figure 6. Images at the middle transverse slice of the real coronary artery specimen reconstructed from 9 projections by different algorithms. From left to right, top row: ground truth, FDK. Middle row: POCS, IHT-POCS. Bottom row: TV-minimization, IHT-TV. Gray-scale window: [0.05,0.6]cm^{-1}.

reconstructing coronary artery images from both few-view simulated data and few-view real data. From the reconstructed images and the quantitative assessment results, it can be observed that these optimization-based, sparsity-exploiting algorithms can yield images with higher fidelity than some of the existing, well-known algorithms. Since these algorithms may yield images with practical utility from significantly reduced amount of data, they have potential implications of shortening data-acquisition time as well as minimizing radiation damage to the imaged subject. These advantages may particularly benefit applications such as longitudinal imaging and dynamic CT. We have carried out additional, extensive studies on image reconstruction and quantitative evaluation, which are not included here due to page limit, and we will present these additional results at the conference.

V. Acknowledgments

This work was supported in part by NIH grants CA120540, EB000225, EB000305, HL65342. Partial funding for the computation in this work was provided by NIH grants S10 RR021039 and P30 CA14599. J. Bian was supported in part by a DoD Predoctoral training Grant BC083239. The authors would like to thank Dr. Hiroyuki Kudo for providing the FORBILD numerical phantom.

References

[1] E. L. Ritman, "Micro-computed tomography – current status and developments," *Annu. Rev. Biomed. Eng.*, vol. 6, pp. 185–208, 2004.

[2] M. Li, H. Yang, and H. Kudo, "An accurate iterative reconstruction algorithm for sparse objects: application to 3D blood vessel reconstruction from a limited number of projections," *Phys. Med. Biol.*, vol. 47, pp. 2599–2609, 2002.

[3] M. Li, H. Kudo, J. Hu, and R. H. Johnson, "Improved iterative algorithm for sparse object reconstruction and its performance evaluation with micro-ct data," *IEEE Trans. Nucl. Sci.*, vol. 51, pp. 659–666, 2004.

[4] E. Hansis, D. Schäfer, O. Dössel, and M. Grass, "Evaluation of iterative sparse object reconstruction from few projections for 3-D rotational coronary angiography," *IEEE Trans. Med. Imaging*, vol. 27, pp. 1548–1555, 2008.

[5] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, pp. 489–509, 2006.

[6] E. Y. Sidky, C.-M. Kao, and X. Pan, "Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT," *J. X-Ray Sci. and Technol.*, vol. 14, pp. 119–139, 2006.

[7] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Phys. Med. Biol.*, vol. 53, pp. 4777–4807, 2008.

[8] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Appl. Comput. Harmon. Anal.*, vol. 27, pp. 265–274, 2009.

[9] E. Y. Sidky, M. A. Anastasio, and X. Pan, "Image reconstruction exploiting object sparsity in boundary-enhanced x-ray phase-contrast tomography," *Appl. Opt.*, In press.

[10] `http://www.imp.uni-erlangen.de/phantoms/`.

[11] S. M. Jorgensen, O. Demirkaya, and E. L. Ritman, "Three-dimensional imaging of vasculature and parenchyma in intact rodent organs with x-ray micro-CT," *Am. J. Physiol. Heart Circ. Physiol.*, vol. 275, no. 10, pp. 1103–1114, 1998.

Multisource inverse-geometry CT : progress update and image noise analysis

Bruno De Man, Jorge Uribe, Randy Longtin, Daniel Harrison, Joseph Reynolds, Bogdan Neculaes, Kristopher Frutschy, Louis Inzinna, Antonio Caiafa, Robert Senzig, Jongduk Baek, Norbert Pelc

***Abstract*—These instructions give you guidelines for preparing papers for IEEE TRANSACTIONS and JOURNALS. Use this document as a template if you are using Microsoft *Word* 6.0 or later. Otherwise, use this document as an instruction set. The electronic file of your paper will be formatted further at IEEE. Define all symbols used in the abstract. Do not cite references in the abstract. Do not delete the blank line immediately above the abstract; it sets the footnote at the bottom of this column.**

***Index Terms*—Computed Tomography, Inverse-geometry CT, Multisource CT**

I. Introduction

ALL modern commercial CT scanners are based on the third-generation architecture: x-rays are emitted from a single x-ray focal spot onto a large arc-shaped or area detector. Also the dual-source CT scanner uses essentially the same architecture but with (almost) duplicated x-ray tubes and detectors. We have recently developed a new distributed source technology, enabling a number of radically novel CT architectures, such as stationary CT [1], inverse geometry CT (IGCT) [2-7], and stationary X-ray tomosynthesis systems [8]. Motivated to further reduce patient radiation dose and to cover an organ-in-rotation without compromising image quality, we are performing an in-depth study of the IGCT concept and are developing a gantry-based IGCT research prototype. We have previously presented several IGCT reconstruction algorithms [9] as well as a mathematical framework for the virtual bowtie [10]. We also presented our gantry-based IGCT research prototype in earlier stages of development [3, 4, 6]. We now present the first rotating-gantry experimental IGCT results, as well as a high-level analysis of x-ray source power and image noise.

II. System Overview

This IGCT research prototype – shown in Fig 1 – is designed to demonstrate large volumetric z-coverage (about 160mm) without cone-beam, without scatter artifacts and with good spatial resolution throughout the entire scan volume. The source has two rows of focal spots separated in z by 100mm. The design consists of four source modules, each with two rows of four focal spots, for a total of 32 focal spots. The corresponding in-plane field-of-view is about 22cm in diameter. Initial experiments were performed with only one source module for an in-plane field-of-view of about 7cm diameter. The gantry rotation speed is one revolution per second.

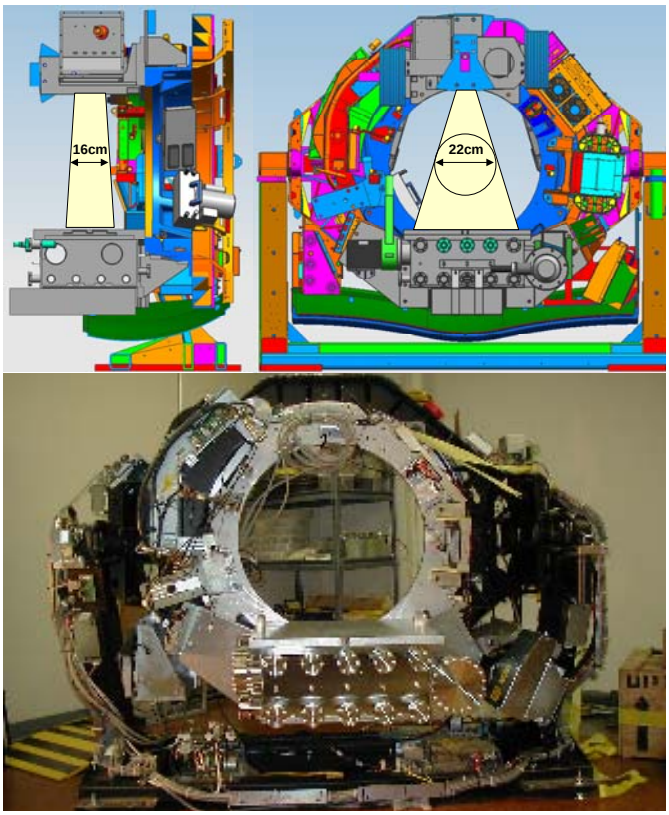


Fig. 1. CAD drawings and picture of the multisource inverse-geometry research prototype.

This work was supported in part by the National Institute of Biomedical Imaging and Bioengineering, part of the National Institutes of Health (NIH) under Grant R01 EB006837.

B. De Man, J. Uribe, R. Longtin, D. Harrison, J. Reynolds, B. Neculaes, K. Frutschy, L. Inzinna, and A. Caiafa are with GE Global Research, Niskayuna, NY 12309 USA (corresponding author : B. De Man, 518-387-7730; deman@ieee.org). R. Senzig is with GE Healthcare, Waukesha, WI 53188 USA. J. Baek and N. Pelc are with Stanford University, Stanford, CA 94305 USA.

III. Experimental Results

The first experiments were performed with stationary gantry and rotating phantom. Some of these first images are shown in Fig 2. The top left image shows a uniform acrylic cylinder mounted on the rotating stage. The top right image is a longitudinal reformat of a hollow plastic tube with acrylic balls taped to the outside for geometric calibration purposes. The bottom left image shows a spheres phantom with a hollow acrylic cylinder filled with a low-density foam and several acrylic spheres to study artifacts and spatial resolution. The bottom right image shows a plastic cylinder filled with trail mix. We plan to acquire the first sinograms with rotating gantry in March and we expect to show rotating-gantry images at the conference.

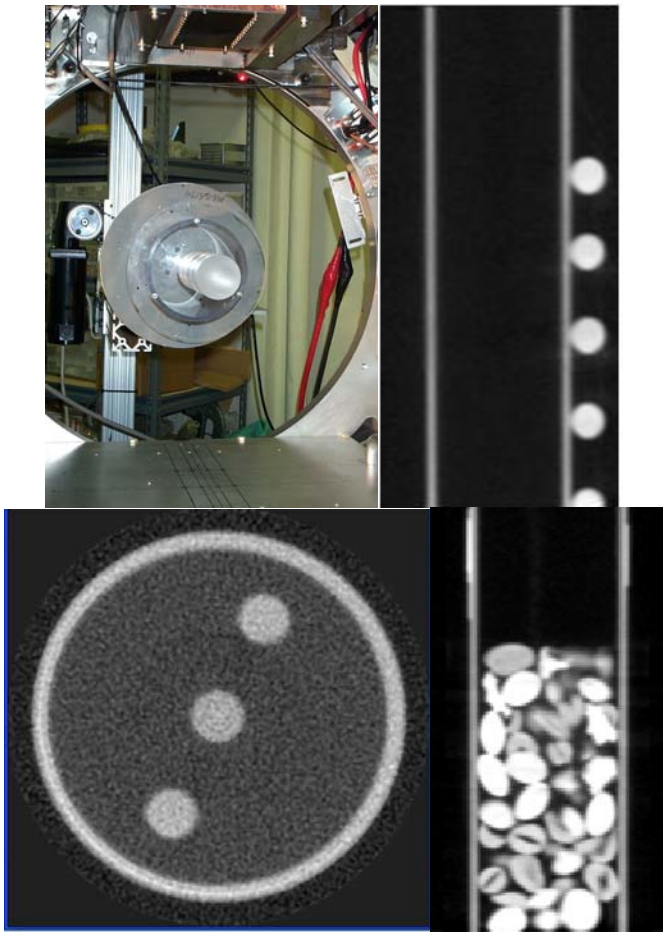


Fig. 2. (top left) Picture of a uniform acrylic cylinder mounted on the phantom rotator inside the multisource CT gantry. The detector active area is seen on top and the source window at the bottom. (top right) Longitudinal reformat of a geometric calibration phantom. (bottom left) First image of a cylindrical test phantoms with spheres for resolution and artifact measurement. (bottom right) Longitudinal reformat of the trail mix phantom.

IV. Gantry balancing and rotation

The CT system consists of a large number of components mounted on a rotating gantry. The x-ray source was developed by this project. Most components are standard GE Healthcare VCT64 and CT750HD product scanner parts. The detector modules are GE Healthcare CT750HD modules but the detector itself was custom designed and built by this project. A number of mechanical fixtures and weight stacks were designed specifically for the IGCT research prototype. The weight stacks were positioned at 5 specific locations on the gantry. Initial balancing was performed analytically and used to define the position of the weight stacks and the nominal value of the weights. For a more refined balancing, the gantry was balanced experimentally using piezo-electric strain gauges attached to the gantry. Fig. 3 shows pictures of the rotating gantry taken at a 4.23s gantry rotation speed. Movies of various gantry rotation speed will be shown at the conference.

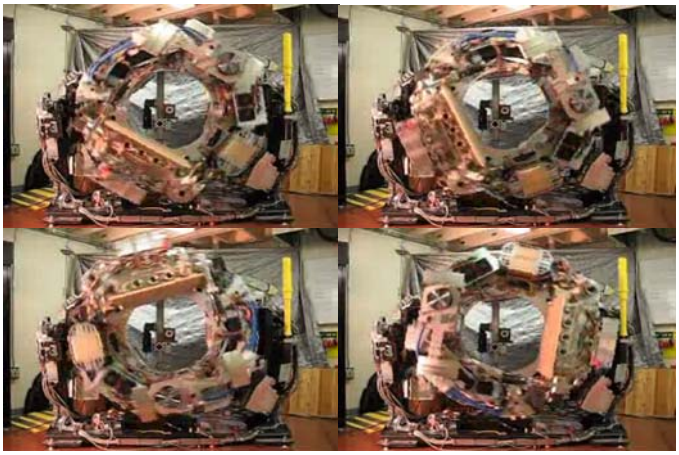


Fig. 3. Four consecutive frames of the gantry rotating at 4.23s per revolution for initial testing.

V. Multisource x-ray tube upgrade

The source was initially tested with a single source module. This staged approach made it possible to validate the module and then replicate it to achieve a larger field-of-view. Fig 4 shows a picture of the source with the original module mounted centrally inside the vacuum chamber. The bottom shows a CAD drawing of the source in its ultimate configuration including 2 rows of 16 spots.

The other source modules are currently being manufactured. Fig 5 first shows a picture of two of the four copper anode blocks. The bottom left picture shows the tungsten target plates that will be brazed onto the copper anode block. The bottom right shows the focusing plates that are part of the cathode assemblies and are designed to focus the electron beams to its desired shape and size.

By the time of the conference the integration of the 32-spot x-ray source should be complete and we will show pictures of the complete assembly as well as possibly some initial validation experiments. Imaging experiments with 32 focal spots are scheduled after the conference.

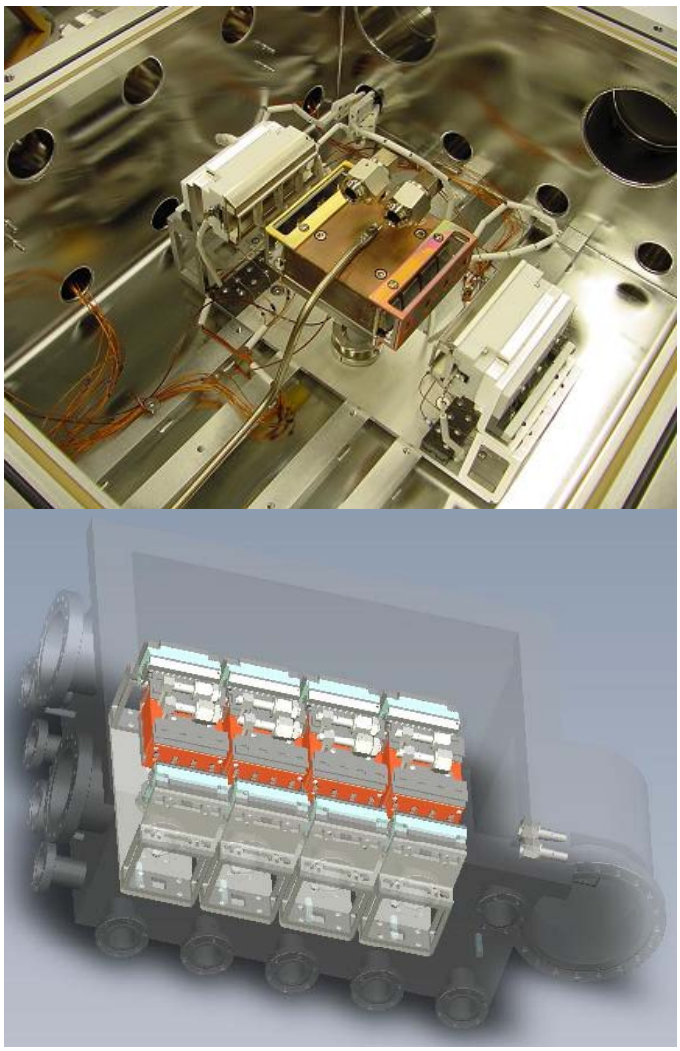


Fig. 4. (top) Picture of the source with the original module mounted centrally inside the vacuum chamber. (bottom) CAD drawing of the source in its ultimate configuration including 2 rows of 16 spots.

VI. Flux and Image Noise Entitlement

The source power and therefore the image noise are limited by several factors. First, the dispenser cathode electron current is limited to about 1200mA without significantly degrading lifetime. This is certainly not a bottleneck in the current multispot x-ray source. Second, HV stability may limit the operation of the tube. In initial tests repeated discharges or "spits" occurred and formed the bottleneck on the power we could achieve so far. We plan to continue seasoning experiments to eliminate spits. The third and most fundamental limiting factor is the target temperature. The source power is ultimately limited by the melting temperature of tungsten. Assuming a cold target, 120kVp tube voltage, a 1mm focal spot (4.62mm^2 thermal area for a 12.5 degree target angle) and a 100us dwell time, to reach the melting temperature of tungsten the tube current in a single x-ray pulse would be 269mA. Beam current can be increased to about 850mA for a single pulse with 10us dwell time. However, for a 10s scan with hundreds of pulses per spot and a repeat time

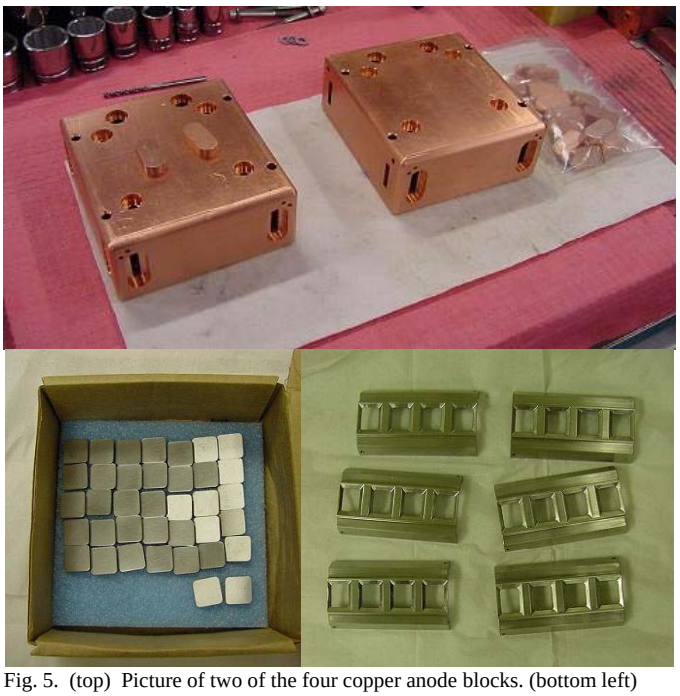


Fig. 5. (top) Picture of two of the four copper anode blocks. (bottom left) Picture of the tungsten target plates that will be brazed onto the copper anode block. (bottom right) Picture of the focusing plates that are part of the cathode assemblies and are designed to focus the electron beams to its desired shape and size.

of 4.3ms the temperature in the focal spots and surrounding area will build up, and the maximum tube current is only 157mA at 120kVp, 1mm focal spot and 100us dwell time. For even longer scan times, the bulk temperature of the source starts to rise and oil cooling becomes critical. We also expect active cooling would reduce the frequency of HV spits. Oil cooling is part of the design but its actual implementation is planned at the end of this project.

For comparison, clinical scanners have tube currents of at least 500mA for this small focal spot size and rotation speeds of 0.35s. Our prototype has a rotation speed of 1s and a voxel duty cycle (fraction of time any given voxel is irradiated) of about 5%. Hence, the integrated flux would be 20 times lower with our research experiments or the noise standard deviation will be about 4 to 5 times higher. Experimentally we can compensate for this by performing slower scan times. Eventually, we foresee an IGCT scanner with reduced source-to-iso distance, improved detector efficiency, reduced dwell times, increased detector size (to improve the voxel duty cycle) and use a combination of the virtual bowtie and statistical reconstruction to achieve acceptable clinical noise levels.

VII. Summary

To our knowledge this is the world's first rotating gantry multi-source inverse-geometry CT system. We have demonstrated multi-source imaging, gantry rotation and by the time of the conference we will have demonstrated the combination of both and completed the integration and initial

testing of a 32-spot x-ray source.

ACKNOWLEDGMENT

The authors would like to thank the machinists at the GE-Research Shop, who have produced many of the multisource CT components., as well as all the other engineers, scientists and support personnel at GE Global Research and GE Healthcare that have contributed to this project.

REFERENCES

[1] B. De Man et al., "Stationary computed tomography system and method," US Patent No. 7,280,631 B2, 2005.

[2] T. G. Schmidt et al, 'An inverse-geometry volumetric CT system with a large-area scanned source: a feasibility study', Med Phys 2004; 9 (31): 2623-7.

[3] B. De Man, S. Basu, D. Bequé, B. E. H. Claus, P. M. Edic, M. Iatrou, J. W. LeBlanc, R. F. Senzig, M. Vermilyea, C. Wilson, Z. Yin, N. J. Pelc , 'Multi-source inverse geometry CT: a new system concept for x-ray computed tomography,' SPIE Medical Imaging Conference 2007 [6510-16], San Diego, CA.

[4] B. De Man, S. Basu, P. Fitzgerald, D. Harrison, M. Iatrou, K. Khare, J. LeBlanc, B. Senzig, C. Wilson, Z. Yin, N. Pelc, 'Inverse geometry CT : the next generation CT architecture ?', IEEE Medical Imaging Conference, Honolulu, 2007.

[5] K. Frutschy, B. De Man, P. Edic, B. Lounsberry, B. Neculaes, Y. Zou, L. Inzinna, K. Conway, X. Zhang, Y. Cao, S. Gunturi, A. Caiafa, J. Reynolds, "X-ray Multisource for Medical Imaging," SPIE Medical Imaging Conference 2009. [725822-1], Lake Buena Vista, FL.

[6] B. De Man, A. Caiafa, Y. Cao, K. Frutschy, D. Harrison, L. Inzinna, R. Longtin, B. Neculaes; J. Reynolds, J. Roy, J. Short, J. Uribe, W. Waters, X. Zhang, Y. Zou, B. Senzig, N. Pelc, "Multi-source inverse-geometry CT: from system concept to research prototype," submitted to IEEE Medical Imaging Conference 2009.

[7] K. Frutschy, B. Neculaes, L. Inzinna, A. Caiafa, J. Reynolds, Y. Zou, X. Zhang, S. Gunturi, Y. Cao, B. Waters, D. Wagner, B. De Man, D. McDevitt, R. Roffers, B. Lounsberry, N. J. Pelc, 'High Power Distributed X-ray Source', Proc. SPIE, San Diego, Feb.2010.

[8] P. Edic, B. Claus, J. Eberhard, J. Miller, D. Langan, M. Vermilyea, J. Manak, C. Unger, "System and method for imaging using distributed x-ray sources", US patent 2007/0009088 A1.

[9] Z. Yin, B. De Man, and J. Pack, "Analytical cone-beam reconstruction using a multi-source inverse geometry CT system," Proceedings of SPIE, Vol. 6510, San Diego, CA, 2007.

[10] J. Sperl, D. Bequé, B. Claus, B. De Man, B. Senzig, and M. Brokate, "Computer Assisted Scan Protocol And Reconstruction (CASPAR) - Reduction of Image Noise and Patient Dose", IEEE Trans. Med. Im., 2009.

4D Iterative Reconstruction in Cardiac CT for Noise & Dose Reduction at Maintained Temporal Resolution

Herbert Bruder, Rainer Raupach, Martin Sedlmair, Karl Stierstorfer
Siemens, HealthCare Division, Forchheim, Germany

Abstract— In this paper, a (non-statistical) four-dimensional (4D) iterative reconstruction scheme is developed to improve noise characteristics and/or reduce radiation exposure in cardiac CT. In our implementation, the update equation of iterative reconstruction is based on Filtered Backprojection (FBP) and the solution is stabilized using nonlinear regularization priors, operating in four dimensions. Image volume datasets are sampled at adjacent temporal positions with respect to a central optimized cardiac phase. The regularization prior operates on a 4D cube surrounding each central voxel in 4D space. Generalized Gaussian Markov Random Field (GGMRF) priors are used to apply non-linear low-pass filtration to the update images, reducing image noise at the same time maintaining image sharpness. Using a local noise estimator the local noise level of the image volume can be accurately estimated to scale the regularization prior. Not only edges in spatial domain but also edges in temporal domain are preserved.

Thus, the temporal resolution of image data is maintained, despite the usage of temporal data that substantially exceed the reconstruction range of a conventional cardiac reconstruction. Consequently, the noise statistics is significantly improved, because non-correlated image data at different temporal positions are utilized.

To reduce the high computational load, the iteration in 4D can be restricted to image space, the regularization operating as a mere iterative image filter followed by a conventional 3D iterative reconstruction.

We demonstrate the potential of noise reduction on basis of clinical cardiac CT data. As an example, for cardiac dual source CT (DSCT) data, a noise reduction of 60% was achieved, while maintaining spatial and temporal image sharpness. Even in case of a very high, irregular heart beat with average heart rate of 115 beats per minute (bpm) the temporal resolution was fully restored. The potential for noise reduction can be utilized to effectively reduce radiation exposure at the same time maintaining image quality.

Index-Terms: FBP-based iterative reconstruction, GGMRF, regularization prior, cone-beam cardiac CT, dual source CT (DSCT)

H. Bruder, SIEMENS, Healthcare Division,
Siemenstr. 1, 91301 Forchheim, Germany,
Phone: +49-9191-188793
Fax: +49-9191-189996
e-mail: herbert.bruder@siemens.com

I. Introduction

In cardiac CT, among others, high temporal resolution and low image noise are prerequisites for accurate diagnosis of coronary artery disease. In addition, cone and spiral artifacts might detoriate image quality, if detectors with large cone-angles are used.

Moreover, cardiac scan modes apply high radiation dose to the patient, hence, any means to reduce the X-ray exposure to the patient is desirable. Basically, image noise can be reduced by increasing the tube current. Yet, at the same time this increases patient dose. Also, image noise can be reduced extending the range of the stream of projection data used for image reconstruction. However, by doing so the temporal resolution of the image is impaired, which might cause motion artifacts. In general, motion artifacts obscure a meaningful diagnosis of coronary disease. The question is how to improve the noise statistics of a cardiac image at constant radiation exposure and maintained temporal resolution? In this paper we present a 4D iterative reconstruction scheme to improve noise characteristics in cardiac CT.

The reconstruction is based on Iterative Weighted Filtered Backprojection (I-WFBP) reconstruction. We have applied the iterative improvement scheme, illustrated in Fig. 1, to the non-exact WFBP method proposed by Stierstorfer et al. [1]. Basically the reconstruction is of the Feldkamp-type with pre-filtration of cone-beam data. In the helical acquisition mode at table speeds below the maximum redundant data are efficiently processed on purpose of optimal dose usage.

To stabilize the solution and to improve the convergence of the iteration loop, regularization priors are used which are based on GGMRF functions [3]. Image sharpness can be preserved by virtue of non-quadratic prior functions. The amount of noise reduction is controlled by the regularization strength and the number of iterations.

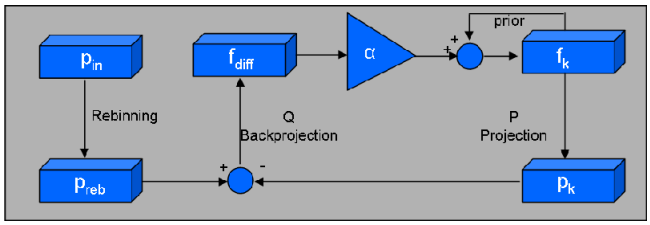


Fig. 1: Illustration of the Iterative Weighted Filtered Backprojection (I-WFBP) method. First, input data are rebinned to semi-parallel geometry. Given an initial image vector $f_0 \in R^N$*, a sequence of image vectors is generated by the update formula* $f_{k+1} = f_k + \alpha \cdot Q \cdot (p_{reb} - P f_k)$*. The matrices* $Q \in R^{NxM}$ *and* $P \in R^{MxN}$ *correspond to the WFBP method and a projection operator modelling the acquisition process, respectively.*

II. Method

A. Iterative Weighted Filtered Backprojection (I-WFBP)

Let N be the number of voxels of the image volume and M the total number of x-ray attenuation measurements. Furthermore, let $p_{reb} \in R^M$ denote input data that have been rebinned to semi-parallel cone-beam data and $f_0 \in R^N$ denote a vector representing the initial voxel volume, which is obtained by WFBP reconstruction. The update step of I-WFBP is then given by

$$(1) \qquad f_{k+1} = f_k + \alpha_k \cdot Q \cdot (p_{reb} - P \cdot f_k)$$

where Q is the reconstruction operator from the previous section and $P \in R^{MxN}$ is the projection operator. In this way, a sequence of voxel volumes $\{f_0, f_1, \ldots\}$ is produced. Due to the pre-filtration of the rebinned projection data with a sharp convolution kernel, the point-spread function of the image signal is estimated and has not to be approximated by iteration. This substantially improves the convergence rate of the iteration. The second term on the right side of eq. (1) is called correction term and manages reduction of cone-beam and spiral artifacts that are introduced by non-exact image reconstruction. Experiments have shown that only a few iterations are needed to obtain significant reduction of cone-and spiral artifacts.

B. Regularization

To stabilize the update image of eq. (1) a penalty term has to be added which imposes constraints on adjacent image pixels. Eq. (1) then reads as

$$(2) \qquad f_{k+1} = f_k + \alpha_k \cdot Q \cdot (p_{reb} - P \cdot f_k) - \beta_k \cdot R(f_k)$$

e.g. R being a regularization filter based on GGMRF priors [3]. This class of regularization priors is characterized by a range filter operating on the gradients of image grey scale values and a domain filter, which can be realized by the inverse distance of adjacent pixels in the simplest case (see eq. (4)). One class of range filters consists of potential functions that can be cast as follows:

$$(3) \qquad V(df) = \frac{|df|^p}{1 + \left|\frac{df}{c}\right|^{(p-q)}}$$

where df denotes the local image contrast. p and q are parameters satisfying $1 \le q \le p \le 2$ for convex functions. c is a local measure of image noise, which can be derived, minimizing local noise variances in the image volume. This class of regularization priors provides high flexibility in terms of adjusting contrast dependent image sharpness and image noise.

Basically, the update equation (2) holds for a 3D voxel volume f. In this case the voxel volume in the *k-th* loop is obtained by the previous update and a weighted sum of the correction image and the regularization image.

The regularization part can be written as:

$$(4) \qquad R(f_k) = \sum_{i=1}^{N} e_i \cdot \sum_{j=1}^{K} d_{ij} \cdot \frac{dV}{df}(f_k(i) - f_k(j))$$

Here, e_i denotes the unit vector with a 1 at position $i \in (1, N)$ voxels. K specifies the number of nodes adjacent to the momentary central voxel i. The central voxel is shifted through the entire image volume.

C. 4D regularization in cardiac CT

In cardiac CT redundant data acquisition is required, to potentially reconstruct volume data at different cardiac phases. For example in spiral acquisition modes, the feed has to be limited, such that for a given z-position any cardiac phase within the cardiac cycle can be reconstructed, leading to a high redundancy of measured data. Thus, it is possible to reconstruct image volumes at adjacent cardiac phases and extend the voxel volume to a 4D voxel volume f_k (e.g. this volume can be organized by sequentially appending the 3D volumes f_k^l with $l \in (1, L)$, where L denotes the total number of adjacent temporal positions at which volume data have been reconstructed). Thus eq. (2) has to be extended:

$$(5) \qquad f_{k+1}^l = f_k^l + \alpha_k \cdot Q \cdot (p_{reb}^l - P \cdot f_k^l) - \beta_k \cdot R_{4D}(f_k)$$

The regularization part can be combined to a 4D regularizer:

$$(6)\qquad R_{4D}(f_k) = \sum_{i=1}^{N\cdot L} e_i \cdot \sum_{j=1}^{K\cdot L} d_{ij} \cdot \frac{dV}{df}(f_k(i) - f_k(j))$$

Here, e_i denotes the unit vector with a *1* at position $i \in (1, N \cdot L)$ voxels. $K{\cdot}L$ specifies the number of nodes adjacent to the momentary central voxel *i*. The central voxel is shifted through the entire 4D image volume f_k.
The cube of nodes adjacent to a central voxel becomes a 4D cube extending also in temporal domain. Basically eq. (5) can be updated slice by slice, using the respective projection data p^l_{reb} for different cardiac phases. The correlation between cardiac phases and projection data is given by the ECG of the patient.
However, the computation load in case of 4D cardiac CT data is very high. A way out of this is a two step approach. In a first step eq. (5) is modified to a mere image filter by setting the correction term to zero, i.e. $\alpha_k = 0 \forall k$. The update equation can be condensed and simplifies to

$$(7)\qquad f_{k+1} = f_k - \beta_k \cdot R_{4D}(f_k)$$

We have to keep in mind that f_k is a 4D image volume. This equation is updated in the vicinity of the selected (optimal) cardiac phase. Basically being a non-linear low-pass filter on 4D image data, the update equation does not converge. The stopping criterion has to be determined empirically. Due to the special type of non-linear regularization temporal and spatial edges can be preserved by a proper selection of the GGMRF prior. At the same time, noise characteristics can be substantially improved on homogeneous regions, due to the correlations induced to the noise in tissue with low contrast. The characteristic line as a function of image contrast depends on the choice of the GGMRF prior.
After this first iteration we take that 3D image volume $\hat{f}$ out of the 4D volume data that corresponds to the optimal cardiac phase. Thus, the iteration problem has been reduced by one dimension. Eq. (2) can be rewritten as

$$(8)\qquad \hat{f}_{k+1} = \hat{f}_k + \alpha_k \cdot Q \cdot (p_{reb} - P \cdot \hat{f}_k) - \beta_k \cdot R(\hat{f}_k)$$

Under certain conditions (see [1]), the iteration converges. Consequently, due to the correction term

$$\alpha_k \cdot Q \cdot (p_{reb} - P \cdot \hat{f}_k),$$

also cone-beam and spiral artifacts can be reduced.

III. Evaluation

In the evaluation of our iterative reconstruction for cardiac cone-beam data, we concentrate on cardiac dual source CT (DSCT) data. In DSCT two X-ray sources and two corresponding measurement systems are mounted on a CT gantry with a mechanical offset of 90^0. One detector (A) covers the entire scan field of view (SFOV), while the other detector (B) is restricted to a smaller, central field of view. Due to the simultaneous data acquisition and the 90^0 angular offset, complementary data segments are measured resulting in reduction of exposure time per image by a factor of 2, and thus the temporal resolution is increased by the same factor. With a rotation time of the CT gantry of *285 ms*, a temporal resolution of *75 ms* is established (method 1). To improve noise characteristics of the DSCT images, the reconstruction ranges of data from system A and B, respectively can be enlarged [4], however, compromising the temporal resolution (method 2, see Fig.2). As derived above, the iterative approach (method 3) utilizes the improved noise statistics due to the enlarged projection ranges going into the update equation (5), at the same time preserving contrast edges in space and time. Fig. 3 shows an example from a patient with extremely high, irregular heart beat. Due to the reduced exposure time of DSCT, the Right Coronary Artery (RCA) is sharply delineated even at this high heart rate. Using method 2 with data ranges of length 180^0 for both systems A and B, noise is actually reduced, but severe motion artifacts are observed, due to the doubled exposure time. Using method 3, temporal resolution is maintained compared to method 1, as well reducing image noise substantially.
Fig.4 shows a second example from a patient with a stable heart beat of average heart rate of 61 bpm. Both in the ascending aorta and in the liver, circular areas in homogenous regions have been evaluated regarding image noise. As expected, method 2 shows noise reduction near to a factor of $\sqrt{2}$ due to the double size reconstruction interval. The noise reduction for method 3 is between *45%* in the aorta to *60%* in the liver region. The large noise reduction is by virtue of the 4D regularization procedure utilizing the non-correlated image data at adjacent cardiac phases.

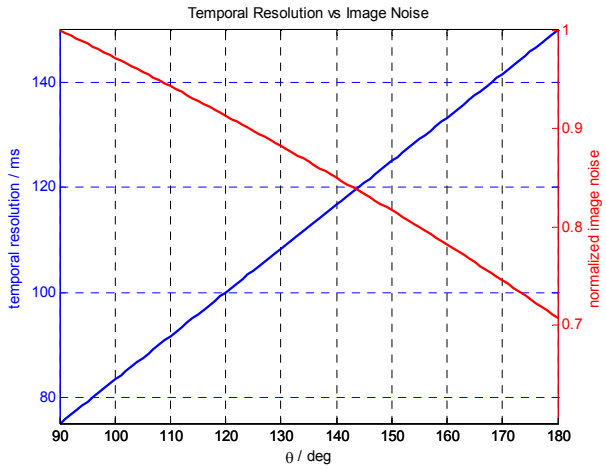


Fig. 2 Temporal resolution and normalized image noise as a function of the reconstruction interval θ of detector A and B – data, respectively. Increased reconstruction interval decreases image noise at the same time increasing the exposure time of the image, and thus decreasing temporal resolution.

IV. Conclusion

We presented a new reconstruction scheme for cardiac CT that enables the utilization of redundant cardiac data to improve the noise statistics without increasing the radiation exposure to the patient, at the same time maintaining temporal resolution. As an example, we addressed cardiac DSCT imaging. In DSCT one major claim is, that coronary artery disease can be diagnosed at any heart rate. However, e.g. in case of obese patients with high image noise, temporal resolution has to be sacrificed in favor of the capacity to decrease image noise. This would restrict the cardiac application of DSCT to moderate heart rates. However, with the new reconstruction scheme image noise can be reduced substantially (*60%*), while maintaining spatial resolution and temporal resolution. The huge potential for noise reduction can also be utilized for reduction of radiation exposure at constant image noise. The presented scheme not only applies to DSCT, but also to Single Source CT (SSCT). In any case, noise can substantially be reduced while maintaining the temporal resolution provided by the fast rotation time of today's CT scanners.

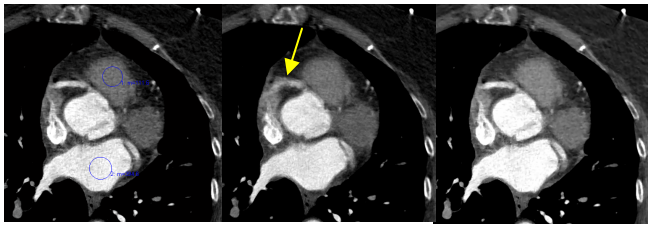
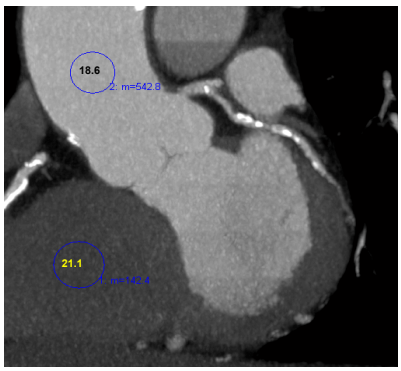


Fig. 3: Axial images at the RCA outlet reconstructed from a cardiac DSCT dataset. The patient showed a highly irregular ECG with an average heart rate of 115 bpm. (left) method1 (reconstruction interval $\theta = 90^0$): cardiac DSCT reconstruction with temporal resolution of 75 ms (middle) method2 ($\theta = 180^0$): cardiac DSCT reconstruction with temporal resolution of 150 ms (right) method 3: proposed iterative reconstruction with temporal resolution of 75 ms. The regularization prior was parameterized with p = 2, q = 1.2, $\beta_k = 1 \forall k$. Please note that for this high heart rate, method 2 fails due to the severe motion artifacts in the proximal part of the RCA (arrow).



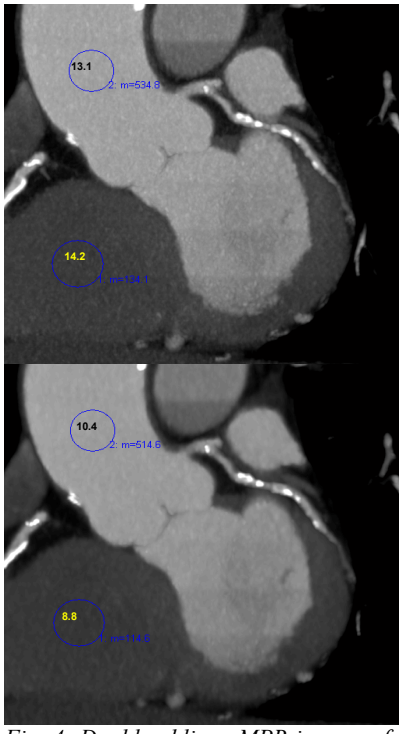


Fig. 4: Double oblique MPR images of a cardiac DSCT data set. The patient showed a regular ECG with an average heart rate of 61 bpm. The initial FBP – reconstruction was performed using a conventional cardiac reconstruction kernel (B46f) (left bottom) method 1 ($\theta = 90^0$): temporal resolution of 75 ms (top) method2 ($\theta = 180^0$): temporal resolution of 150 ms (middle) method 3: proposed iterative reconstruction with temporal resolution of 75 ms. The regularization prior was parameterized with p = 2, q = 1.2, $\beta_k = 1 \forall k$. Both in the ascending aorta and the liver, circular areas in homogenous regions have been evaluated with respect to image noise. As expected, method 2 shows noise reduction near to a factor of √2 due to the double size reconstruction interval. The noise reduction for method 3 was between 45% in the aorta and 60% in the liver region. The large noise reduction is by virtue of the 4D regularization procedure utilizing the (non-correlated) image data at adjacent cardiac phases.

V. References

1. J. Sunnegård, Combining Analytical and Iterative Reconstruction in Helical Cone-Beam CT, Thesis No 1301, Institute of Technology, Linkøping (2007)
2. K. Stierstorfer, A. Rauscher, J. Boese, H. Bruder, S. Schaller and T. Flohr, Weighted FBP - a simple approximate 3DFBP algorithm for multislice spiral CT with good dose usage for arbitrary pitch, Phys. Med. Biol. **49**, 2209–2218 (2004)
3. J. Thibault et al., A three-dimensional statistical approach to improved image quality for multislice helical CT, Med. Phys. **34**, 4526 (2007)
4. T. Flohr et al., Image reconstruction and image quality evaluation for a dual source CT scanner, Med. Phys. **35** 5882 (2008)

ECG-Gated Cardiac Reconstruction for Non-Periodic Motion

Christopher Rohkohl, Günter Lauritsch, Lisa Biller and Joachim Hornegger

***Abstract*—The 3-D reconstruction of cardiac vasculature using C-arm CT (rotational angiography) is an active and challenging field of research. Current reconstruction algorithms do strongly depend on a high degree of cardiac motion periodicity for working properly. In the interventional environment patients often do have arrhythmic heart signals or cannot hold breath during the complete acquisition. This represents one important class of clinical cases that cannot be reconstructed with current approaches.**

We present a motion compensated gated reconstruction algorithm consisting of a 4-D time-continuous affine motion model which is capable of reconstructing datasets with highly non-periodic motion patterns. A time-correlated objective function is introduced which measures the error between the measured projection data and the dynamic forward projection of the motion compensated gated reconstruction. Reconstruction results are provided using the data of the open CAVAREV-platform. Further, a clinical case of coronary arteries corrupted by breathing motion is investigated. The results show that the proposed algorithm provides excellent reconstruction quality in cases where classical approaches achieve totally degraded image quality.

I. INTRODUCTION

A. Purpose of this Work

ONE key component of image guidance in the field of interventional cardiology is three-dimensional image information before, during and after interventional procedures. Three-dimensional image data can support complex interventional procedures, such as transcutaneous valve replacement, interventional therapy of atrial fibrillation, implantation of biventricular pacemakers and the assessment of myocardial perfusion [1], [2].

With the technology of C-arm CT it is possible to reconstruct intraprocedural 3-D images from angiographic projection data [3]. Currently, the major limitation of this technology is its insufficient temporal resolution which limits the visualization of fast moving parts of the heart. Due to the long acquisition time of several seconds, at which a couple of heart beats and breathing motion can occur, motion related image artifacts, e.g. blurring or streaks are observed. Therefore it is essential to develop algorithms that can cope with cardiac and respiratory motion.

B. State-of-the-Art

An established technique for time-resolved cardiac reconstruction is to record the electrocardiogram (ECG) during the data acquisition. Based on the ECG-signal a relative cardiac phase is assigned to each projection image assuming a cyclic heart motion [4]. The phase information is used for a phase-correlated reconstruction by gating or motion estimation and compensation. A gated reconstruction takes only those images into account that lie inside a defined phase window, that is centered at the targeting cardiac phase to be reconstructed [5], [6], [7], [8]. This is, however, not ideal in terms of missing data for single-run acquisitions. The incomplete data leads to streak artifacts and a poor signal-to-noise ratio. To increase the data usage, the cardiac motion is estimated and motion compensated reconstruction algorithms are applied [9], [5], [10], [11].

The previous methods were shown to provide reasonable results in the presence of regular heart rates without breathing or other patient motion. However, in the field of intraprocedural cardiac reconstruction, the patients suffer from heart diseases and cannot completely hold breath, stay still or have irregular heart beats. Those aspects do conflict with the periodicity assumption of ECG-based methods.

In literature these problems were addressed by approximate 2-D corrections in the projection image [12]. In a previous work [13] we presented an algorithm that is able to estimate 4-D non-periodic motion patterns using a time-continuous B-spline motion model. Both algorithms require an initial 3-D reference image as a priori information. This reference is easy to obtain for slight breathing motion or arrhythmic cardiac motion by ECG-gated reconstruction. However, strong breathing motion can make it impossible to obtain a sufficient initial image. This weak point is to be solved in this paper.

C. Outline

In this paper a method for motion estimation and compensation of an ECG-gated reconstruction for non-cyclic motion patterns is proposed. It is based on the assumption that breathing motion and heart irregularities around a certain heart phase can be assumed to be an affine motion. For that a time-continuous affine motion model is introduced which is estimated in an optimization procedure. The motion parameters are calculated such that the similarity between the measured projection data and the dynamic forward projection of the corresponding 3-D reconstruction is maximized.

It is demonstrated that the algorithm is capable to estimate cardiac motion and breathing motion for obtaining reasonable reconstruction results of cardiac vasculature.

C. Rohkohl and J. Hornegger are with the Pattern Recognition Lab, Department of Computer Science, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany, Email: christopher.rohkohl@informatik.uni-erlangen.de. G. Lauritsch and L. Biller are with the Siemens AG, Healthcare Sector, Forchheim, Germany.

II. Affine 4-D Motion Estimation and Reconstruction Algorithm

A. General Idea

The quality of ECG-gated reconstructions highly depends on the periodicity of the motion. If the physical motion state of the heart varies for the scanned heart beats, this poses a problem for reconstruction algorithms that assume periodicity. Two major sources of non-periodicity can be identified. First, for varying heart rates, i.e. arrythmias, the ECG-phases cannot be correlated exactly to a physical motion state of the heart [4]. Furthermore, breathing motion adds a second motion component leading to a non-periodic motion.

However, in previous investigations, e.g. by Shechter et al. [14] it has been shown that heart phase variations and respiratory motion of cardiac vasculature can be modelled using global transformations, i.e. rigid body transformations or affine transformations. In the following sections an algorithm for the estimation of such an affine motion between the different heart beats is presented.

B. Affine 4-D Time-Continous Motion Model

We assume a time-continous motion model that maps a voxel $\boldsymbol{x} = (x_0, x_1, x_2)^T$ to a new voxel location $\boldsymbol{x}'$ for each time when a projection image is acquired. It is conceptionalized by a function $M : \mathbb{N} \times \mathbb{R}^3 \times \mathbb{S} \mapsto \mathbb{R}^3$ with $M(i, \boldsymbol{x}, \boldsymbol{s}) = \boldsymbol{x}'$ transforming the voxel coordinate $\boldsymbol{x}$ at the time of the i-th projection image. The mapping is based on the motion model paramters $\boldsymbol{s} \in \mathbb{S}$. In this work a global affine motion model is used.

For that a set of time points is created. Each time point is assigned 12 affine parameters that describe the affine transformation at that moment. The set of time points is determined from the ECG-signal. We are choosing two time points per heart beat. In particular it is the reference heart phase h_r and an additional phase $h_\mathrm{r} + \Delta h$. For both heart phases a nearest-neighbor gating is performed, i.e. only the projection which is closest to the desired point in the cardiac cycle for each acquired heart beat is selected. For a heart phase h this set of projection images is denoted $\mathcal{N}_h$. The complete set L of temporal control points is then given by

$$L = \{1, N\} \cup \mathcal{N}_{h_\mathrm{r}} \cup \mathcal{N}_{h_\mathrm{r}+\Delta h}\,, \tag{1}$$

where the number of projections ranges from 1 to N. The first and last projection image is added such that no boundary problems will occurre in the following.

The affine parameters $\boldsymbol{s}_l \in \mathbb{R}^{12}$ for a single time point $l \in L$ is a 12-element vector with the following components: $\boldsymbol{s}_l = (t_0, t_1, t_2, \alpha_0, \alpha_1, \alpha_2, a_0, a_1, a_2, b_0, b_1, b_2)^T$, where t_i represents the translation along, α_i the rotation around, a_i the scaling along and b_i the shearing of the i-th coordinate axis. The complete parameter vector $\boldsymbol{s} \in \mathbb{S}$, $\mathbb{S} = \mathbb{R}^{12|L|}$ is then given by

$$\boldsymbol{s} = (\boldsymbol{s}_{L_1}, \ldots, \boldsymbol{s}_{L_{|L|}})^T \tag{2}$$

with L_i beeing the i-th smallest element of L.

For an arbitrary projection image i the affine transform parameters $\tilde{\boldsymbol{s}}_i$ are then obtained by temporal interpolation of each component. In our work a cubic B-spline interpolation [15] has been used.

The final motion model is then formally given by

$$M(i, \boldsymbol{x}, \boldsymbol{s}) = \boldsymbol{x}' \quad \text{with} \quad \boldsymbol{A}_{\tilde{\boldsymbol{s}}_i} \begin{pmatrix} \boldsymbol{x} \\ 1 \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}' \\ 1 \end{pmatrix} \tag{3}$$

where $\boldsymbol{A}_{\tilde{\boldsymbol{s}}_i}$ is the affine transformation matrix in homogenous coordinates for the affine parameters $\tilde{\boldsymbol{s}}_i$. The creation of the affine transformation matrix can be found in standard math or computer vision literature, e.g. [16].

C. Motion Compensated ECG-Gated Reconstruction

1) Projection Image Preprocessing: For motion estimation we are only interested in the motion of the cardiac vasculature. Therefore we apply a background reduction technique previously proposed by Hansis et al. [8]. By applying a morphological top-hat filter the vasculature remains and the background is removed to a large degree. In the following this preprocessed projection data will be referred to by the function $p : \mathbb{N} \times \mathbb{R}^2 \mapsto \mathbb{R}$ where $p(i, \boldsymbol{u})$ returns the value of the i-th preprocessed projection image at the pixel $\boldsymbol{u}$.

2) Reconstruction Algorithm: For motion estimation and correction a dynamic reconstruction algorithm $f(\boldsymbol{x}, \boldsymbol{s})$ is defined. The function f returns the reconstructed object value at a voxel $\boldsymbol{x}$ based on the motion model parameters $\boldsymbol{s}$. In principle, any dynamic reconstruction algorithm could be used. In this paper, an extension of the FDK reconstruction method for moving objets is utilized [5], [10]. The ECG-gating is performed by applying a weighting factor λ to each image which is calculated from the closeness to the reference heart phase. The dynamic ECG-gated FDK reconstruction $f_{h_\mathrm{r}} : \mathbb{R}^3 \times \mathbb{S} \mapsto \mathbb{R}$ is then given by

$$f_{h_\mathrm{r}}(\boldsymbol{x}, \boldsymbol{s}) = \sum_i \lambda(i, h_\mathrm{r}) \cdot w(i, M(i, \boldsymbol{x}, \boldsymbol{s})) \cdot p_F(i, A(i, M(i, \boldsymbol{x}, \boldsymbol{s}))) \tag{4}$$

The function $w : \mathbb{N} \times \mathbb{R}^3 \mapsto \mathbb{R}$ is the distance weight of the FDK-formula. The pre-processed, filtered and redundancy weighted projection data is accessed by the function $p_F : \mathbb{N} \times \mathbb{R}^2 \mapsto \mathbb{R}$ where $p_F(i, \boldsymbol{u})$ returns the value of the i-th image at the pixel $\boldsymbol{u}$. The pixel location $\boldsymbol{u}$ is determined by the perspective projection $A : \mathbb{N} \times \mathbb{R}^3 \mapsto \mathbb{R}^2$, where $A(i, \boldsymbol{x}) = \boldsymbol{u}$ maps a voxel $\boldsymbol{x}$ to a pixel location $\boldsymbol{u}$ in the i-th projection image.

The function λ is a weighting function that is used for obtaining an ECG-phase correlated reconstruction for the heart phase $h_\mathrm{r} \in [0, 1]$. It is given by

$$\lambda(i, h_\mathrm{r}) = \begin{cases} \cos^\beta\left(\frac{d(h(i), h_\mathrm{r})}{\omega}\pi\right) & \text{if } d(h(i), h_\mathrm{r}) \le \frac{\omega}{2} \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

where $h(i)$ is the heart phase of the i-th projection image and $\omega \in (0, 1]$ is the width of the non-zero support region of the weighting function. The parameter $\beta \in [0, \inf)$ controls the shape of the support region, e.g. a value of zero corresponds to a rectangular shape [10]. The function d is a distance function which measures the distance between two motion phases. For a relative heart phase it is defined as $d(h_1, h_2) = \min_{c \in \{0,1,-1\}} |h_1 - h_2 + c|$.

D. Objective Function for Motion Estimation

Motion estimation is formulated as a multi-dimensional optimization problem where the motion model parameters $\hat{\boldsymbol{s}} \in \mathbb{S}$ maximizing the objective function $\mathcal{L} : \mathbb{S} \mapsto \mathbb{R}$ need to be estimated, i.e.

$$\hat{\boldsymbol{s}} = \arg\max_{\boldsymbol{s}\in\mathbb{S}} \mathcal{L}(\boldsymbol{s}) \,, \tag{6}$$

The objective function introduced in this paper is motivated by the basic relationship of the motion compensated reconstruction f with the measured projection data p. Maximum intensity projections can be created from a reconstruction $f(\boldsymbol{x}, \boldsymbol{s})$ by dynamic forward projection:

$$r(i, \boldsymbol{u}, \boldsymbol{s}) = \max_{\boldsymbol{x}\in L_{i,\boldsymbol{u}}} f_{h_\mathrm{r}}\left(M^{-1}(i, \boldsymbol{x}, \boldsymbol{s}), \boldsymbol{s}\right). \tag{7}$$

The function $r : \mathbb{N} \times \mathbb{R}^2 \times \mathbb{S} \to \mathbb{R}$ returns the dynamic maximum intensity forward projection of the ECG-gated and motion compensated reconstruction f_{h_r}. The voxels on the straight measurement ray $L_{i,\boldsymbol{u}}$ of the i-th image hitting the detector at pixel $\boldsymbol{u}$ are transformed by the inverse motion model to consider the motion state observed at the projection image i.

The matching of the measured and preprocessed data p and the forward projected data r is assessed by computing the average normalized cross-correlation (NCC). Formally, our objective function is then given by:

$$\mathcal{L}(\boldsymbol{s}) = \tfrac{1}{\nu}\textstyle\sum_i^N \left(\lambda(i, h_\mathrm{r}) \sum_{\boldsymbol{u}} \frac{(p(i,\boldsymbol{u})-\mu_{p_i})(r(i,\boldsymbol{u},\boldsymbol{s})-\mu_{r_{i,s}})}{\sigma_{p_i}\sigma_{r_{i,s}}}\right) \tag{8}$$

with the normalizing factor $\nu = (I_\mathrm{p}-1)\sum_i^N \lambda(i, h_\mathrm{r})$, I_p being the number of image pixels $\boldsymbol{u}$ of a projection and μ, σ being the mean and standard deviation of the subscripted image.

The value of the objective function ranges $\mathcal{L}(\boldsymbol{s}) \in [-1, 1]$ with the maximum value representing a perfect linear relationship of the measured and forward projected data. The NCC of the i-th projection image pair is weighted by the gating function λ as it characterizes the influence on the dynamic ECG-gated reconstruction.

E. Optimization Strategy

For maximization of Eq. (8) a stochastic gradient asscent method is used. In each iteration one of the 12 affine parameters is selected with a certain probability. For initialization, all parameters are assigned the same probability. The gradient is computed using finite differences by varying the selected parameter for all time points. Next, one step is taken into the gradient direction with a fixed step size. The probability of the parameter being selected in the next iteration, is set proportional to the increase of cost function value. Optimization stops after a certain number of iterations or if the convergence ratio drops below a certain threshold.

This kind of stochastic selection of parameters for the gradient-based optimization procedure has the benefit that it picks out the most beneficial parameters. In that way a fast convergence with as few as possible cost function evaluations was found. This especially holds because mainly the non-periodic parts of the motion are caused by translational or rotational components which are preferred during optimization if they gain a greater gain in objective function value.

F. Implementation Details

One evaluation of the objective function Eq. (8) comprises an ECG-gated reconstruction, an ECG-gated forward projection and the computation of the quality measure. Each step is very well parallelizable on the graphics card [17], [18]. The algoritm was implemented on the GPU using the CUDA 2.2 programming environment [19]. The backprojection of the FDK-reconstruction and forward projection are based on projection matrices. The affine matrix $\boldsymbol{A}_{\tilde{s}_i}$ in Eq. (3) only depends on the projection geometry and is independent from the voxel location. This allows us to replace the voxel-wise computation of the motion transform M by a right-side multiplication of the projection matrix with the affine transformation matrix. Consequently, no additional overhead is introduced during forward or backward projection operations.

III. Experimental Setup

A. Cavarev-*Platform*

The proposed algorithm has been applied to one of the datasets of the open Cavarev-Platform [20]. It is an freely available phantom dataset based on true patient dataset with combined cardiac and breathing motion. The standard ECG-gated reconstruction and the motion compensated reconstruction are visually compared. The number of iterations was set to 300. The reconstruction parameters were set to $h_\mathrm{r} = 0.95$, $h_\mathrm{r} + \Delta h = 0.45$, $\omega = 0.25$, $\beta = 0$. As a baseline, the same dataset with a strictly periodic motion is reconstructed with the same parameters. Image reconstruction is performed on an image volume of 20^3 cm^3 distributed on a 256^3 voxel grid.

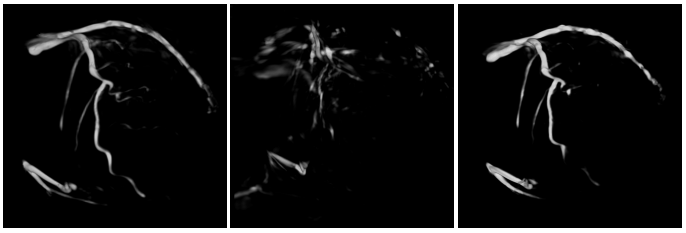
B. Clinical Data

The presented algorithm has been applied to the reconstruction of a right coronary artery dataset which was corrupted by breathing motion because the patient ignored the breathing command. The dataset was acquired on an Artis Zeego C-arm systems (Siemens AG, Healthcare Sector, Forchheim, Germany). It consists of $N = 133$ projection images acquired in 5 seconds with a size of 960×1240 pixels at an isotropic resolution of 0.32 mm/pixel. The standard ECG-gated reconstruction and the motion compensated reconstruction are visually compared. The number of iterations was set to 300. The reconstruction parameters were set to $h_\mathrm{r} = 0.7$, $h_\mathrm{r}+\Delta h = 0.2$, $\omega = 0.2$, $\beta = 0$. Image reconstruction is performed on an image volume of 20^3 cm^3 distributed on a 256^3 voxel grid.

IV. Results and Discussion

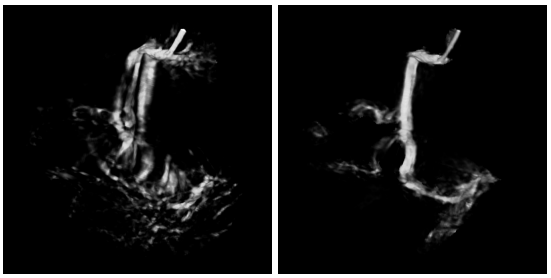
A. Cavarev-*Platform*

In Fig. 1 the reconstruction results are depicted. It can be seen that the standard ECG-gated reconstruction leads to very bad image quality for the non-periodic motion pattern. The vasculature cannot be located in the resulting 3-D reconstruction. In contrast, the proposed method can fully recover the vascular structure. The reconstruction quality is comparable to the reference with a periodic motion.



(a) Baseline (b) Standard (c) Proposed

Fig. 1: Reconstruction results for the dataset of the CAVAREV-platform. (a) ECG-gated reconstruction for the strictly periodic dataset as baseline. (b) ECG-gated 3-D image for the dataset corrupted by breathing motion. (c) Result for the proposed algorithm for the dataset corrupted by respiratory motion.



(a) Standard (b) Proposed

Fig. 2: Reconstruction results for the clinical dataset of the coronary arteries.

B. Clinical Data

In Fig. 2 the reconstruction results are depicted. For the clinical data similar results are obtained as for the CAVAREV phantom study. The proposed algorithm outperforms the standard ECG-gating as it can successfully handle the non-periodic motion pattern of the dataset.

The runtime for the presented setups was in average 20 minutes on a NVIDIA Quadro FX5600 graphics card.

V. CONCLUSIONS AND OUTLOOK

A major clinical challenge of C-arm based cardiac vasculature reconstruction is non-periodic motion. In this paper a framework for the ECG-gated 3-D reconstruction with affine motion correction for non-cyclic motions has been introduced. It is shown in a simulation study and on clinical datasets that the method is able to perfectly reconstruct datasets where classical approaches fail.

Disclaimer: The concepts and information presented in this paper are based on research and are not commercially available.

REFERENCES

[1] J. Rieber, C. Rohkohl, G. Lauritsch, H. Rittger, and O. Meissner, "Application of C-arm computed tomography in cardiology (Kardiale Anwendung der C-Arm-Computertomographie)," *Der Radiologe*, vol. 49, no. 9, pp. 862–867, September 2009.

[2] H. Hetterich, T. Redel, G. Lauritsch, C. Rohkohl, and J. Rieber, "New X-ray imaging modalities and their integration with intravascular imaging and interventions," *The International Journal of Cardiovascular Imaging*, p. online (to be printed), 2009.

[3] N. Strobel, O. Meissner, J. Boese, T. Brunner, B. Heigl, M. Hoheisel, G. Lauritsch, M. Nagel, M. Pfister, E.-P. Rührnschopf, B. Scholz, B. Schreiber, M. Spahn, M. Zellerhoff, and K. Klingenbeck-Regn, *Multislice CT*, 3rd ed. Springer Berlin Heidelberg, 2009, ch. 3D Imaging with Flat-Detector C-Arm Systems, pp. 33–51.

[4] B. Desjardins and E. Kazerononi, "ECG-Gated Cardiac CT," *American Journal of Roentgenology*, vol. 182, no. 4, pp. 993–1010, April 2004.

[5] D. Schäfer, J. Borgert, V. Rasche, and M. Grass, "Motion-compensated and gated cone beam filtered back-projection for 3-d rotational x-ray angiography." *IEEE Transactions on Medical Imaging*, vol. 25, no. 7, pp. 898–906, July 2006.

[6] G. Lauritsch, J. Boese, L. Wigström, H. Kemeth, and R. Fahrig, "Towards cardiac c-arm computed tomography." *IEEE Transactions on Medical Imaging*, vol. 25, no. 7, pp. 922–934, July 2006.

[7] E. Hansis, D. Schäfer, M. Grass, and O. Dössel, "An iterative method for the reconstruction of the coronary arteries from rotational x-ray angiography," in *Medical Imaging 2007: Physics of Medical Imaging. Proceedings of the SPIE.*, J. Hsieh and M. J. Flynn, Eds., vol. 6510, no. 1, San Diego, CA, USA, 2007, p. 651026.

[8] E. Hansis, D. Schafer, O. Dossel, and M. Grass, "Evaluation of iterative sparse object reconstruction from few projections for 3-d rotational coronary angiography," *IEEE Transactions on Medical Imaging*, vol. 27, no. 11, pp. 1548–1555, November 2008.

[9] C. Blondel, G. Malandain, R. Vaillant, and N. Ayache, "Reconstruction of coronary arteries from a single rotational x-ray projection sequence," *IEEE Transactions on Medical Imaging*, vol. 25, no. 5, pp. 653–663, May 2006.

[10] C. Rohkohl, G. Lauritsch, A. Nöttling, M. Prümmer, and J. Hornegger, "C-arm ct: Reconstruction of dynamic high contrast objects applied to the coronary sinus," in *IEEE NSS-MIC (Nuclear Science Symposium-Medical Imaging Conference)*, Dresden, Germany, October 2008, pp. M10–328.

[11] E. Hansis, H. Schomberg, K. Erhard, O. Dössel, and M. Grass, "Four-dimensional cardiac reconstruction from rotational x-ray sequences: first results for 4d coronary angiography," E. Samei and J. Hsieh, Eds., vol. 7258, no. 1. Lake Buena Vista, FL, USA: SPIE, March 2009, p. 72580B.

[12] E. Hansis, D. Schäfer, O. Dössel, and M. Grass, "Projection-based motion compensation for gated coronary artery reconstruction from rotational x-ray angiograms," *Physics in Medicine and Biology*, vol. 53, no. 14, pp. 3807–3820, July 2008.

[13] C. Rohkohl, G. Lauritsch, M. Prümmer, and J. Hornegger, "Interventional 4-d motion estimation and reconstruction of cardiac vasculature without motion periodicity assumption," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009*, ser. Lecture Notes in Computer Science, G.-Z. Yang, D. Hawkes, D. Rueckert, A. Noble, and C. Taylor, Eds., vol. 5761. Springer, 2009, pp. 132–139.

[14] G. Shechter, C. Ozturk, J. R. Resar, and E. R. McVeigh, "Respiratory motion of the heart from free breathing coronary angiograms," *IEEE Transactions on Medical Imaging*, vol. 23, no. 8, pp. 1046–1056, 2004.

[15] M. Unser, "Splines: A perfect fit for signal and image processing," *IEEE Signal Processing Magazine*, vol. 16, no. 6, pp. 22–38, November 1999, IEEE Signal Processing Society's 2000 magazine award.

[16] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, ISBN: 0521540518, 2004.

[17] H. Scherl, B. Keck, M. Kowarschik, and J. Hornegger, "Fast GPU-Based CT Reconstruction using the Common Unified Device Architecture (CUDA)," in *Nuclear Science Symposium, Medical Imaging Conference 2007*, E. C. Frey, Ed., 2007, pp. 4464–4466.

[18] C. Rohkohl, B. Keck, H. Hofmann, and J. Hornegger, "RabbitCT - an open platform for benchmarking 3D cone-beam reconstruction algorithms," *Medical Physics*, vol. 36, no. 9, pp. 3940–3944, 2009.

[19] NVIDIA, "Cuda programming guide 2.2," 2009.

[20] C. Rohkohl, "Cardiac vasculature reconstruction evaluation platform," 2009, http://www.cavarev.com.

Development of a 4-D Digital Beating Heart Phantom with a Detailed Coronary Arterial Tree for C-arm and CT Angiography Simulation

George S.K. Fung, W. Paul Segars, Katsuyuki Taguchi, Alexander I. Veress, Grant T. Gullberg, *Fellow, IEEE*, and Benjamin M.W. Tsui, *Fellow IEEE*

***Abstract* – With the continuous resolution improvement in C-arm and CT angiography, increasingly smaller blood vessels are being imaged. To study the advancement in hardware, algorithms, and methodologies, a 4-D digital heart phantom with fine details of coronary arterial tree is needed. In this paper, we developed a new 4-D digital beating heart phantom by combining a detailed computer generated coronary arterial tree with the anatomic and motion realistic 4-D XCAT heart phantom. To generate the coronary arterial tree, the proximal branches of the coronary arteries were first defined based on a set of gated CT images of a normal heart. Then, an iterative rule-based algorithm based on anatomic, physiologic, and morphometric properties of the normal coronary arterial tree was used to extend the vasculature to cover the whole heart. The resulting vasculature was organized in a tree structure of arterial segments which had individual end points and radii. To superimpose the coronary arterial tree on the 4-D XCAT beating heart phantom, the locations of the end points of these arterial segments were spatially interpolated or extrapolated from the realistic motion vectors of the myocardium region of XCAT according to the time point of the specific frame. To apply the new 4-D digital heart phantom to x-ray imaging, an x-ray projection simulator, which allowed flexible x-ray tube and detector settings, were employed to generate realistic projection data of the phantom. In this study, we simulated several C-arm coronary fluoroscopies of a contrast-enhanced coronary angiography of LAD, LCX, and RCA in RAO and LAO projection positions, and rotational angiography. For each simulated fluoroscopy, thirty frames of projection images of one heart cycle were generated using flat panel detector, which were assessed and concluded that the images compared well with data from typical patient studies. In conclusion, the new 4-D digital beating heart phantom with a superimposed detailed coronary arterial tree is a unique tool allowing the generation of realistic x-ray projection data that mimic those from C-arm and CT angiography. It provides the exact status of anatomy and motion in simulation study to evaluate different scanning protocols and reconstruction algorithms.**

I. INTRODUCTION

To evaluate and optimize the image reconstruction and scanning protocols of medical imaging systems, computer simulation has become one of the essential components to perform medical imaging research in both academic and industry. In CT imaging, the first sub-process of this computer simulation is the generation of projection data, which involves of applying a imaging process of certain geometry and physics settings on a mathematical or voxelized phantom. The main advantage of using digital phantoms in computer simulation studies is that the anatomy and physiological information of the phantom are exactly known, thus providing essential data to perform accurate and quantitative evaluation on medical imaging systems and image reconstruction algorithms. To have direct impact to produce significant practical value to actual clinical situation, it requires the phantom to be anatomically and physiologically realistic enough in comparing with the targeted patient and clinical conditions.

With the continuous resolution improvement in C-arm and CT angiography, increasingly smaller blood vessels are being imaged. To study the advancement in hardware, algorithms, and methodologies through computer simulation studies, a 4-D digital heart phantom with fine details of coronary arterial tree is needed. In this paper, we developed a highly detailed coronary arterial tree and incorporate it into the beating heart model of the 4-D XCAT phantom [1] making it more applicable to C-arm and CT angiography imaging research, an area where the phantom can find great use to optimize and improve imaging techniques.

Manuscript received February 1, 2010. This work was partly supported by the National Institute of Health under Grant No. R01-EB00168 and R01-EB00121.

G. S. K. Fung, K. Taguchi, and B. M. W. Tsui are with the Department of Radiology, Johns Hopkins University, Baltimore, MD 21287 USA (e-mail: gfung2@jhmi.edu, ktaguchi@jhmi.edu, btsui1@jhmi.edu).

W. P. Segars is with the Department of Radiology, Duke University, Durham, NC 27710, USA (e-mail: paul.segars@duke.edu).

A. I. Veress is with the Department of Mechanical Engineering, University of Washington, Seattle WA 98192 USA (e-mail: averess@u.washington.edu).

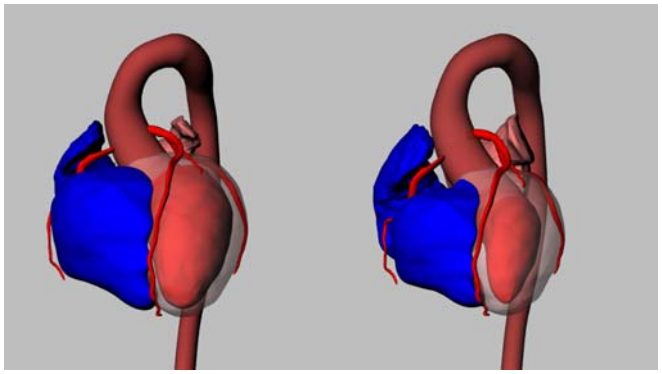
G. T. Gullberg is with E.O. Lawrence Berkeley National Laboratory, Berkeley, CA 94720 USA (e-mail: GTGullberg@lbl.gov).

II. METHODS

A. 4-D XCAT Heart Phantom

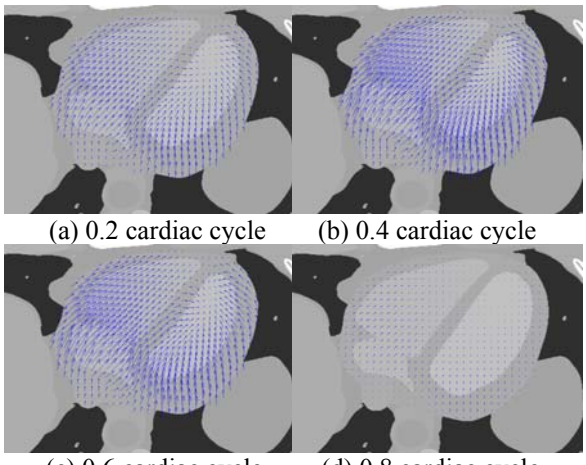
The 4-D extended cardiac-torso (XCAT) phantom, a computer model of human anatomy and physiology that is both realistic and flexible, was previously developed in our laboratory [1] and has been under further improvement and extension in various aspects. In XCAT, non-uniform rational b-splines (NURBS) surfaces were used to construct the high level of anatomical details of the organ shapes. The XCAT was extended to four dimensions to model the cardiac and respiratory motions using 4-D cardiac-gated tagged MRI and multi-slice CT data and 4-D respiratory gated CT data respectively [2].

Retrospective ECG-gated high-resolution dual-source CT [3] coronary angiographic image data of a contrast enhanced normal human heart was obtained from Siemens Healthcare. The end-diastole and end-systole phase of the DSCT data were used to define the anatomical details of the cardiac layers and the proximal branches of the coronary arterial tree as shown in Fig. 1. The motion vector fields of different phases in respect to end-diastolic heart were calculated from previous tagged MRI data and were depicted as in Fig. 2.



(a) End-diastole (b) End-systole

Fig. 1. The XCAT heart phantom at end-diastole and end-systole phases



(a) 0.2 cardiac cycle (b) 0.4 cardiac cycle

(c) 0.6 cardiac cycle (d) 0.8 cardiac cycle

Fig. 2. The dense motion fields (in a transaxial plane) of the XCAT heart phantom at certain cardiac cycles after the end-diastolic phase

B. Computer Generated Coronary Arterial Tree

B.1. Morphometry of Coronary Arterial Tree

In Kassab's statistical morphometric data [4], a diameter-defined Strahler ordering scheme was employed to systematically define the orders or generations of the coronary artery segments. The arterial segments were categorized into non-overlapping diameter ranges, from the largest arteries, order 11, to the smallest precapillary arterioles, order 1. The diameters, lengths, and connectivity probabilities were defined for each vessel order of each main artery. Since the morphometric measurements were based on porcine hearts, the diameters and the lengths of vessel segment were required to be linearly and quadratically (power of 2) scaled respectively to nominal human coronary arteries measurements according to scaling laws.

B.2. Fluid Dynamic Bifurcation Constraints

The relationship between the flow rate through and diameter of a branch segment was defined as $Q = Cd^n$, where Q is the flow rate, C is a constant that depends on the organ and the fluid, d is the diameter, and n, within the range 2 to 3, is a constant called the diameter exponent. Using this fundamental equation, the optimal branching angles of the daughter branches were then determined by the assumption of the minimization of shear stress for bifurcation as derived by Zamir [5].

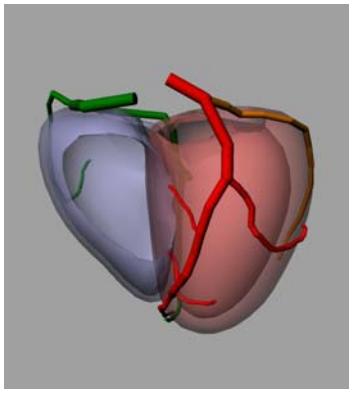
B.3.Rule-based Generation Algorithm

Given the morphometry of coronary arterial tree and fluid dynamic constraint for bifurcations, an iterative generation algorithm, which included self avoidance and boundary avoidance algorithms, was used to set of rules to guide the growth directions of the daughter branches at bifurcations. The details could be found in [6]. The rules were summarized as follows:

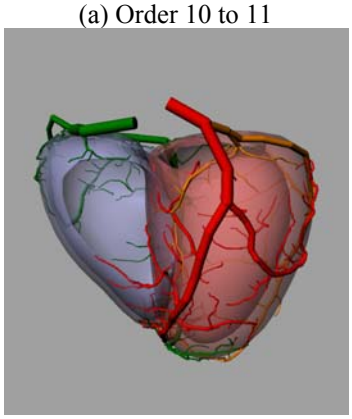
1) two daughter branches are produced for each parent branch;
2) the daughter branches lie in the branching plane;
3) the flow rate of the parent branch is conserved after branching;
4) the diameters of the daughter branches are statistically determined by the connectivity probability and nominal diameters of the morphometric data;
5) the flow-dividing ratio and branching angles are determined by the diameters of the daughter branches;
6) the branch length of a given order is determined statistically by "scaled" nominal length of the morphometric data;
7) the larger branches (order 9 or larger) are confined to the epicardium layer and the smaller branches are allowed to penetrate into the myocardium layer;
8) the combined branching vector is determined by the self avoidance and boundary avoidance algorithms; and
9) the normal vector of the branching plane is calculated from the parent branch vector and the combined branching vector.

The generation algorithm successively generated two daughter segments at each bifurcation as outlined above and iterated until the segments of largest six orders were constructed. A coronary arterial tree model, including the LAD, LCX and RCA, was generated. The generated coronary

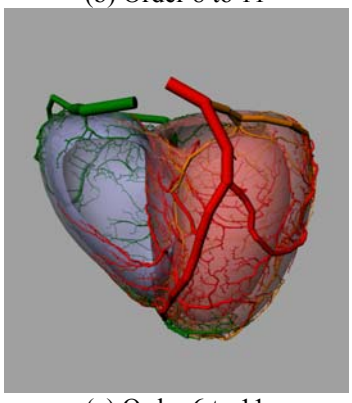
arterial tree including down to order 10, order 8, and order 6 branches were displayed through the transparent ventricular wall in Fig. 3 (a), (b) and (c) respectively. The six largest order model of the coronary arterial tree, which had diameters down to 120μm, consisted of 3940, 1232, and 3902 segments for LAD, LCX, and RCA respectively.



(a) Order 10 to 11



(b) Order 8 to 11



(c) Order 6 to 11

Fig. 3. The computer generated coronary arterial at different levels of detail

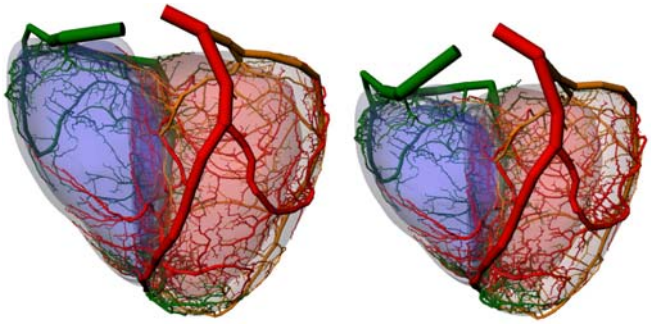
C. X-ray Projection Simulator

An x-ray projection simulator, called CT project, was previously developed in our lab. It calculated attenuation ray sum directly from NURBS surfaces of XCAT phantom, which has the advantage to avoid alias artifact caused by voxelized phantom, especially in high spatial resolution projection data generation. The details of the implementation could be found in [1]. The simulator had the flexibility to set the x-ray source position and energy spectrum, the detector position and geometrical shape, and the geometry and speed of the trajectory.

III. Results and Discussions

We generated a new digital phantom with the computer generated coronary arterial tree of over 9000 arterial segments and superimposed it on the XCAT cardiac phantom at end-diastolic phase. By using the motion vector fields of XCAT phantom at different cardiac phase, the coronary arteries and the heart could be deformed to any cardiac phase by spatial and temporal interpolation. The digital phantoms at end-diastolic and end-systolic phases were shown in Fig 4(a) and (b) respectively.

We simulated three C-arm coronary fluoroscopies of a contrast-enhanced coronary angiography of LAD, LCX, and RCA in RAO and LAO projection positions, and rotational angiography. The tree was filled with the mixture of iodine contrast and blood. For each simulated fluoroscopy, thirty frames of projection images of one heart cycle were generated using flat panel detector geometry. The image sequences of the RAO, LAO and rotational angiography were depicted in Fig 5,
Fig 6, and Fig 7 respectively. For rotational angiography scan, we set the timing to one rotation per heart cycle. Attached gif animated



(a) End-diastole (b) End-systole

Fig. 4. The new digital beating heart phantoms with a detailed coronary arterial tree at end-diastolic and end-systole phases

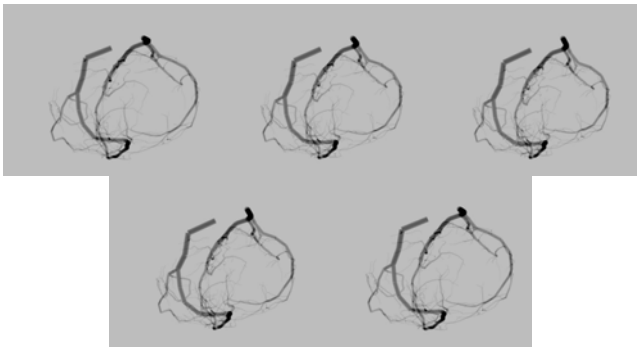


Fig. 5. The image sequence of the simulated fluoroscopy in LAO projection position

[Download animated gif at http://dmip1.rad.jhmi.edu/~gskfung/CAT_LAO.gif]

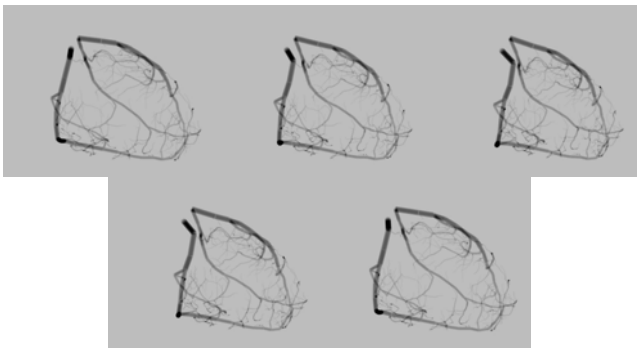


Fig. 6. The image sequence of the simulated fluoroscopy in RAO projection position
[Download animated gif at http://dmip1.rad.jhmi.edu/~gskfung/CAT_RAO.gif]

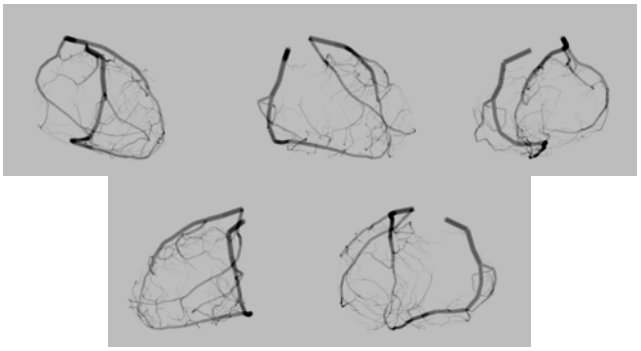
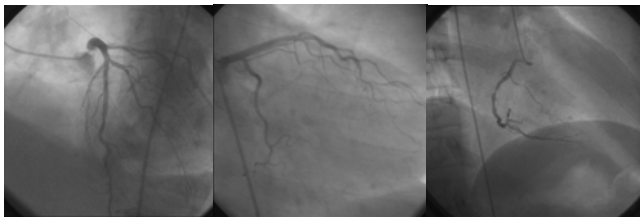


Fig. 7. The image sequence of the simulated fluoroscopy in rotational angiography
[Download animated gif at http://dmip1.rad.jhmi.edu/~gskfung/CAT_RA.gif]



(a) LCA in LAO (b) LCA in RAO (c) RCA in LAO
Fig. 8. Some typical samples of coronary C-arm angiography

The simulated flouroscopies were qualitatively assessed by two cardiac imaging experts and one C-arm angiography expert. In both RAO and LAO projections, all three experts agreed the right-dominant anatomy and the non-right deformation motion of the coronary arterial tree were realistic in comparing to typical coronary C-arm angiography. The orientation of the arteries and the projection overlap between the arterial branches could be clearly interpreted from the angiograms.

The experts had a number of comments to further improve the realistic of the simulation: (1) there were no gaps at bifurcations; (2) the curvature of the arteries should be continuous; (3) surrounding background organs should be included; and (4) image noise at practical situations should be added. To address the above issues: for (1) and (2), the spline curves, which pass through the bifurcation points and have variable diameters, will be used to model the anatomy of coronary arterial tree; for (3), the surrounding organs, such as spine, rib, lung, and liver, of XCAT phantom will be added; and for (4), the image noise can be added based on the mAs per projection.

IV. Conclusion

In conclusion, the new 4-D digital beating heart phantom with a superimposed detailed coronary arterial tree is a unique tool allowing the generation of realistic x-ray projection data that mimic those from C-arm and CT angiography. By applying image reconstruction algorithms to the generated projection data, in the form of rotational angiography or sonogram, we will perform further studies on CT angiography using the new phantom.

References

[1] W. P. Segars, M. Mahesh, T. J. Beck, E. C. Frey, and B. M. W. Tsui, "Realistic CT Simulation using the 4D XCAT Phantom," *Medical Physics*, vol. 35, no. 8, pp 3800-3808, Aug. 2008.

[2] W. P. Segars, "Development of Application of the New Dynamic NURBS-based cardiac-torso (NCAT) phantom," *Ph.D. thesis*, Biomedical Engineering, University of North Carolina, Chapel Hill, NC, 2001.

[3] T.G. Flohr, et al., "First Performance Evaluation of a Dual-Source CT (DSCT) System," *European Radiology*, vol. 16, pp. 256-268, 2006.

[4] G. S. Kassab, C. A. Rider, N. J. Tang, and Y. C. B. Fung, "Morphometry of Pig Coronary Arterial Tree," American *Journal of Physiology*, vol. 265, pp. H350-H365, 1993.

[5] M. Zamir, "Optimality Principles in Arterial Branching," *Journal of Theoretical Biology*, vol. 62, pp. 227-251, 1976.

[6] G. S. K. Fung, W. P. Segars, A. I. Veress, G. T. Gullberg, and B.M. W. Tsui, "Toward Modeling of Regional Myocardial Ischemia and Infraction: Generation of Realistic Coronary Arterial Tree for the Heart Model of the XCAT Phantom," *SPIE Proceedings of Medical Imaging*, vol. 7262, paper no. 7262-15, 6 pages, Feb 2009.

Dynamic CT Technique for Assessment of Musculoskeletal Joint Instabilities

Shuai Leng, PhD*, Kristin Zhao, MS**, Mingliang Qu, MD*, Kai-Nan An, PhD**, Cynthia McCollough, PhD*
*Department of Radiology, **Biomechanics Laboratory, Division of Orthopedic Research, Mayo Clinic
200 1st St. SW
Rochester, MN 55905 USA

INTRODUCTION

Clinical techniques to diagnose abnormal joint motion are highly subjective and have a high false positive rate [1]. Additionally, routine radiographic examinations are not capable of demonstrating dynamic instabilities [2]. This is a significant clinical problem, since individuals suffering from joint instabilities experience progressive degenerative disease that results in functional disabilities or osteoarthritis (OA) [3]. However, if physicians are able to diagnose dynamic joint instabilities at an early stage, conservative or surgical intervention can restore normal function before the onset of arthritis or static deformities [4, 5].

Dynamic, three-dimensional (3D) measurements of joint kinematics have been shown to provide significant information regarding joint function [6, 7]. Biplanar fluoroscopy methods can accurately determine the arthrokinematics of articulations in real time for joints such as the hip, shoulder, and knee, where limited superposition of bone geometry occurs within the two-dimensional (2D) fluoroscopy images. However, radiographic superposition in joints such as the carpal bones of the wrist and the tarsal bones of the ankle precludes the application of this technique to these joints.

In this paper, we will present novel imaging methodology using 4-dimensional (4D; 3D + time) CT techniques to detect subtle joint instabilities that can only be observed during joint motion. Two separate dynamic CT imaging methods were investigated and image quality and radiation dose were assessed using cadaveric wrist joints. Our results demonstrate that high spatial resolution and high temporal resolution images could be obtained for qualitative and quantitative joint dynamics analysis and pathology diagnosis.

MATERIALS AND METHODS

A. MOTION SIMULUTOR AND CADAVERIC SPECIMEN SETUP

A custom motion simulator was fabricated to simulate radial-ulnar deviation at the wrist joint. We exposed the proximal ends of the radius and ulna bone in the cadaveric forearm and firmly mounted it to the device (Fig. 1). The hand was attached to an acrylic paddle via a single plastic screw through the 2nd intermetacarpal space, just proximal to the deep intermetacarpal ligament. Two linear slides under the paddle provided composite motions in the x and z-axes. A programmable stepper motor (connected to a laptop) produced belt-driven motion of the paddle in the x-axis, with free motion of the paddle in the z-axis, allowing the hand to perform periodic radioulnar motion through a maximum arc of 30° (10° of radial deviation and 20° of ulnar deviation). The wrist was programmed to move at 30 cpm [8, 9], representing a typical wrist motion speed [10].

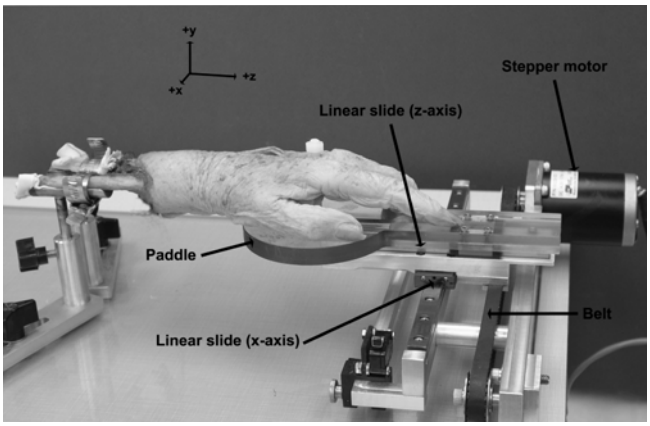


Fig. 1. Cadaveric wrist on a custom motion simulator.

B. DYNAMIC 4D CT IMAGING PROTOCOLS

Two separate dynamic 4D CT imaging methods were investigated to scan cadaveric wrist specimens: a retrospectively gated scan technique, and a non-gated sequential scan technique.

The retrospectively gated technique is similar to the spiral cardiac scans with a small helical pitch. Image slices corresponding to the same motion phase, while from different motion cycles, were retrospectively sorted together to form a 3D volume of the moving joint at a specific phase. 4D data sets were obtained by reconstructing 3D volumes at different motion phases. A programmable electronic trigger produced a periodic 10 ms voltage spike, which was synchronized to the motion profile of the stepper motor and used by the scanner's ECG monitor for reconstruction of the retrospectively-gated 4D CT [11]. We scanned the cadaveric wrist on a 64-slice CT scanner (Sensation 64, Siemens Healthcare, Forchheim, Germany). Images were reconstructed using a partial scan reconstruction algorithm with a temporal resolution of 165 ms. Each motion phase was evaluated for overall image quality and for the presence of motion artifacts by two independent, blinded raters – a hand surgeon and a radiologist.

In the retrospectively gated technique, periodicity was required for joint motion which may not be perfectly maintained in *in-vivo* patient studies. Therefore, we proposed a non-gated sequential scanning mode in which periodicity was not required. The cadaveric wrist was imaged using a Siemens Definition FLASH scanner (Siemens Healthcare, Forchheim, Germany). In this mode, there was no table translation between successive scans. Two seconds of data were acquired (one motion cycle). Scan parameters were 0.28 second gantry rotation, 2 x 64 x 0.6 mm detector collimation (38.4 mm z-axis coverage), 140 kVp, and 200 mAs per rotation. Twenty images were reconstructed over the 2 second cycle using the commercially implemented dual-source cardiac reconstruction algorithm. Reconstruction parameters were 75 ms temporal resolution, 150 mm scan FOV, 0.6 mm slice thickness for both medium sharp (B40) and sharp (B70) reconstruction kernels. 3D images at each of 20 time points in the motion cycle were generated using volume rendering techniques (VRT) on the scanner's image processing workstation.

C. Dose estimation and reduction

Skin dose was estimated for the non-gated sequential scans based upon CTDIvol obtained from the scanner console using the established relationship between skin dose and CTDIvol [12]. This estimated skin dose was compared with published thresholds for skin effects to determine whether deterministic effect is a concern. To follow the ALARA guideline, we investigated dose reduction for the 4D scanning techniques by scanning the cadaveric wrist at different dose levels. Image quality was evaluated at each dose level to determine the minimal dose at which the scans still had sufficient diagnostic image quality.

Results

A. Retrospectively gated technique

Excellent image quality was observed for motion phases 1 and 4 (Fig. 2) and fair image quality for motion phases 2 and 3 [13] due to band artifacts [14] and motion artifacts [15, 16]. The distal scaphoid had a total displacement of 12.4 mm from motion phase 1 to 4 (20^o ulnar deviation to 10^o radial deviation) during half of the motion cycle (1 sec), yielding an average motion velocity of 12.4 mm/sec for a typical radial-ulna wrist motion. Our previous results regarding the relationship between motion velocity and image quality [17], and the improved temporal resolution of the new Flash CT system, indicate that gated dual-source 4D CT can image objects moving with velocities up to 44 mm/sec. Therefore, 4D CT using a dual-source Flash CT system can freeze joint motion in dynamic wrist examinations. However, band artifacts occur in gated 4D CT scanning [14] when motion is not perfectly periodic or not ideally synchronized to the CT acquisition. The non-periodic motion leads to mismatch between scans from successive motion cycles, which corresponds to nearby slices in the longitudinal direction and presenting as band artifacts. Band artifacts can be partially corrected by retrospective editing of the "ECG"-signal, provided that a clear object contour is visible in the image to assist with realignment (Fig. 3). For in-vivo studies, patients with pathologies might have difficulty maintaining perfect periodic motion, which limits the wide application of this technique.

Fig. 2. Images from phase 1 and phase 4 of the gated 4D CT data set, acquired during radial-ulnar deviation of the wrist.

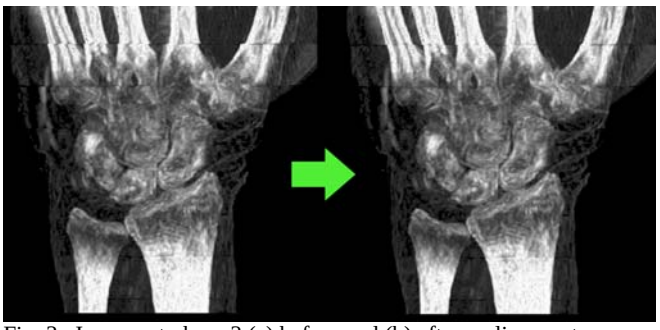


Fig. 3. Images at phase 3 (a) before and (b) after realignment.

B. Non-gated sequential technique

Carpal bones, distal radius and ulna, and joint spaces were clearly delineated in the VRT images, without motion blurring or banding artifacts, in all motion phases. 4D movies were generated from these 3D images to visualize the motion of each carpal bone and the change in joint spaces and contact areas throughout the motion cycle (movie file included in supplementary materials). Using the workstation's interactive visualization software, the carpal bones and joint motion can be visualized from any arbitrary view angle. Radiologists and clinicians can interactively select the best angle(s) with which to view any articulating structure. For example, a palmar view can be used to show overall wrist motion for all carpal bones (Fig. 4a-c), while radial views can be used to evaluate scaphoid motion (Fig. 4d-f).

Additionally, various surface opacities and color schemes can be applied, one of which generates a virtual radiograph at any arbitrary view from the CT data (Fig. 5). VRT images and movies from this cadaveric study were shown to an experienced hand surgeon for qualitative evaluation of image quality and joint motion. Image quality was considered to be excellent and joint motion and articulating surfaces were clearly visualized. These high spatial and temporal resolution 3D images will be used to perform quantitative assessment of the joint motion.

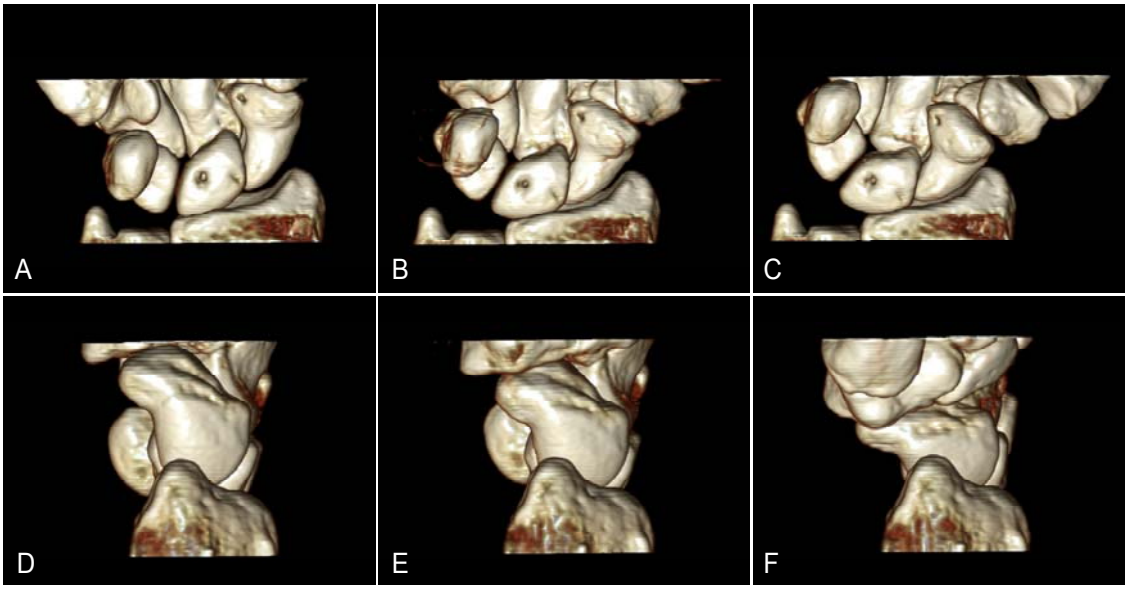


Fig 4. Volume rendered images (top row: palmar view; bottom row: radial view) of cadaver wrist scanned with a dynamic scanning mode on a dual source CT scanner. Images in ulnar deviation (A, D), neutral (B, E), and radial deviation (C, F) positions show individual carpal bones and joint spaces clearly in three dimensions.

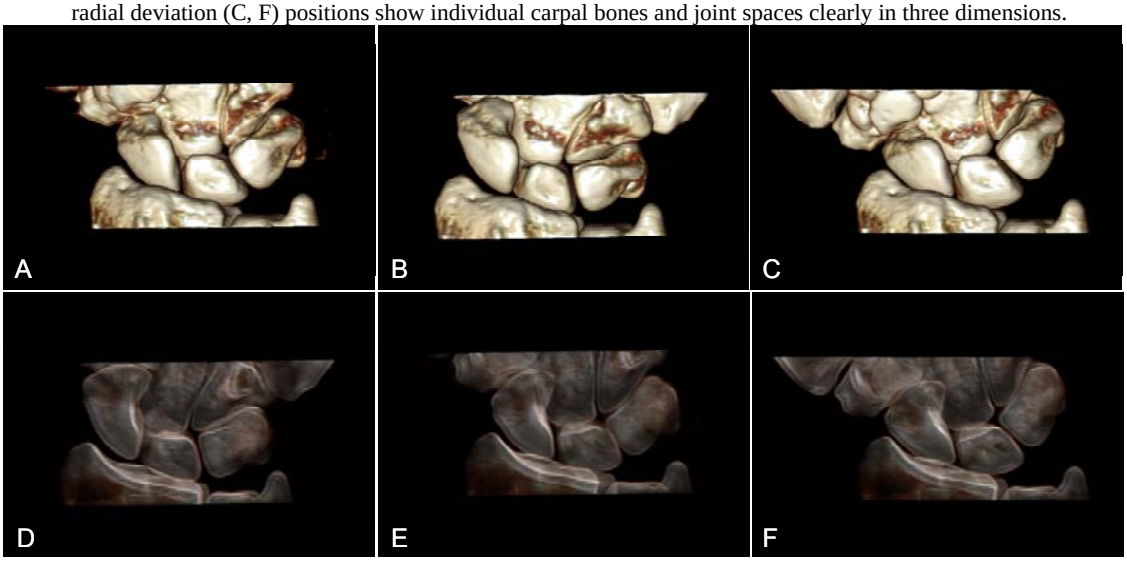


Fig 5. Volume rendered images (top row, dorsal view) and virtual radiography images (bottom row) of the cadaveric wrist at radial deviation (A, D), neutral (B, E), and ulnar deviation positions (C, F).

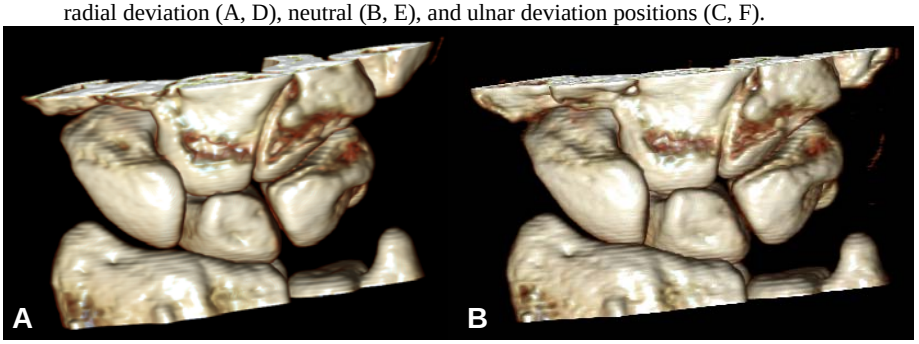


Fig 6. Volume rendered images of a cadaveric wrist acquired at skin doses of (A) 55 mGy and (B) 9 mGy. These data demonstrate the ability to dramatically decrease radiation dose without compromising the ability to clearly visualize the bones and joint spaces.

.

C. DOSE ESTIMATION

The estimated skin dose in this preliminary study, with scanning parameters of 140 kVp, 200 mAs per rotation, and a 2 second scan time, was about 55 mGy, which is a factor of 40 lower than the minimum threshold for skin effects (2 Gy) [18]. Thus, no deterministic skin injury is possible. Potential cancer risk is similarly negligible because of the small exposure volume and insensitivity of the exposed tissues to radiation. Potential dose reduction techniques were evaluated and found to enable a factor of 6 reduction in dose (Fig. 6)

DISCUSSIONS AND CONCLUSIONS

One of the greatest problems facing clinicians evaluating patients with suspected joint instability is the lack of a reliable diagnostic tool. While static instabilities can be diagnosed using routine imaging techniques, dynamic instabilities do not demonstrate abnormalities on routine radiographic examinations, even though these patients continue to have disabling pain. In this study, we have proposed two dynamic 4D CT imaging methods to detect joint motion and validated our methods using a cadaveric wrist and a custom motion simulator. The retrospectively gated method allows us to scan the whole joint with excellent image quality if the periodicity of joint motion is maintained. However, the method suffers banding artifacts in the cases where the joint motion is not perfectly periodic. The non-gated sequential technique, without the requirement of periodic motion, generates high spatial and high temporal resolution (75ms) images of the joint due to the two data acquisition systems on the dual source CT scanner. Radiation dose from this technique is much lower than the threshold for any potential deterministic effect and the stochastic effect is extremely low due to the absence of critical organs. One limitation of the non-gated technique is that the longitudinal coverage is limited by the detector size, which is close to 4 cm for this scanner. However, this is sufficient for imaging of small joints, including the wrist and carpometacarpal joint.

ACKNOWLEDGMENT

This work is partially supported by Mayo Novel Methodology DA Award 1 UL1 RR024150-01. The authors would like to thank Lawrence J. Berglund for his assistance in motion simulator fabrication.

REFERENCES

[1]. Wolfe SW, Crisco JJ. Mechanical evaluation of the scaphoid shift test. J Hand Surg Am 1994; 19:762-768.

[2]. Werner FW, Short WH, Green JK. Changes in patterns of scaphoid and lunate motion during functional arcs of wrist motion induced by ligament division. Journal of Hand Surgery - American Volume 2005; 30:1156-1160.

[3]. Watson HK, Weinzweig J, Zeppieri J. The natural progression of scaphoid instability. Hand Clin 1997; 13:39-49.

[4]. Adams BD, Berger RA. An anatomic reconstruction of the distal radioulnar ligaments for posttraumatic distal radioulnar joint instability. Journal of Hand Surgery - American Volume 2002; 27:243-251.

[5]. Walsh JJ, Berger RA, Cooney WP. Current status of scapholunate interosseous ligament injuries. Journal of the American Academy of Orthopaedic Surgeons 2002; 10:32-42.

[6]. Bey MJ, Zauel R, Brock SK, Tashman S. Validation of a new model-based tracking technique for measuring three-dimensional, in vivo glenohumeral joint kinematics. Journal of Biomechanical Engineering 2006; 128:604-609.

[7]. Wang S, Passias P, Li G, Li G, Wood K. Measurement of vertebral kinematics using noninvasive image matching method-validation and application. Spine 2008; 33:E355-361.

[8]. Flohr T, Stierstorfer K, Raupach R, Ulzheimer S, Bruder H. Performance evaluation of a 64-slice CT system with z-flying focal spot. Rofo 2004; 176:1803-1810.

[9]. Kyriakou Y, Kachelriebeta M, Knaup M, Krause JU, Kalender WA. Impact of the z-flying focal spot on resolution and artifact behavior for a 64-slice spiral CT scanner. Eur Radiol 2006; 16:1206-1215.

[10]. Marras W. Wrist motions in industry. Ergonomics 2003; 36:341-351.

[11]. Ohnesorge B, Flohr T, Becker C, et al. Cardiac imaging by means of electrocardiographically gated multisection spiral CT: Initial experience. Radiology 2000; 217:564-571.

[12]. Bauhs JA, Vrieze TJ, Primak AN, Bruesewitz MR, McCollough CH. CT dosimetry: Comparison of measurement techniques and devices. Radiographics 2008; 28:245-253.

[13]. Tay SC, Primak AN, Fletcher JG, et al. Four-dimensional computed tomographic imaging in the wrist: Proof of feasibility in a cadaveric model. Skeletal Radiol 2007; 36:1163-1169.

[14]. Taguchi K, Chiang BS, Hein IA. Direct cone-beam cardiac reconstruction algorithm with cardiac banding artifact correction. Med Phys 2006; 33:521-539.

[15]. McCollough CH, Bruesewitz MR, Daly TR, Zink FE. Motion artifacts in subsecond conventional CT and electron-beam CT: Pictorial demonstration of temporal resolution. Radiographics 2000; 20:1675-1681.

[16]. Mori S, Endo M, Tsunoo T, et al. Physical performance evaluation of a 256-slice CT-scanner for four-dimensional imaging. Med Phys 2004; 31:1348-1356.

[17]. Tay S, Primak A, Fletcher J, et al. Understanding the relationship of image quality and motion velocity in gated-CT imaging: Preliminary work for 4d musculoskeletal imaging. JCAT 2007; 32:634-639.

[18]. Stecker MS, Balter S, Towbin RB, et al. Guidelines for patient radiation dose management. J Vasc Interv Radiol 2009; 20:S263-273.

AUTHOR INDEX

Author Index