Design and Analysis of Low-cost X-ray Imaging System Incorporating Add-on Mouse Rotation System in Pre-clinical Research

by

Huanyi Zhou

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
August 6, 2022

Keywords: CT, System Calibration, Model-based Image Reconstruction, Deep Learning

Approved by

Stanley J. Reeves, Chair, Professor of Electrical and Computer Engineering
Thomas Denney, Professor of Electrical and Computer Engineering
Yanzhao Cao, Professor of Mathematics and Statistics Department
Peter P. Panizzi, Professor of Drug Discovery and Development, Harrison School of Pharmacy

Abstract

Computed tomography (CT) has become an essential tool in diagnostic and sports medicine. It is able to generate a three-dimensional rendering that can be used to interrogate a given region or desired structure from any orientation. However, in pre-clinical research, its deployment remains limited due to relatively high upfront costs. Besides, existing integrated imaging systems that provide X-ray modalities also dwarf its popularity in small laboratories. This lag of pre-clinical use of CT has resulted in small academic labs having to run pre-clinical CT scans on clinical veterinary scanners.

This dissertation seeks to generate CT images using an existing small-animal X-ray imager with a specialized add-on specimen rotation system, the *MiSpinner*. This setup conforms to the cone-beam computed tomography (CBCT) geometry, which requires high spatial accuracy. The key problem in retrofit studies is that the relative position of portable add-on devices varies by case. Therefore, in this research, we pursue three aims based on projection-based image processing methods and neural network methods: 1. System geometry calibration, 2. Low-dose projection denoising, and 3. sparse sampling enrichment.

System geometry calibration is achieved by our proposed structure tensor-based two-step online (ST-TSO) calibration method. It combines with a comprehensive edge detection technique to detect spatial misalignment information on projections. Specifically, the first processing step employs a modified projection matrix-based calibration algorithm to estimate the relevant geometric parameters, and the second processing step fine-tunes the parameters with an iterative strategy based on the symmetry property of the sum of projections. We showed that by using this calibration method we can successfully reconstruct the image, and its accuracy outperforms other online calibration method.

Low-dose projection image denoising is achieved with a model-based iterative reconstruction (MBIR) method. We consider the noise generation statistical model under a maximum a posterior probability (MAP) framework with a structure tensor based total directional variation

(ST-TDV) image prior regularization term. We demonstrated that the proposed method can effectively remove the noise in low-dose cases using a digital phantom study.

Sparse sampling introduces extra streak artifacts into reconstructed images that degrade the image quality. We propose an adversarial learning-based spatial transformer network for projection image synthesis, which aims to improve the quality of reconstructed images by reasonably increasing the number of projection images based on the existing projection data. Simulation and experiments have shown that this data-driven model can give rise to competitive results compared with conventional algorithms.

Additionally, due to the lack of electron density reference standards, we have tried an image style transfer network, cycle generative adversarial network (cycle-GAN), to improve the reconstructed image contrast. The deployed network that only requires two group of image datasets instead of image pairs has shown promise in retrofit studies, and our testing has confirmed its promising competitive performance.

Acknowledgments

I am deeply indebted to many years of phenomenal mentorship received at Auburn University. It is a remarkable experience to feel myself going from ignorance to being capable of independently finishing tasks. The faculty and staff here are all brilliant engineer-scientists as well as educators. It is nothing short of a privilege to work in such a friendly environment.

First of all, I would like to express my sincere gratitude to my supervisor, Prof. Stanley J. Reeves, for his unreserved guidance and support in my study and life. It is my luck and honor to have such a wonderful professor for leading me to many scenic views on this journey and for being patient and nurturing along the way. Likewise, I would like to express the same gratitude to my co-supervisor, Prof. Peter Panizzi, for his inclusiveness and full support throughout my Ph.D. study. Furthermore, I would like to thank Prof. Thomas Denny and Prof. Yanzhao Cao for serving on my thesis committee.

Moreover, I would also like to express my special thanks to other professors and professionals who have helped me deepen my understanding of the field: Prof. Chengying-Chou, for her systematic education on CT imaging and Neural Networks; Prof. Wenlu Wang, for her support and insight on Neural Network training; Prof. Dale Watson, for his introduction to professionalism in career development; Dr. Linxi Shi, for her offering the opportunity to connect with cutting-edge scientists in the world; Manager Tao Zhao, for his deep conversation with me to drag me out from being lost; Prof. Mark Nelms, for his department's financial support.

In addition, I would like to thank my friends for their company: Jueting Liu, for eight-year roommate company from Masters to Ph.D. and help in all computer science areas; Qianman Peng, for her help on career development; Jingjing Li, for her help on neural network learning; Chao Yang, for his help in paper writing; Shuwen Xue, for her help on Linear Algebra learning; Junwei Ma, Xiangyu Wang and Zhitao Yu, for their help in network training; Jiahui Xu and Xin Wei, for taking me to various activities. I have also benefitted greatly from the company

of Zerui Dong, Yibo Lyu, Dehua Li, Xingyu Yang, Xue Xia, as well as many friends I met at Auburn. Special thanks to Dr. Qiuyang Lin for keeping on pushing me to move forward.

At last, my utmost thanks to my parents and other family members for their full trust and support.

Contents

List of Figures

List of Tables

List of Abbreviations

ART	Algebraic Reconstruction Technique
BPD	basis pursuit denoising
CBCT	cone beam computed tomography
CBP	Convolution Back Projection
CCD	charged coupled device
CNN	convolutional neural network
COR	center of rotation
CS	Compressed Sensing
CT	Computed Tomography
DI	directional interpolation
DL	Deep Learning
DSD	distance from the source to the detector plane
DSO	distance from the source to the object center
DTL	deep transfer learning
EAS	Evolutionary Annealing Simplex
FBP	Filtered Back Projection

FDK Feldkamp, Davis and Kress algorithm

FISTA fast iterative shrinkage algorithm

FOPD first order primal-dual

FOV field of view

GAN generative adversarial network

HDTV higher degree total variation

HLCC Helgason-Ludwig consistency condition

HU Hounsfield Unit

IFT inverse Fourier transform

IR iterative reconstruction

IVIS In Vivo Imaging System

LASSO least absolute shrinkage and selection operator

MAE mean absolute error

MAP maximum a posterior probability

MBIR model based iterative reconstruction

ML-CSC multi-layer convolutional sparse coding

MRF Markov random field

NGM noise-generating-mechanism

NN neural network

PM projection matrix

POCS projection onto convex sets

PSNR peak signal-to-noise ratio

PST Projection-Slice Theorem

PWLS penalized weighted least-square

ReLU rectified linear unit

RIP restricted isometry property

ROF Rudin-Osher-Fatemi

ROI Region-of-Interest

SA simulated annealing

SIN sinogram inpainting network

SLSR-Net semi-supervised learned sinogram restoration network

SNR signal-noise-ratio

SOP sum of projections

SSIM structure similarity index measure

ST-TDV structure tensor based total directional variation

ST-TSO structure tensor-based two-step online

STP standard pressure and temperature

TDV total directional variation

TGV total generalized variation

TV Total Variation

VAE variational autoencoder

WLS weighted least squares

Chapter 1

Introduction to Small-Animal Imaging

The goal of this thesis is to develop new systems, models, and analysis methods ensuring that a retrofit imaging system with an add-on device works properly. The work focuses on applying image processing methods to yield high-quality CT-like reconstruction images. In this chapter, we review the fundamentals of tomographic imaging and the imaging system.

1.1 Background

For decades of development, Computed Tomography (CT) has become an essential tool in diagnostic and sports medicine. Its use has revolutionized the management of the progress of chronic diseases in the clinical field, such as demineralization during osteoporosis [1], joint degradation during osteoarthritis [2], and hemorrhage detection during stroke [3]. In pre-clinical research, its application in small-animal imaging can help to understand human molecular biology and pathophysiology using virus injected or genetically engineered mice with spontaneous diseases that closely mimic human disease [4].

CT imagers typically use a juxtaposed X-ray source and multiple detector arrays that maintain their relationship to each other during a scan to generate X-ray projections from various perspectives as they travel either circularly around or in a spiral path (a.k.a. helical CT) [5]. The development of higher-resolution imaging techniques, such as micro-CT [6], PET/CT [7, 8], and SPECT/CT [9] have enabled the study of molecular interactions in disease progression that assist in evaluating therapeutic efficacy. Furthermore, newer CT reconstruction algorithms and procedures have improved spatial resolution and eliminated errors caused by the collection methods and object scattering [10, 11].

However, pre-clinical use of CT has experienced a lag that is likely related to relatively high upfront instrument cost [12], which has sometimes forced academic labs to run pre-clinical CT scans on clinical veterinary scanners. Although recently developed micro-CT imaging systems have become more cost-effective and widely available [13], a price ranging from $100,000 to $1,000,000 remains a concern for many small laboratories [14]. Moreover, in lieu of dedicated CT imagers, pre-clinical labs settle for planar systems that allow X-rays to be combined with fluorescence, bioluminescence, or Cherenkov imaging. The availability of these systems dwarfs the population of CT imagers in the pre-clinical space [15]. Furthermore, there are inherent logistical considerations in conducting sequential imaging sessions with immune-compromised animals on shared equipment that has been used immediately prior to viral or bacterial pathogen work.

To address these limitations in the pre-clinical field and to provide avenues for researchers to collect meaningful endpoint 3D structural information, we sought to evaluate the use of a limited number of X-ray projections to generate CT-like data using a small-animal X-ray system.

1.2 Instrument

The imaging instrument used in this thesis consists of two components, a main imaging system and an add-on object rotation device, termed the *MiSpinner*, seen in Fig. 1.1.

The imaging system, the In Vivo Imaging System (IVIS) Lumina XRMS machine (PerkinElmer, Waltham, MA), is a high-sensitivity, in vivo imaging platform that provides the opportunity to view the inside of the body without harming it and to track cellular and genetic activity within a living organism through optical and X-ray modalities. A complete imaging system, Fig. 1.1 Left, consists of a pre-configured computer with Living Image software installed, an imaging chamber for taking pictures and an XGI-8 Gas Anesthesia Delivery System providing isoflurane to anesthetize small animals inside the chamber. Specifically, the main component, the imaging chamber, includes a fixed-position X-ray source at the bottom, a scintillation plate, and a charged coupled device (CCD) camera at the top. It provides three modalities for imaging: bioluminescence, fluorescence, and X-ray.

A movable stage provides a platform for the sample rotation system, which fits snugly under the scintillation plate. The scintillation screen is used to convert the X-rays into photons that can be collected by the super-cooled camera. The stage has four set increments that essentially allow some moderate magnification of a given region. X-ray projection images can be acquired at different contrast levels under a fixed 10×10 cm^2 field of vision.

The add-on object rotation device, the *MiSpinner*, is made up of a 3D printed stage, an actuated motor,a 50-mL conical mouse holder, a foam insert, a cap connector, and a controller, shown in Fig. 1.1(a). The specific assembly process, as Fig. 1.1(b) shows, is the motor attached to a customized 3D-printed stage, stabilizing the animal holder during the rotation process. Moreover, the actuated motor rotates in multi-angle selection mode, which offers at most 96 intervals for a full 360° scan, and the holder is connected to the central drive shaft of the motor with a 3D-printed cap with a central square hole that mounts the holder directly to the actuated motor. Initial trials had a square punch used to modify a 50-mL conical tube cap. Because this was not reproducible, we chose to 3D print a conical tube cap with a perfectly centered square hole. In summary, we used a FlashForge Creator Dual Extrusion 3D printer equipped with a polylactic acid filament reel (1.75 mm) to print the mounting cap. The printing was driven by the use of the Simplify3D software (version 4.1).

With the help of the *MiSpinner*, specimen scanning can be taken through a sequence of angles, seen in Fig. 1.1(c). The actuator motor rotates the mouse holder to form an iso-center projection image stack, which conforms to the geometry of projections obtained from cone beam computed tomography (CBCT) .

The movable stage can be raised to different preset levels, providing different magnifications within a given 10×10 cm^2 field of view, marked as A (268.37 mm from the CCD), B (179.44 mm), C (85.75 mm), and D (4.76 mm). There is also a Z mode that is available at the A level but using a zoom lens. This paper selects the B position as the appropriate place for data collection. Detailed instrument information is shown in Table 1.1.

Table 1.1: The scanning protocol of IVIS Lumina XRMS in X-ray mode

X-ray Source	36 kV, 100 mA
Source to the MiSpinner distance	32.1 cm
Source to the Plate distance	37.4 cm
Field of view	10×10 cm^2
Detector pixel size	$13\times13\mu$m
Detector resolution	1024×1024
Angular Sampling	1 projection per 3.75 degree
Angular coverage	360 degree
Exposure time	7.8 s per projection

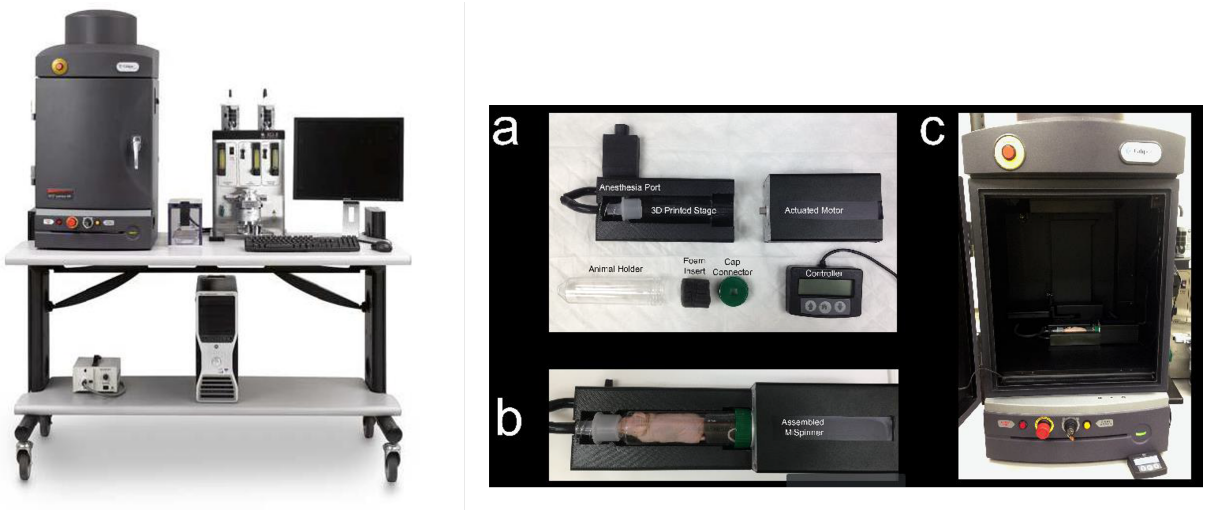


Figure 1.1: Instrument introduction of small-animal imaging. Left: Complete IVIS Lumina XRMS imaging system. Right: Display of the add-on object rotation device, the *MiSpinner*. (a) shows the detail of the *MiSpinner* for each component, (b) is a complete view of the device and (c) shows its placement in the imaging system.

1.3 CT Fundamentals

CT imaging aims to reconstruct cross-sectional images of the Region-of-Interest (ROI) from the data (projections) measured by the detector at different angle, which can help user observe inside information of the specimen without cutting. In this section, we introduce related fundamental physics in CT imaging to assist in implementation.

1.3.1 The Beer-Lambert Law

In transmission tomography, an external X-ray source sends a narrow X-ray beam passing through the object and detected by a 2D flat-panel detector. The object can be modelled by

its density map μ in R^2 (single slice reconstruction) or R^3(volume reconstruction). When an object is irradiated, X-rays with initial intensity I_0 will be absorbed by the object following the Beer-Lambert law [16]

$$I = I_0 e^{-\int \mu(l)dl}, \tag{1.1}$$

where I denotes the outgoing X-ray intensity recorded by the detector and l denotes the unit length of the photon traveling path. Since our interest is the distribution of density map μ, a characteristic feature of the object describing its attenuation coefficients, a negative log transformation will be applied to the collected data first no matter what scanning mode is used.

1.3.2 Sinogram: Forward Projection

Several basic integral equations model the variety of possible scanning geometries, e.g. Radon transform [17], which maps a density function to the set of its integrals over hyperplanes, called a sinogram in the CT field.

Definition 1.1 (The Radon Transform)**.** Let $\mu \in R^n$, the Radon transform is defined as

$$\Re_\mu(\vec{\theta}, s) = \int_{\vec{x}\cdot\vec{\theta}} \mu(\vec{x})d\vec{x} = \int_{R^n} \delta(\vec{x}\cdot\vec{\theta} - s)\mu(\vec{x})d\vec{x}, \vec{\theta} \in S^{n-1}, s \in R, \tag{1.2}$$

S^{n-1} refers to the set of unit vectors in R^n, which has dimension n-1.

Note that the variables of the Radon transform can identify one hyperplane where $\vec{\theta}$ is the normal unit vector and s is its distance to the origin. We can observe that in the 2D case, a hyperplane is a line and the Radon transform suits the physical reality of X-ray measurements, seen in Fig. 1.2(a). While in the 3D case, it is a plane but does not model X-ray acquisition. Therefore, we introduce the X-ray transform in the 3D case.

Definition 1.2 (The X-ray Transform)**.** Let $\mu \in R^3$, the X-ray transform is defined as

$$P_\mu(\vec{\theta}, \vec{y}) = \int_R \mu(\vec{y} + l\vec{\theta})dl, \vec{\theta} \in S^{n-1}, \vec{y} \in \vec{\theta}^\perp. \tag{1.3}$$

Note that the previous two transforms are parallel acquisition cases, which means all X-rays are parallel in one projection. For the case where the rays diverge from a point, we introduce the divergent-beam transform for X-ray emission from a point source.

Definition 1.3 (The Divergent-beam Transform)**.** Let $\mu \in R^3$, and $\vec{S_\lambda}$ denotes the 1D trajectory of an X-ray source. Then the divergent-beam transform is defined as

$$\mathfrak{D}_\mu(\lambda, \vec{\alpha}) = \int_0^\infty \mu(\vec{S_\lambda} + t\vec{\alpha})dt, \lambda \in [0, 2\pi], \vec{\alpha} \in S^{n-1}, \tag{1.4}$$

λ is in polar coordinates and indicates the source position, and $\vec{\alpha}$ is in spherical coordinate representing the X-ray path.

In this case, we assume the source is outside the ROI and the range of the divergent beam is wide enough to cover the region. Without loss of generality, we consider the circular trajectory centered on the origin of the ROI, shown in Fig. 1.2(b). In the 3D case, it refers to CBCT imaging. Generally, we use p to denote a projection from the parallel beam case and g to represent the fan beam case. Then, the relationship between parallel projection and fan beam projection is

$$g(\lambda, \alpha) = p(\lambda - \alpha, R\sin\alpha). \tag{1.5}$$

The process of beam re-binning maps divergent-beam data to parallel-beam data, which is crucial in fan beam reconstruction.

1.3.3 Image Reconstruction: Backward Projection

In this section, we introduce several CT image reconstruction techniques in the 2D case, including the projection-slice theorem, filtered back projection algorithm in parallel mode, re-binning in fan beam mode and revised volume reconstruction algorithm in the 3D case.

Recall that the Radon Transform is the same as the X-ray transform in the 2D case, noted as $p(R, \theta)$ and re-written in 2D Cartesian coordinates:

$$p(R, \theta) = \iint f(x, y)\delta(x\cos\theta + ysin\theta - R)dxdy \tag{1.6}$$

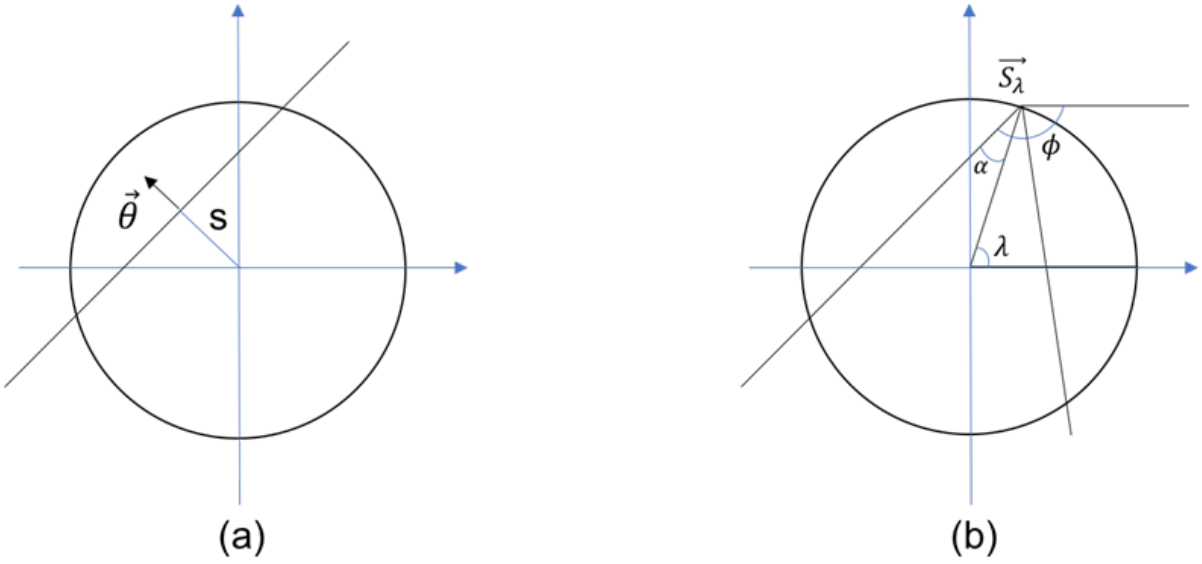


Figure 1.2: Sketch of forward projection: (a) Radon transform in the 2D case, (b) divergent-beam transform in the 2D case.

Projection-Slice Theorem (PST): The PST is proposed based on the principle that the 1D Fourier transform of a projection equals one line at an angle of the 2D Fourier transform of the density function. The proof is shown below:

$$\begin{aligned} G(\rho,\theta) &= \mathcal{F}[p(R,\theta)] = \int_{-\infty}^{\infty} p(R,\theta) e^{-2\pi j\rho R} dR \\ &= \iint f(x,y) e^{-2\pi j\rho(x\cos\theta + y\sin\theta - R)} dxdy \\ &= \mathcal{F}[f(x,y)]|_{u=\rho\cos\theta}^{v=\rho\sin\theta}. \end{aligned} \tag{1.7}$$

Thus, the basic implementation of PST is:

1. Take the 1D FFT of each projection.

2. Use the PST to arrange data in 2D Fourier space.

3. Interpolate those data onto a regularly-spaced grid.

4. Take the inverse FFT to obtain the density of attenuation coefficients.

The PST method has two main advantages. First, it reduces the computational cost compared with 2D Fourier transform calculation from k×M×N×(log(M×N)) to k×M×N×(log M). Second, Fourier transforms can take place during measurement. Theoretically the PST method could provide accurate results given a full 2D scan with an infinite number of slices. However, in the limited projection case, inaccuracy and high time cost interpolation between adjacent projections makes it less desirable for limited projections.

Filtered Back Projection (FBP): The FBP algorithm is an alternative reconstruction method in CT that aims to simplify Steps 2 and 3 in the PST method.
Given the 2D inverse Fourier transform (IFT) equation

$$f(x,y) = \iint \mathcal{F}(u,v) e^{2\pi j(\mu x + vy)} du dv, \tag{1.8}$$

and combining with the PST method, we have the following reconstruction

$$\begin{aligned} f(x,y) &= \iint G(\rho,\theta) e^{2\pi j(\mu x+vy)} du dv|_{u=\rho\cos\theta}^{v=\rho\sin\theta} \\ &= \int_0^{2\pi}\int_0^{\infty} G(\rho,\theta) e^{2\pi j\rho(x\cos\theta + y\sin\theta)} \rho d\rho d\theta \\ &= \int_0^{\pi} d\theta \int_{-\infty}^{\infty} G(\rho,\theta) e^{2\pi j\rho(x\cos\theta+y\sin\theta)} |\rho| \, d\rho. \end{aligned} \tag{1.9}$$

Rewriting the IFT form, we obtain

$$\begin{aligned} f(x,y) &= \int_0^{\pi} d\theta \int_{-\infty}^{\infty} G(\rho,\theta) e^{2\pi j\rho(x\cos\theta+y\sin\theta)} |\rho| \, d\rho \\ &= \int_0^{\pi} d\theta \int_{-\infty}^{\infty} [\int_{-\infty}^{\infty} G(\rho,\theta) |\rho| \, e^{2\pi j\rho R} d\rho] \delta(x\cos\theta + y\sin\theta - R) dR \\ &= \int_0^{\pi} d\theta \int_{-\infty}^{\infty} g(R,\theta)\delta(x\cos\theta + y\sin\theta - R) dR, \end{aligned} \tag{1.10}$$

where

$$g(R,\theta) = \int_{-\infty}^{\infty} G(\rho,\theta) |\rho| \, e^{2\pi j\rho R} d\rho. \tag{1.11}$$

The difference between the projections $g(R,\theta)$ and $p(R,\theta)$ is the former is the result of multiplying the original projection by an operator $|\rho|$, the ramp filter, in the frequency domain, as

a means of filtering out low-frequency components. After that, we can obtain accurate reconstruction by smearing the projection data back into the spatial domain.
The basic procedure is:

1. Take the 1D Fourier transform of projection data.
2. Multiply the operator in the frequency domain.
3. Take the inverse Fourier transform of the result of Step 2.
4. Smear back across the image.
5. Loop the procedure to all projection data.

Convolution Back Projection (CBP): The CBP method is a modification of the FBP algorithm in the spatial domain. It modifies Steps 2 and 3 in the FBP with convolution as follows:

$$\mathcal{F}^{-1}[\mathcal{F}\{g(R,\theta)\}\,|\rho|] = g(R,\theta) * \mathcal{F}^{-1}\{|\rho|\}, \tag{1.12}$$

where $*$ represents convolution. Let $C(R) = \mathcal{F}^{-1}\{|\rho|\}$, then

$$f(x,y) = \int_0^{\pi} d\theta \int_{-\infty}^{\infty} \{g(R,\theta) * C(R)\}\delta(x\cos\theta + y\sin\theta - R)dR. \tag{1.13}$$

The drawback of the FBP and CBP algorithms is that $|\rho|$ is non-integrable in continuous space. Therefore, it will produce non-negligible degradation problem in discretization. Classic processing methods add a window function on $|\rho|$ for approximation, either the R-L window, S-L window or the Hamming window. Here we will use R-L filter, or ramp filter, to maintain low computational cost for simplicity.

Beam Re-binning Strategy: In fan beam reconstruction, a common solution is to apply beam re-binning to convert the divergent beam projection data into corresponding parallel data based on the mapping relationship introduced before. The advantage is convenience, while the drawback is sparsity and uneven sampling in the corresponding parallel case simulation.

A general solution in the fan beam case can be derived from the CBP algorithm, and Fig. 1.3 shows an example. Assume detectors on a flat plane that are spatially distributed evenly, distance from the source to the object center (DSO) is D, and distance from the source to the detector plane (DSD) is $D^{'}$. For any incident X-ray, e.g., line SB in Fig. 1.3, assume its distance from the origin O to the projected point A is s. Then we have the relationship:

$$P(R, \theta) = q_{\beta}(s), \tag{1.14}$$

where

$$\theta = \alpha + \beta,\ R = s\ \cos\alpha = \frac{sD}{\sqrt{s^2 + D^2}}. \tag{1.15}$$

Rewriting the CBP algorithm in polar coordinates,

$$f(x, y) = f(r, \phi) = \frac{1}{2}\int_0^{2\pi}\int_{-R_m}^{R_m} P(R, \theta) \times C(r\cos(\theta - \phi) - R)dRd\theta, \tag{1.16}$$

then we have fan beam reconstruction

$$f(r, \phi) = \frac{1}{2}\int_0^{2\pi}\int_{-s_m}^{s_m} q_{\beta}(s) \times C[r\cos(\beta + \alpha - \phi) - \frac{sD}{\sqrt{s^2 + D^2}}]\frac{D^3}{\sqrt{s^2 + D^2}}dsd\beta, \tag{1.17}$$

where

$$\alpha = \arctan\frac{s}{D} \tag{1.18}$$

For simplicity, we can modify Eq. (1.17) as

$$f(r, \phi) = \int_0^{2\pi}\frac{1}{U^2}\{[q_{\beta}(s) \times \frac{D}{\sqrt{D^2 + s^2}}] * \frac{1}{2}C(s)\}d\beta \tag{1.19}$$

where

$$U(r, \phi, \beta) = \frac{D + r\sin(\beta - \phi)}{D}. \tag{1.20}$$

The equation is similar to the CBP algorithm. The difference is it pre-processes the projection data before with a weighted function related to its position and its numerical implementation:

1. Pre-process the projection data by multiplying by a weighted function.

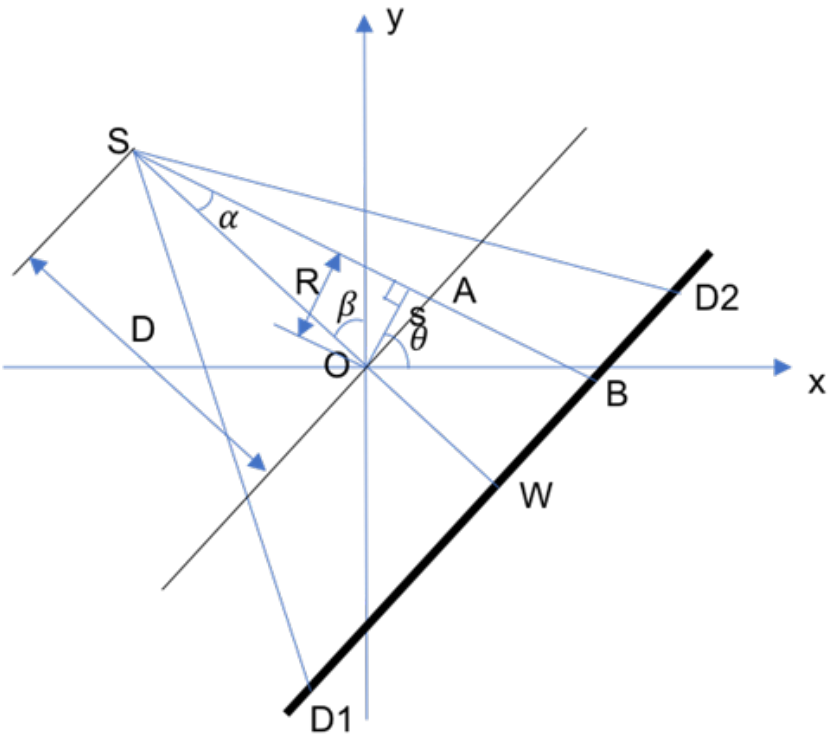


Figure 1.3: Sketch of fan beam geometry on a flat panel

2. Convolve with a modified filter function.

3. Add another weighted function and then backproject as FBP does.

FDK reconstruction: An analogue for volumetric reconstruction from 2D projections is the Feldkamp, Davis and Kress (FDK) algorithm [18, 19], which contains additional weights to account for fan and cone angles. It was first proposed by Feldkamp *et al.* in 1984, solving a cone-beam structure problem with a circular orbit of the X-ray source in 3D image reconstruction. It implements the reconstruction row by row with three steps:

1. Adding pre-weighting factor to the projection data: a pre-weighting factor, dependent on both the fan angle and cone angle, is applied to the projection data to approximate equivalent 2D fan beam data as

$$g_\theta^{'} = \frac{R_D}{\sqrt{R_D^2 + u^2 + v^2}} g_\theta(u, v), \tag{1.21}$$

where $g_\theta(u, v)$ is the 2D projection data of 3D object and u,v are coordinates in the detector plane. R_D is the distance from the X-ray source to the detector plane.

2. One-dimensional filtering: FDK applies a one-dimensional filter to process the weighted projection with convolution:

$$G_\theta(u, v) = h(u) * g_\theta^{'}(u, v), \tag{1.22}$$

where $h(u)$ is a ramp filter.

3. 3D reconstruction: after obtaining the filtered projection data, multiply by a weight function, and then reconstruct the 3D object volume slice by slice along the normal vector direction perpendicular to the fan beam plane.

$$\mu(x, y, z) = \int_0^{2\pi} \frac{R^2}{(R + x\cos\theta + y\sin\theta)^2} G_\theta(u', v')d\theta, \tag{1.23}$$

where

$$u' = \frac{R_D(-x\sin\theta + y\cos\theta)}{R + x\cos\theta + y\sin\theta},\ v' = \frac{R_D z}{R + x\cos\theta + y\sin\theta}, \tag{1.24}$$

$\mu(x, y, z)$ denotes the reconstructed 3D object attenuation coefficient and R is the distance between X-ray source and object center.

In addition, Parker weighting was introduced to account for short scans (210° arc) generated by extremity CBCT prototypes [20].

1.4 Thesis Outline

Chapter 2 describes a multi-level thresholding algorithm for system calibration inspired by global variance assessment criteria. A version of this chapter is published in:

- Zhou, H., Reeves, S.J. and Panizzi, P.R., Center of Rotation Correction for in vivo Computed Imaging System with Limited Projections. In 2020 IEEE 17th international symposium on biomedical imaging (ISBI 2020) . IEEE.

- Zhou, H., Reeves, S.J. and Panizzi, P.R., 2021, November. Estimating the Center of Rotation of Tomographic Imaging Systems with a Limited Number of Projections. In 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 3157-3160). IEEE.

Chapter 3 has further introduced a novel geometry based system calibration algorithm that utilizes scanning projection data directly. Furthermore, we have extended our research into more general case with extra constrained condition. The contents of this chapter can be found in:

- Zhou, H., Chou, C., Reeves, S.J., Brannen, A., and Panizzi, P.R. Online Geometry Calibration for Retrofit Computed Tomography from a Mouse Rotation System and a Small-Animal Imager. Medical physics

Chapter 4 shows how we improve the image quality after obtaining the correct reconstruction result. It includes two parts, model based denoising algorithm for projection pre-processing and machine learning based image synthesis algorithm for solving sparse sampling problem. This chapter includes:

- Zhou, H., Wang, W., Chou, C., Reeves, S.J. and Panizzi, P.R., Projection Image Synthesis Using Adversarial Learning Based Spatial Transformer Network for Sparse Angle Sampling CT. In 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE

Chapter 5 concludes our contribution to the retrofit system study and highlights the challenges associated with further improvement and reconstructing high-resolution results.

Chapter 2

Multi-level Thresholding Algorithm for 2D Online Calibration

X-ray imaging system with an add-on *MiSpinner* conforms to the CBCT geometry, which is sensitive to errors in the scanning geometry. Incorrect geometry will lead to severe image quality degradation problems, noted by image blurring and other structure artifacts. This chapter introduces the projection-based calibration algorithm to correct the geometry error. Test results have validated its accuracy in system correction, thus demonstrating the potential of the described system to be modified and adapted for pre-clinical research.

2.1 Introduction

A complete view of the retrofit imaging system is displayed in Fig. 2.1. We present our scanning mode simulation model which conforms to the CBCT imaging principle. In CBCT, the complete scanning geometry can be described by nine parameters. [21] In the special case of object-rotation small-animal imaging, the flat detector relative position to the X-ray source is always fixed during scanning. Therefore, the number of parameters is reduced to seven [22] since the detector coordinate-related parameters are aligned with the X-ray source translation coordinates; therefore, they can be removed from the estimation. However, not all parameters have the same impact on the reconstruction quality. Xi *et al.* [23] have pointed out that bias in the perpendicular direction to the rotation axis and in-plane rotation angle error can lead to more severe degradation than the other five parameters. More precisely, a 0.5-unit bias or a 0.5-degree misalignment is large enough to observe visible artifacts in the reconstruction images, while other parameters can tolerate a wider error range [22, 24].

In our imaging system, geometry errors are most likely to appear in the object rotation area, e.g., inaccurate placement of the *MiSpinner* or inclination of the animal holder. This can lead to object deformation in forward projections. Furthermore, the sample in the spinning system may not be placed precisely or consistently relative to the imaging coordinate system, the actual center of rotation (COR) is often uncertain. In addition, the sample may not consistently spin around an assumed axis because of variations in the attachment to the holder or due to gravity. Therefore, the specimen presented in the projection images may appear to be tilted or off-center. As a result, the reconstruction of such datasets may show ringing or streaking artifacts. Therefore, calibration of the geometry is essential before adequate image reconstruction can occur.

2.2 Related Work

Generally, geometry calibration approaches can be divided into two categories, offline methods and online methods. For offline methods, researchers use a precisely designed phantom with known spatial information to estimate the errors in geometric parameters and then calibrate the geometry either through adjustment of mechanical equipment or accounting for errors with reconstruction algorithms. For example, Cho *et al.* [25] used 24 steel beads as a phantom to get a set of projections for parameter estimation. Yang *et al.* [26] reduced the phantom to 12 steel beads on two planes in a cylinder phantom for error assessment. Sun *et al.* [27] combined prior relative position information with only 4 beads in a single projection to calculate the results. Ouadah *et al.* [28] proposed a 3D-2D registration algorithm that utilized a projection matrix (PM) to characterize system geometry. Parameters are optimized by comparing the similarity of scans and forward projections of previously acquired 3D volumes. In such methods, ample markers and related position locating algorithms (PM method) are crucial for accurate parameter estimation. This approach requires the design of a dedicated phantom that is difficult for other users to duplicate. Another challenge is the changing scanning geometry in pre-clinical research. Our *MiSpinner* device may not always be placed in the imaging system precisely and consistently. Hence, one-time offline calibration using a phantom may not be appropriate for continuously scanning in pre-clinical research.

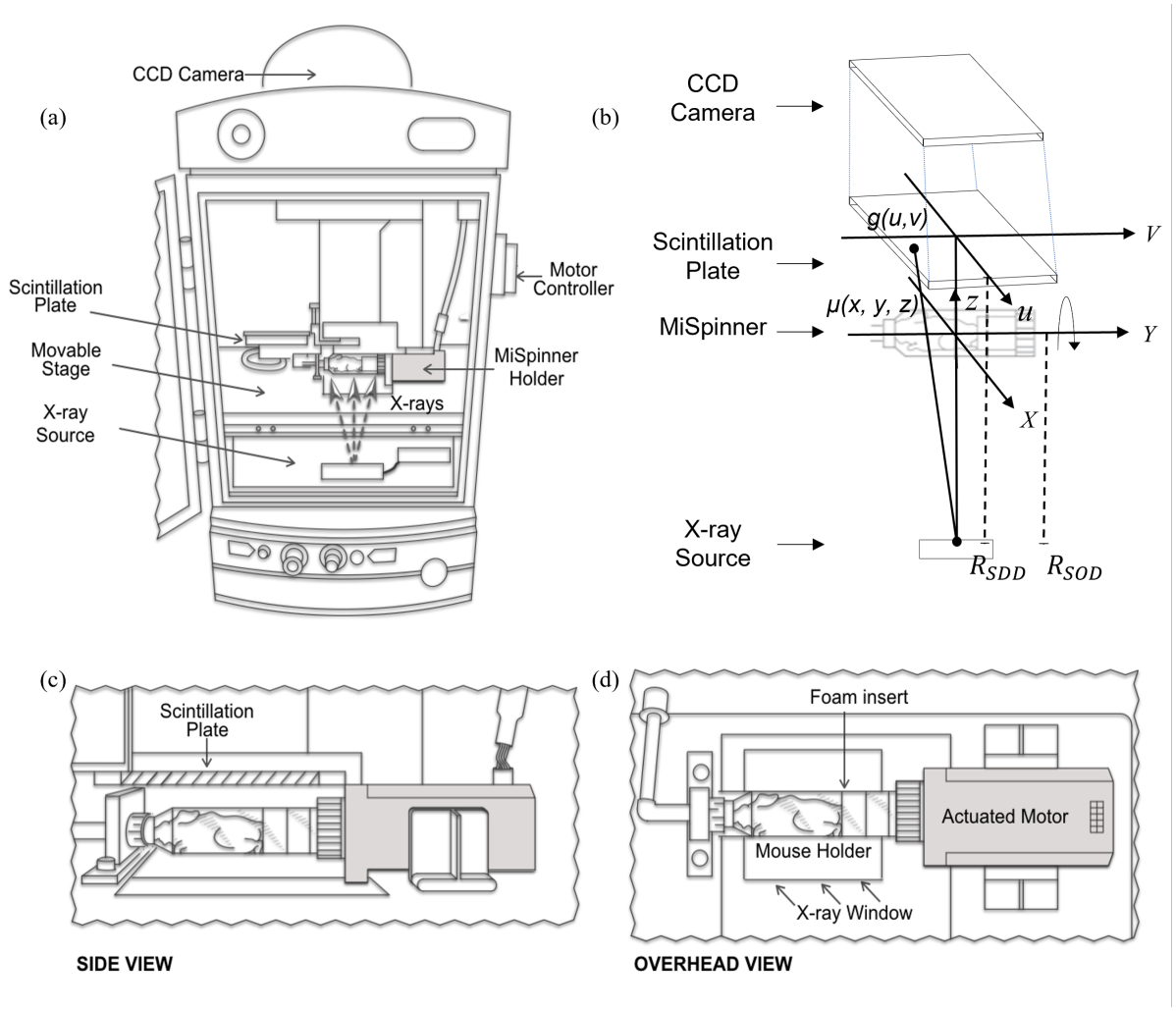


Figure 2.1: MiSpinner design and orientation in the pre-clinical scanner. (a) Imaging chamber of IVIS Lumina XRMS machine, fixed position of an X-ray source and scintillation plate, movable stage where for the *MiSpinner*. (b) 3D geometry of projection image taken in Lumina XRMS machine; the origin is defined at the center of the object, and (x,y,z) is the defined coordinate axis system. Object rotates around the x-axis. (c-d) The *MiSpinner* structure when placed in Lumina XRMS machine; ideally, the actuated motor can rotate mouse holder to form an iso-center projection image group.

In contrast, online calibration methods directly evaluate the projection image parameters without requiring any dedicated calibration phantoms. These methods fall into two classes, one that takes advantage of data redundancy in full scanning and the other that uses iterative optimization based on a designed cost function. For data redundancy, Donath *et al.* [29] proposed an image registration method, which estimates the maximum correlation on two opposite projections, for the COR correction, while Zhu *et al.* [30] utilized the centroid method, which is an application of the Helgason-Ludwig consistency condition (HLCC). Furthermore, Yu *et al.* [31] extended the HLCC into fan-beam cases to evaluate the angle errors. For iterative algorithms, Nghia *et al.* [32] developed a quality assessment algorithm based on the observation that key magnitude coefficients of the sinogram image Fourier transform lie in a particular region in the frequency domain. Dong *et al.* [33] proposed a global variance-based estimation matrix for COR calibration. Kyriakou *et al.* [34] optimized the system by minimizing the entropy. However, an obvious flaw of these algorithms is that they only focus on correcting one or two key geometric parameters. In a recent development, Meng *et al.* [22] developed a sum of projections (SOP) method, which could estimate five geometric parameters. However, execution time is an unavoidable problem. To make it converge faster, the authors limited the error bounds of the parameters in their simulation.

Given the challenge that our add-on device is not placed permanently, this research mainly focuses on solving the problem of online calibration. We first review existing work in detail.

2.2.1 Image Registration Method

The image registration method, also known as the opposite-angle method, is an analytical solution that aims to calculate COR bias. Fig. 2.2 shows an example of COR error in scanning. We can observe that in a full scan, a specific incident X-ray, the central ray, crosses the same region after 180 degree rotation. Thus it will give rise to the same projection value on the detector. In Fig. 2.2, this incident X-ray is the line SO that transmits from the source S to the rotation center O. Accordingly, the central ray can be determined by identifying the pair of angles that generate the same projections on the detector. Every angle has a corresponding pair with the relationship that $\theta_2 = \theta_1 + \pi$ and $\theta_1 \in [0, \pi)$.

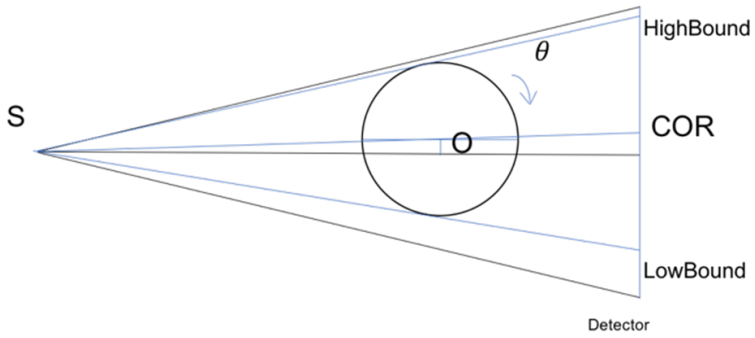


Figure 2.2: Illustration of image registration method in COR determination

Following this relationship, we can divide the full scan projections into two parts and then calculate the cross-correlation coefficient of projection data along the detector bins as Eq. (2.1) shows. If we find the maximum value of Q, we can determine the COR position.

$$Q(A,B) = \frac{\sum_i (A_i, B_i)}{\sqrt{(\sum_i A_i^2)(\sum_i B_i^2)}} \tag{2.1}$$

The drawback of this approach is it requires one or several π-pair-projections in measurements which will be violated in sparse sampling. Besides, this method is sensitive to the noise as well as low signal-noise-ratio (SNR) case.

2.2.2 Center-of-Mass method

The center-of-mass method, also known as the centroid method, is an application of HLCC [35]. Recalling the Definition 1.1, we denote the projection data as $p(R,\theta)$. Then HLCC is defined as

$$J_k(\vec{\theta}) = \int p(\theta, R) R^k dR = \sum_{r=0}^{k} \frac{k!}{r!(k-r)!} \cos^r(\theta) \sin^{k-r}(\theta), \tag{2.2}$$

where

$$m_{i,j} = \iint x^i y^j f(x,y) dx dy,\ k = i + j, \tag{2.3}$$

is called the geometry moment of the ROI region.

According to the centroid definition, for projection data at angle θ, its centroid calculation is

$$\vec{S_\theta} = \frac{\int p(R,\theta) R dR}{\int p(R,\theta) dR}. \tag{2.4}$$

Meanwhile, the center of mass of the ROI is

$$\bar{x} = \frac{\iint \mu(x,y) x dx dy}{\iint \mu(x,y) dx dy},\ \bar{y} = \frac{\iint \mu(x,y) y dx dy}{\iint \mu(x,y) dx dy}. \tag{2.5}$$

Based on a physical interpretation, we can draw the conclusion that the integral of the projection data equals the sum of the ROI region, that is,

$$\int p(R,\theta) dR = \iint \mu(x,y) dx dy. \tag{2.6}$$

Referring to Fig. 2.3, we define a new coordinates system, s, t coordinates, where the *s*-axis is along the detector bin direction and the *t*-axis is along view angle direction. Then we have

$$s = x\cos\beta + y\sin\beta,\ t = -x\sin\beta + y\cos\theta, \tag{2.7}$$

where β is the view angle in Fig. 2.3 which is the same as θ in other equations and s equals R in previous equations. Hence, we can infer that

$$\begin{aligned} \vec{S_\theta} &= \frac{\int p(s,\theta) s ds}{\int p(s,\theta) ds} = \frac{\iint \mu(x,y)(x\cos\beta + y\sin\beta) dx dy}{\iint \mu(x,y) dx dy} \\ &= \frac{\iint \mu(x,y) x\cos\beta dx dy}{\iint \mu(x,y) dx dy} + \frac{\iint \mu(x,y) y\sin\beta dx dy}{\iint \mu(x,y) dx dy} \\ &= \bar{x}\cos\beta + \bar{y}\sin\beta. \end{aligned} \tag{2.8}$$

It can be concluded that the centroid coordinate of a projection is the projection of the mass center of the ROI projected at that angle. Compared with the HLCC definition Eq. (2.2), Eq. (2.8) is a simplified version with $m = 1$.

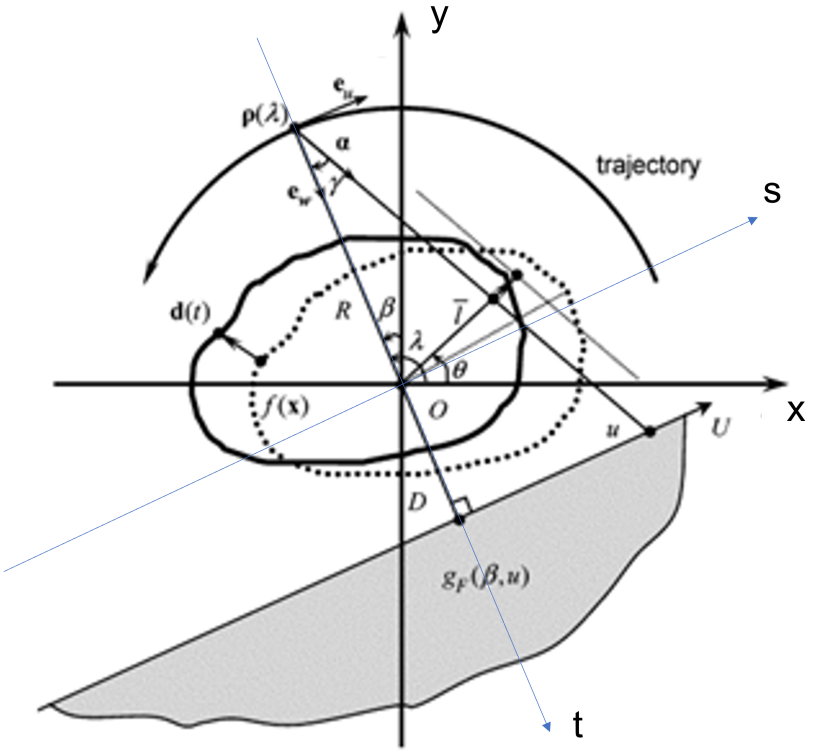


Figure 2.3: Illustration of Center-of-Mass method in COR determination

Next we extend the HLCC into the fan beam case following the beam re-binning strategy; that is,

$$J_k(\vec{\theta}) = \int_{-\infty}^{\infty} g_F(\theta - \arctan\frac{u}{D}, u)(\frac{Ru}{\sqrt{D^2+u^2}} - d_1(t)\cos\theta - d_2(t)\sin\theta)^k \times(\frac{RD^2}{D^2+u^2}^{3/2} + \frac{D(d_1(t)\cos\theta + d_2(t)\sin\theta)}{\omega(D^2+u^2)})du, \tag{2.9}$$

where $d(t) = (d_1(t), d_2(t))$ is COR translation in two directions that varies with time t, ω is gantry rotation speed. The derivation is detailed in the Appendix A. Estimation of $d(t)$ can be found in Yu *et al.* [31].

2.2.3 Frequency-Based Method

The frequency-based method utilizes the Fourier transform of the sinogram in parallel beam tomography. The sinogram is the projection of the ROI data in the 2D case. Given projection

data $p(R,\theta)$, its Fourier transform in the (R,θ) domain is

$$\tilde{P}(\omega,n)=\frac{1}{2\pi}\int_0^{2\pi}\int_{-\infty}^{\infty}p(R,\theta)dRd\theta. \tag{2.10}$$

Assume each point $\mu(x,y)$ in the ROI is independent. Then

$$\mu(x,y)=\delta(x-r_p\cos\phi,y-r_p\sin\phi), \tag{2.11}$$

where (r_p,ϕ) are the corresponding polar coordinates. $\tilde{P}(\omega,n)$ becomes

$$\begin{aligned} p(s,\theta)&=\delta(s-r_p\cos(\phi-\theta)),\\ \tilde{P}(\omega,n)&=\frac{1}{2\pi}\int_0^{2\pi}e^{-i[\omega r_p\cos(\phi-\theta)+n\theta]}d\theta. \end{aligned} \tag{2.12}$$

This is similar to the Bessel integral representation [36, 37],

$$\begin{aligned} \tilde{P}(\omega,n)&=J_n(\omega r_p)e^{-in(\phi+pi/2)},\\ J_n(x)&=\frac{1}{2\pi}\int_{-\pi}^{\pi}e^{i(x\sin\tau-n\tau)d\tau}. \end{aligned} \tag{2.13}$$

Refer to the Bessel function property introduced in [37]; i.e., for large $|n|$, the function $J_n(x)$ decreases rapidly in magnitude as $|\omega r_p|$ decreases below $|n|-1$. Accordingly, we have a double-wedge region at the boundary $n=\omega r_p$ and $n=-\omega r_p$ in the frequency domain where significant Fourier coefficients of the sinogram are found in the region

$$|n|\leqslant|\omega|\,r_p, \tag{2.14}$$

where ω is the spatial-frequency variable, n is the angular harmonic number, and r_p is the outmost radius of the ROI. A demonstration of this property is shown in Fig. 2.4.

A typical application of this property includes reconstruction image quality improvement by interpolating extra information into the sinogram at certain places [38], while another application is COR correction [32] which is based on the assumption that sinogram data in range $[0,\pi)$ and $[\pi,2\pi)$ are strictly antisymmetric. Fig. 2.5 shows an example of the double-wedge

property application in Shepp-Logan phantom. We create the second half sinogram data, which refers to $[\pi, 2\pi)$, by reflecting the first half one, and then form a complete sinogram as Fig. 2.5(b) shows. Then we can observe that Fourier transform of the artificial sinogram will show non-negligible Fourier coefficients outside the double-wedge region in Fig. 2.5(d), which is mainly because the broken part of the sin-like information will create non-negligible high frequency information in Fourier domain. A general calibration procedure could be implemented as:

1. Synthesize complete sinogram in $[0, 2\pi)$ by reversing the $[0, pi)$ sinogram as described.

2. Apply 2D Fourier transform to the synthetic sinogram and set an assessment function Q_{SF} to evaluate the non-negligible Fourier coefficients. To simplify the calculation, a binary mask function $M(u, v)$ in frequency domain is multiplied to ensure our focus on the region outside the double-wedge:

$$Q_{SF} = \frac{\iint real(\tilde{P}(u, v))M(u, v)dudv}{\iint M(u, v)dudv}, \tag{2.15}$$

   where

$$M(u, v) = \begin{cases} 1, & else \\ 0, & |v| \leq |u|\, r \end{cases} \tag{2.16}$$

3. Repeat 1 and 2 in search range and find the global minimum Q_{SF}.

2.2.4 Metric Evaluation Method

Metric evaluation, also known as scoring of reconstructions, is derived from the basic idea that the best approximation to the true COR will give the best reconstruction result.

In the past, scoring of a reconstruction was completed manually, selected with the fewest artifacts in the researcher's subjective opinion. This procedure has gradually developed to design a scalar metric to assess the reconstructed image quality. The best COR is found through optimization of the metric value with respect to the shift distance variable. We will introduce several metrics applied in this field:

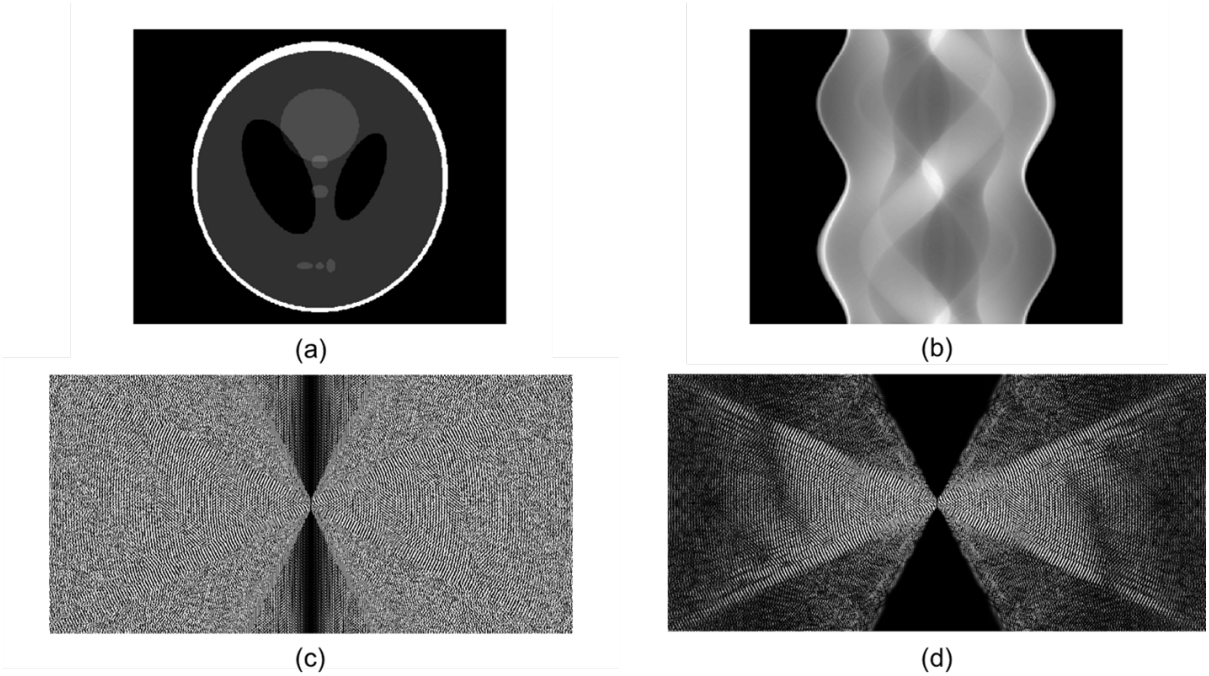


Figure 2.4: Illustration of double-wedge property of Bessel function in image frequency domain. (a) Shepp-Logan Phantom; (b) Corresponding $[0, 2\pi)$ parallel sinogram; (c) Fourier transform of the sinogram; (d) Significant Fourier coefficients after 25% thresholding.

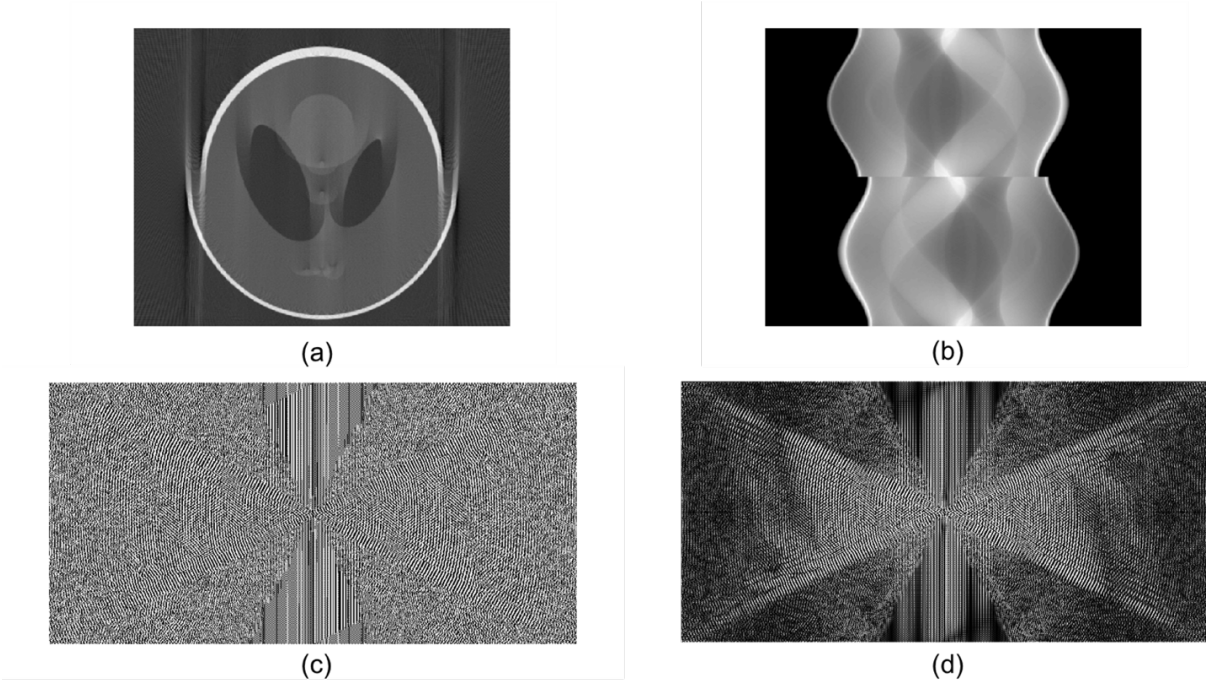


Figure 2.5: Illustration of frequency based method in COR correction (a) Reconstructed Shepp-Logan Phantom with 10 pixels COR bias; (b) Corresponding $[0, 2\pi)$ parallel sinogram; (c) Fourier transform of the sinogram; (d) Significant Fourier coefficients after 25% thresholding

Metric Q_{IA} (Integral of Absolute Value) The metric is defined as

$$Q_{IA} = \frac{1}{m_0} \iint \left| \tilde{f}(x,y) \right| dxdy, \tag{2.17}$$

where $\tilde{f}(x,y)$ represents the reconstructed image pixels and

$$m_0 = \iint f(x,y)dxdy = \int p(\theta,l)dl. \tag{2.18}$$

This is based on the mathematical explanation that if $f(x,y) \geqslant 0$,

$$\iint |f(x,y)|\, dxdy = \left| \iint f(x,y)dxdy \right| = \left| \iint \tilde{f}(x,y)dxdy \right| \leqslant \iint \left| \tilde{f}(x,y) \right| dxdy. \tag{2.19}$$

Hence, the best estimate of the COR can be reached when Q_{IA} reaches a minimum value.

Metric Q_{IN} (Integral of Negativity) The metric is defined as

$$Q_{IN} = -\frac{1}{m_0} \iint u[-\tilde{f}(x,y)]\tilde{f}(x,y)dxdy, \tag{2.20}$$

where

$$u(\alpha) = \begin{cases} 0, & \alpha < 0 \\ 1, & \alpha \geqslant 0 \end{cases} \tag{2.21}$$

Similar to the Q_{IA} method, the best COR can be found at the minimum value of Q_{IN}.

Metric Q_H (Histogram Entropy) The metric is defined as

$$Q_H = -\int p(g)\log_2(p(g))dg = -\sum_{k=1}^{G} p_k \log_2(p_k). \tag{2.22}$$

This originates from the basics of information theory in image processing, i.e., measuring the amount of information in the image. It is valid for the assumption that a wrong value of the COR will lead to extra structures and gradients in the reconstruction, and smear out the histogram of the image, which means it is in the most randomly distributed state and requires the most

information to represent when the COR shifts. Therefore, the best COR can be found at the minimum value of Q_H.

Metric Q_{TV} (Total Variation) The metric is defined as

$$Q_{TV} = \sum_{y=1}^{m} \sum_{x=1}^{n} \sqrt{I_x^2 + I_y^2} \tag{2.23}$$

where

$$[I_x, I_y] = [\frac{I(x+1, y) - I(x-1, y)}{2}, \frac{I(x, y+1) - I(x, y-1)}{2}]. \tag{2.24}$$

Total Variation (TV) is another way to measure the "information", i.e., the sum over all pixels of squared differences from neighboring pixels, in other words, measuring gradient information in the spatial domain. It is based on the assumption that TV is large if the image contains high-frequency signals such as noise and edges, and a tomographic image with errors (biases of the rotational axis and errors in the projection angle) contains extra artifacts of arcs and lines. Similarly, the best approximation of the COR can be found at the minimum value of Q_{TV}.

Metric Q_V (Global Variance) The metric is defined as

$$Q_v = \frac{\sum \sum (I(x, y) - \bar{I})^2}{\bar{I} \times N^2}. \tag{2.25}$$

This metric is based on the observation that small COR shifts lead to blurring of the image which degrades the variance of the image. Moreover, large COR errors will segment the useful information into small pieces, which will also decrease the variance of the image. Thus, the best COR can be found at the maximum value of Q_V.

2.3 Multi-level Thresholding Algorithm

As mentioned in the Introduction section, both in-plane rotation angle error and the off-centering bias error can highly affect the reconstruction image quality. However, analytical algorithms, e.g., the registration method or the centroid method, are sensitive to noise. Inspired by the

work of "automated motion correction" [30], we adapt the proposed coarse-fine search strategy in our research to improve the COR determination efficiency and meanwhile address the in-plane rotation angle error simultaneously.

2.3.1 Material

A CT phantom was used to verify the proposed method in practice, which is shown in Fig. 2.6. Specifically, it consists of two components, the QRM and the GmbH phantom. The QRM phantom, seen in Fig. 2.6(d), comprises two perpendicular 5×5 mm^2 silicon chips. Each chip contains a bar (trench) and a point pattern with diameters from 5 to 150 μm line/point thickness. The GmbH phantom, seen in Fig. 2.6(f), incorporates five electron densities inside, which approximates five different human tissues: air, muscle, water, adipose, and bone.

Furthermore, for this study, 4 C57BL/6 mice were produced by an in-house breeding colony maintained under protocol number 2018-3438, as approved by the Institutional Animal Care and Use Committee of Auburn University. All procedures were designed in accordance with the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. In this study, no living mice were imaged. The imaged carcasses were obtained from normal maintenance of a breeding colony. All mice were euthanized with CO_2 asphyxiation.

2.3.2 Coarse Search

In the coarse search, Zhu *et al.* [30] applied the center-of-mass algorithm to estimate the raw COR position given the assumption that the system works in parallel projection mode. However, in small-animal research, this assumption is not satisfied.

Instead, a coarse estimation of COR can be simply realized by calculating the average of two edges. The edges of the conical tube projection can be used as an indicator of the upper and lower boundaries, which are approximately equally displaced from the nominal axis of rotation.

$$COR_{raw} = \frac{P_{upboundary} + P_{botboundary}}{2} \tag{2.26}$$

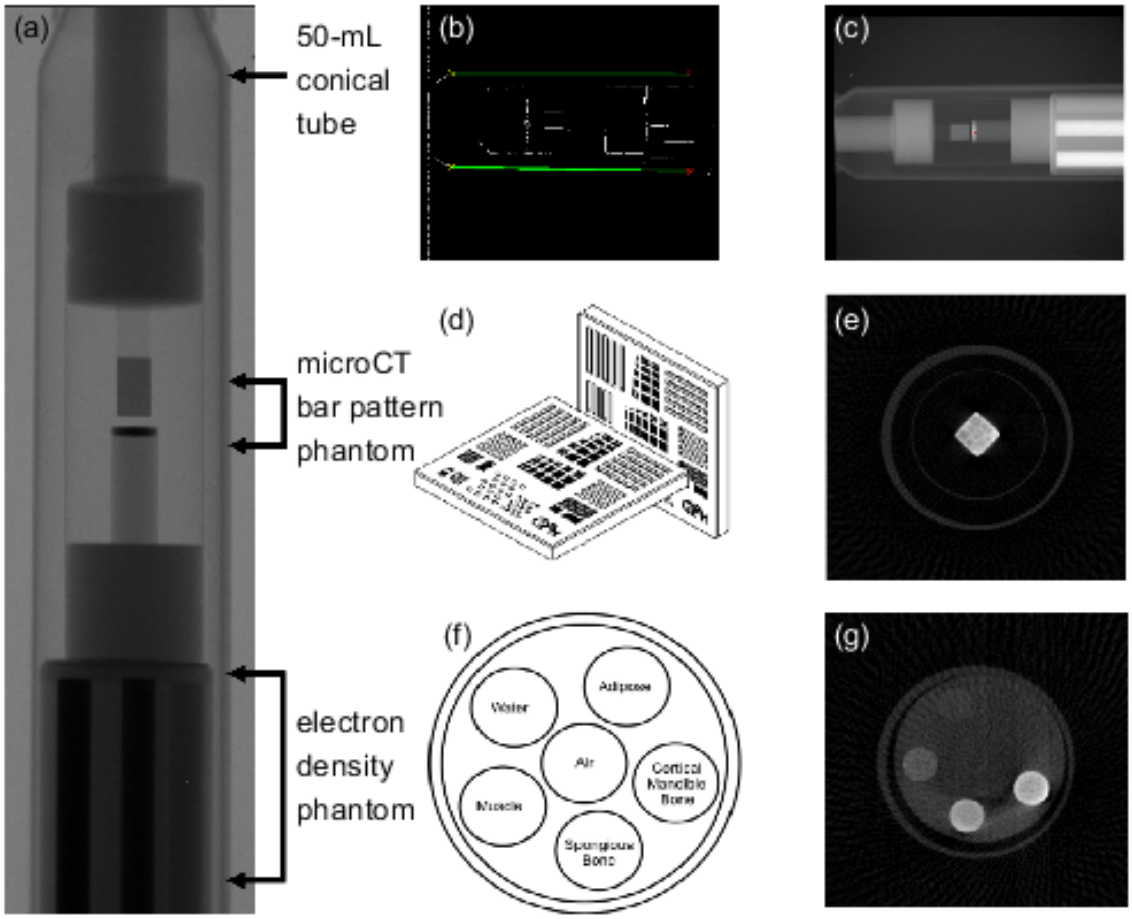


Figure 2.6: Arrangement of CT phantoms. (a) Single planar X-ray projection through a 50-mL cylindrical tube. (b) Edge detection on the image (a), where yellow and red dots are the endpoints of the detected lines.(c) Corresponding calibrated and centered projection image. (d) QRM micro-CT bar pattern chip phantom. (e) Reconstruction slice of QRM phantom with ST-TSO calibration method. (e) GmbH phantom with the indicated five human tissue equivalents densities: air, muscle, water, adipose, and bone. (f) Reconstruction slice of GmbH phantom after calibration

The boundaries of the holder have low contrast in the projection images. To address this problem, we proposed an edge detection algorithm as Fig. 2.7 shows. It is an auto boundary selection strategy combining the Canny detection algorithm [39, 40], the Hough Transform [41] and the RANSAC selection technique [42] that includes three steps:

1. Apply classic Canny kernel on scanned projection to detected boundary information. See Fig. 2.7(a).

2. Apply the Hough transform to automatically detect the horizontal lines. See Fig. 2.7(c).

3. Regard inner boundary coordinates as undesired points and select the desired boundaries with the RANSAC algorithm. See Fig. 2.7(d).

An average on 0°, 90°, 180° and 360° projection images is calculated to further remove the error caused by light scattering or device vibration during measurement. After rotation axis (COR) position is determined, in-plane rotation degree is estimated by calculating the angle between the rotation axis and the lateral axis simultaneously.

Fig. 2.8 shows an example of the proposed procedure in in-plane angle correction on mouse and CT phantom data. Numerical performance before and after the calibration is shown in Table 2.1. Since raw estimation is designed to find a suitable starting point to narrow down the search range in a later iterative algorithm in the fine search step, high accuracy is not required. Instead, three main advantages of the proposed coarse search step are as follows: 1) easy to realize and robust; 2) independent of the scanning mode; and 3) applicable to in-plane rotated projection images.

Table 2.1: In-plane rotation angle estimation before and after calibration

CT Phantom(#)	1	2	3	4	5	6	7	8	9	10
before	0.51	0.48	0.43	0.42	0.47	0.51	0.49	0.48	0.51	0.52
after	0.04	-0.05	0.07	0.08	0.15	0.01	0.12	0.22	0.17	0.08
Mouse Data(#)	1	2	3	4	5	6	7	8	9	10
before	-0.74	-0.65	-0.55	-0.66	-0.62	-0.60	-0.54	-0.69	-0.59	-0.60
after	0.01	-0.11	-0.05	-0.18	-0.08	0.02	-0.05	-0.12	-0.09	-0.25

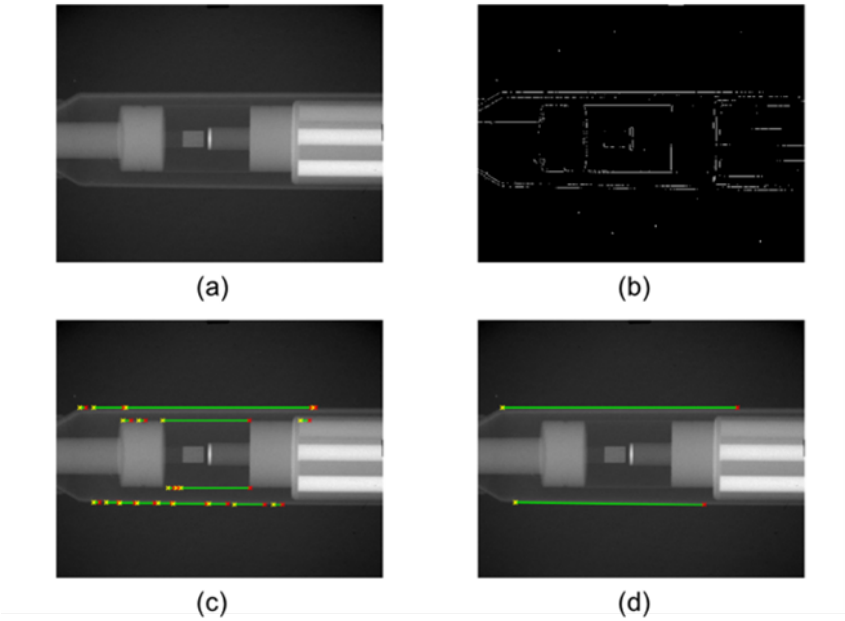


Figure 2.7: Demonstration of edge detection method. (a) Detected CT phantom projection, (b) Detected edge information with Canny kernel, (c) Auto-discovered edges with Hough Transform, (d) Determined boundaries with RANSAC selection algorithm

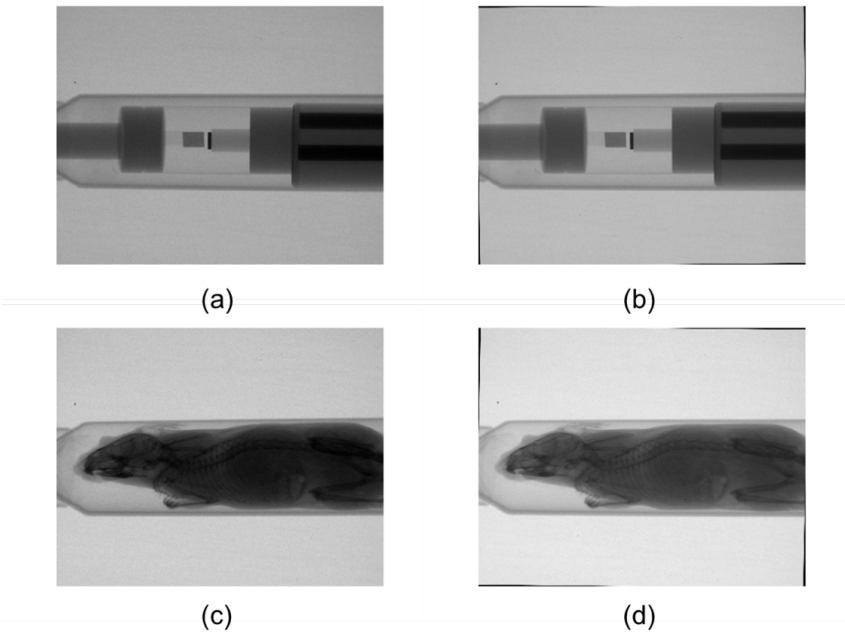


Figure 2.8: In-plane angle estimation and calibration with proposed edge detection algorithm. (a) Detected CT phantom projection, (b) CT phantom after calibration, (c) Detected Mouse projection, (d) Mouse projection after calibration.

2.3.3 Fine Search

Once the raw COR of the holder is determined, the fine search region is set to ± 30 pixels. Then we evaluate the reconstruction image quality by the metric evaluation algorithm for each possible COR position. In [30], the evaluation is defined on the global variance of the reconstruction image, based on the assumption that the maximum variance occurs when reconstructing with the correct COR [43]. Unfortunately, this does not hold for the low-dose case where global variance estimation is highly affected by the presence of scattering and blurring as well as the extra background structure caused by limited projections.

Therefore, we use a multi-level thresholding on the image histogram with Otsu's method to distinguish the ROI from the background under two assumptions: 1. background artifact values are generally smaller than specimen values, and 2. most variance contributions are from high-value components. To better demonstrate our proposed multi-level thresholding algorithm, an image analysis is presented in Fig. 2.9. For phantom data, we can segment the image into three groups: a background region that equals 0, streak artifacts that oscillate around 0, and useful signal which is the object region, as Fig. 2.9(a) shows. For the object image in Fig. 2.9(b), ROI classification includes three parts as well: bone structures, soft tissue region, and foam region, whose intensity is close to the background region. In contrast to the phantom data, the mouse data shows greater artifacts in the object region due to the existence of quantum noise. In addition, when COR error is small, global variance will be decreased by blurring in theory. However, extra streak artifacts from the background region will compensate this loss, which leads to global variance estimation being inaccurate. Segmenting the ROI from the background can effectively reduce the quantum noise and artifact impact on variance estimation.

Our revised variance estimation metric can be described as:

$$Q_{M-V} = \frac{\sum\sum(\tilde{I}(x,y) - \tilde{I})^2}{\tilde{N}^2}, \tag{2.27}$$

where

$$\begin{aligned} \tilde{I} &= Otsu(I,n) * I, \\ Otsu(I,n) &= \arg\max\{(\sigma_B^{'})^2\{t_1, t_2, ..., t_n - 1\}\}, \\ (\sigma_B^{'})^2 &= \sum_{k=1}^{n} \omega_k \mu_k^2, \end{aligned} \tag{2.28}$$

I is the reconstruction result, $\tilde{I}$ is the reconstruction after multilevel thresholding, $(\sigma_B^{'})^2$ is the modified between-class variance and $t_1, t_2, \ldots, t_{n-1}$ are multilevel thresholds in the histogram. The multi-level thresholding can be used to focus the variance calculation on image contents rather than background areas with noise.

For each reconstruction in the fine search range, the maximum value of Q_{M-V} indicates the correct COR position in the system, and its position in the fine search range can help calculate the bias of the COR.

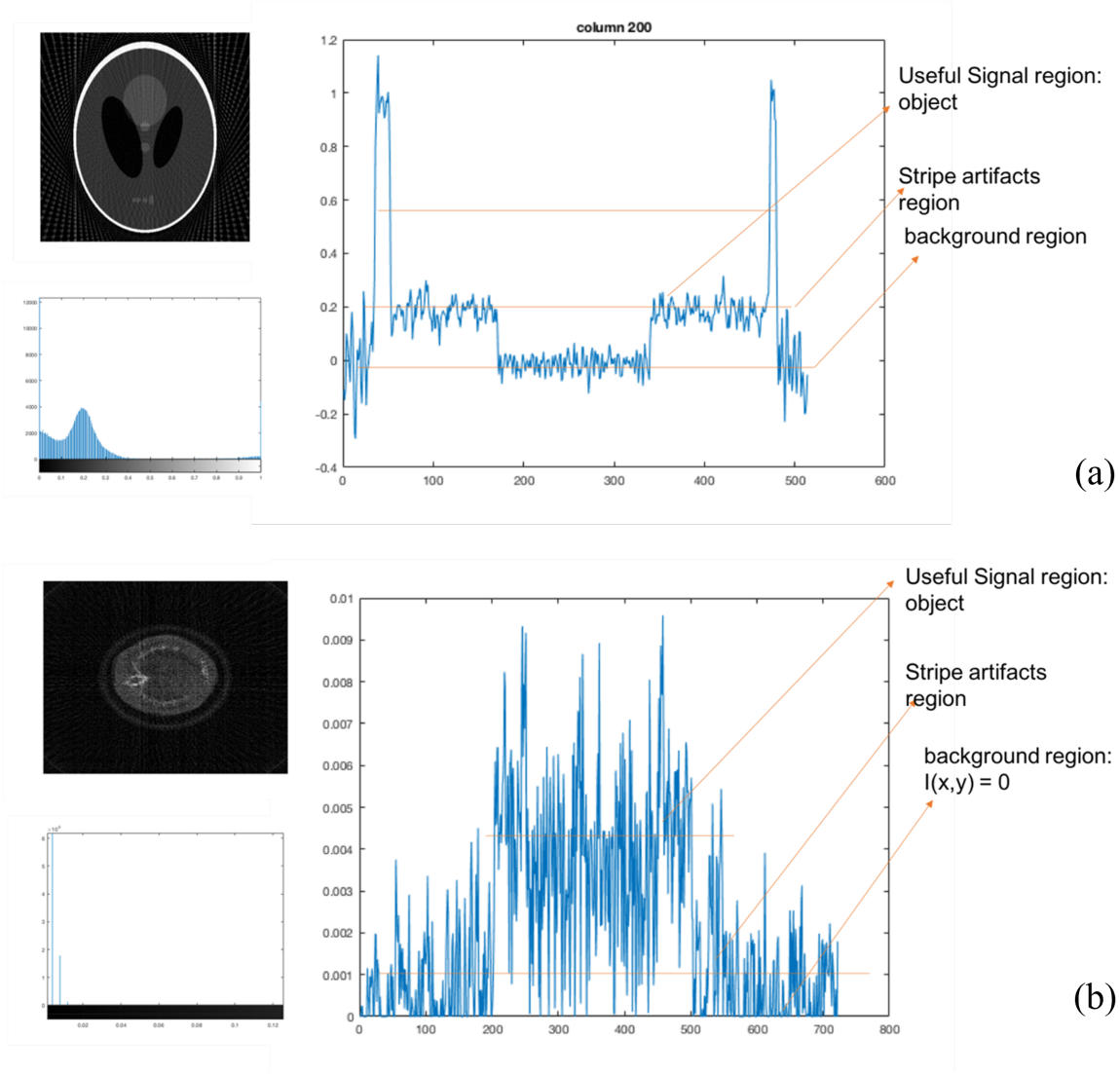


Figure 2.9: Analysis of S-L Phantom and Mouse reconstructions in intensity histogram distribution and line profile. A three-level thresholding was determined visually to demonstrate the multi-level OTSU method.

2.4 Simulation and Experiment

2.4.1 Shepp-Logan Phantom Simulation

To compare the performance of the proposed method, both variance-based and multi-level thresholding methods were tested on the Shepp-Logan phantom in an interval ± 15 pixels away from the correct COR position with sufficient projections. As Fig. 2.10 shows, an evaluation curve was obtained after an exhaustive estimation on 31 reconstruction images. We can observe that both algorithms are able to find the correct COR position corresponding to the largest value

in its estimation curve, indicating that our proposed criterion can produce similar performance compared with the widely used global variance estimation algorithm in high-contrast situations.

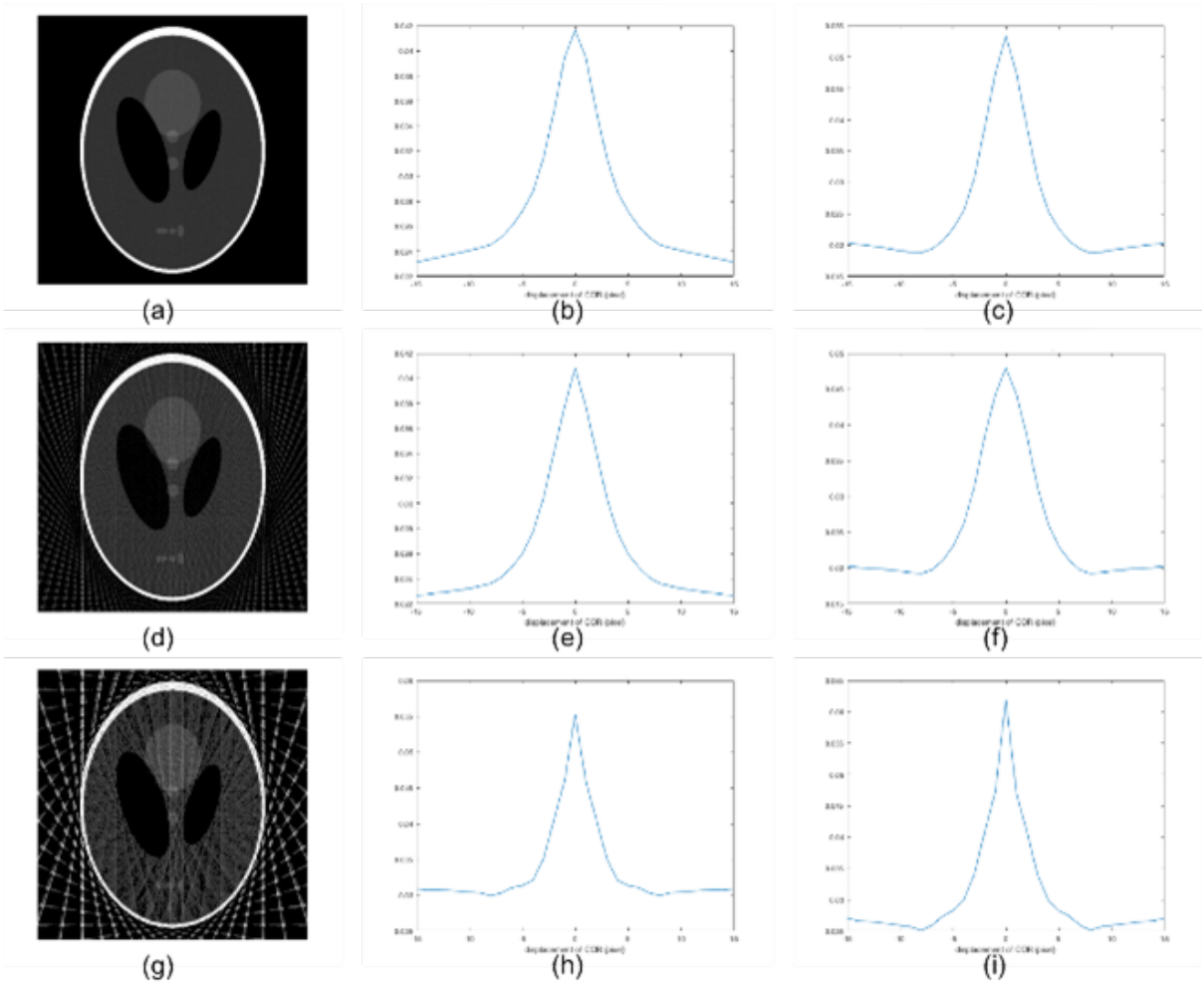


Figure 2.10: Fine search test on Shepp-Logan phantom with limited number of projections. (a), (d), (g) are correct reconstruction images from 360, 97, and 40 projections. (b), (e), (h) are estimation results from variance-based algorithm in [30], (c), (f), (i) are results from our method. X-axis is the search range from -15 to 15 pixels.

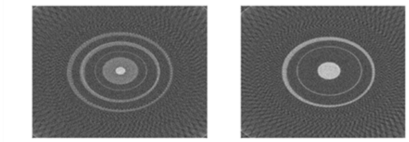
2.4.2 Actual Projection Data Test

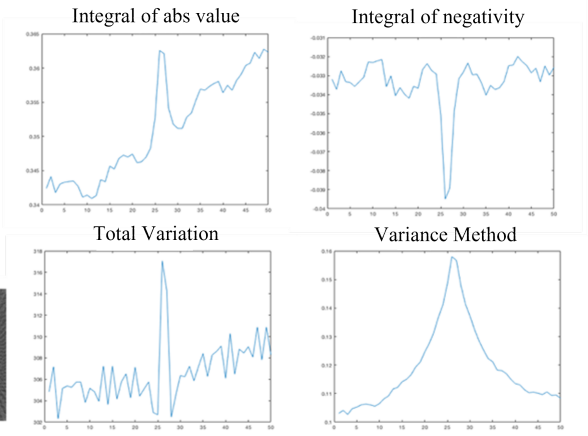
To further verify the validity of the proposed method in practice, we tested the proposed algorithm on actual scanned projection data. In summary, the IVIS imaging system provided a total of 97 projections in a full scan, and size of each projection is 1024×1024 pixels.

For comparison, those aforementioned objective functions were evaluated as well, and their performance is presented in Fig. 2.11 and Fig. 2.12. Due to the lack of a ground truth image, we ran an exhaustive search along the bias direction and determined the best reconstruction based on human observation, labeled ”Human Decision” in figures. In Fig. 2.11, which represents high contrast and simple structure, we could observe that the variance-based algorithm outperformed other algorithms with a much smoother objective function curve. Apparently, streak artifacts caused by sparse sampling can be considered as a pattern in background images with the variance-based algorithm, which also deviates when the COR changes. Therefore, variance reaches the maximum value when the image is reconstructed at the correct COR position. However, in Fig. 2.12, i.e., the mouse projection data, the variance curve around the correct position showed some difficulty in accurately determining the COR location.

We also tested our proposed algorithm on the mouse data. In order to compare the performance, the same coarse search was conducted first to make sure they exhaustively searched the same region. The raw FBP reconstruction result shown in Fig. 2.13(a) has significant artifacts. Meanwhile, a rough estimate of the center of the tube by a coarse search is pixel coordinate 551, which becomes the center point for the fine search. The curve in Fig. 2.12(c) describes the global variance estimated by the normalized variance equation, and the curve in Fig. 2.12(d) is local variance estimation after multi-level thresholding to eliminate the effect from the background. As a scoring method, its maximum position maps to the desired COR position. In Fig. 2.12(c), the maximum value occurred at -1, indicating the axis equals 550. In Fig. 2.12(d), the value occurred at -5, indicating a value of 546. Fig. 2.12(e) and Fig. 2.12(f) are corresponding reconstruction slices.

CT phantom	COR shift distance
Human Decision	27
Image Registration	34
Centroids Method	13
Integral of abs value	49
Integral of negativity	26
total variation	26
Variance based method	26





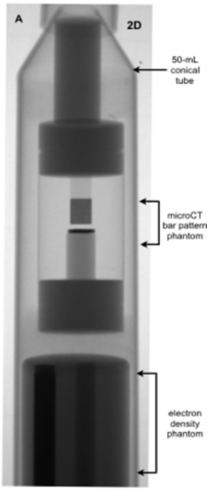
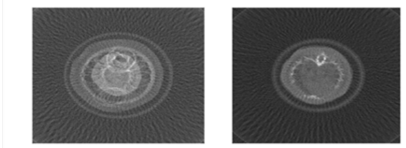
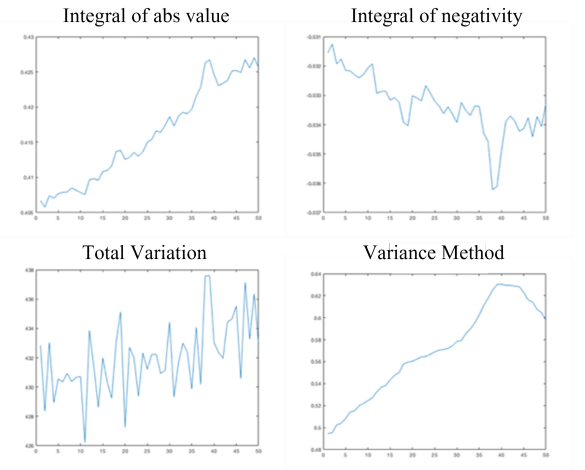


Figure 2.11: Test of aforementioned algorithms on CT phantom data: upper left table presents the final result of each algorithm on COR location, lower left image shows the result before and after calibration, middle plots show the curve of each metric-based algorithm, right shows the material in the test

Mouse	COR shift distance
Human Decision	43
Image Registration	42
Centroids Method	23
Integral of abs value	49
Integral of negativity	38
total variation	39
Variance based method	40





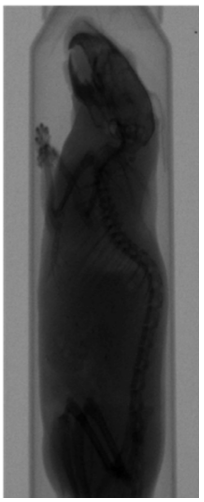


Figure 2.12: Test of aforementioned algorithms on mouse data: upper left table presents the final result of each algorithm on COR location, lower left image shows the result before and after calibration, middle plots show the curve of each metric-based algorithm, right shows the material in the test

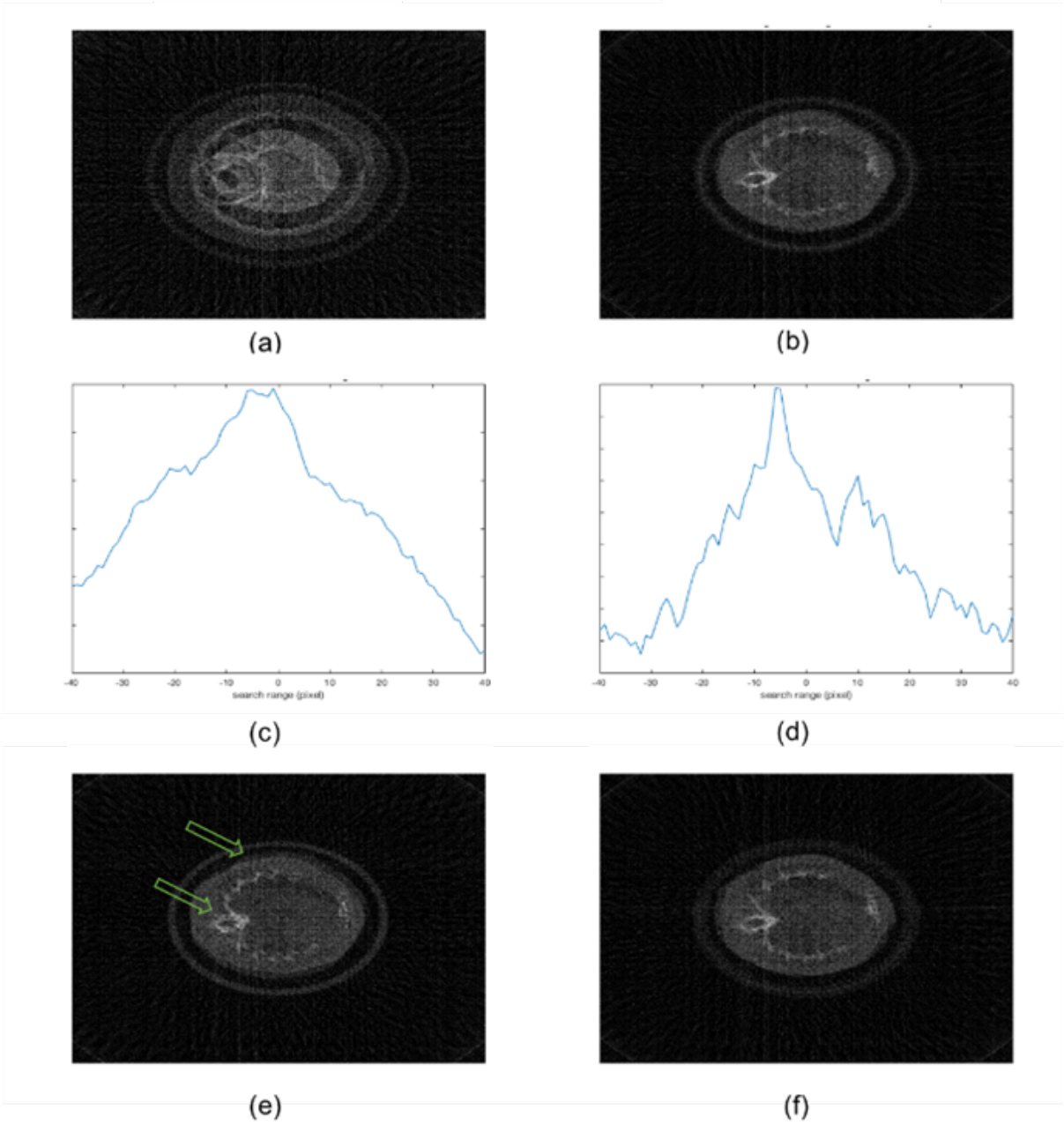


Figure 2.13: Application of coarse-fine search strategy on mouse dataset. (a) is FBP result without any correction, (b) is the best reconstruction result found within the search interval, (c) is variance estimation curve from [30], (d) is the curve from the proposed multi-level thresholding method, (e) is the image reconstructed at the maximum point of the curve (c), mis-aligned structure are pointed out by green arrows, (f) is the result at the maximum point of the curve (d), mis-aligned structure are fixed.

2.5 Conclusion and Discussion

Structure mismatch can occur in both bone structure and soft tissue regions when COR errors are made, as in Fig. 2.13(e). Besides, in low SNR cases, blur can act as a smoothing of the original signal that actually decreases the variance estimation when approaching the correct COR position. This can be observed as the flat region in the interval [-6,0] in Fig. 2.12(c), where the correct COR position is obscured. Multi-level thresholding can successfully address this drawback by separating the ROS from the background to increase the usefulness of variance estimation at lower SNR and provide a clear and accurate maximum for COR position estimation.

In summary, this work has shown that our revised method has superior performance in solving in-plane angle error and predicting the COR given a limited number of projections in the low SNR case.

Chapter 3

Structure Tensor Based Two-step Online Method for System Geometry Calibration

Although multi-level thresholding has shown improvement in system calibration, its focus on solving the dominant geometry parameters limits its accuracy. This work aims to develop an algorithm to calibrate more parameters. Therefore, we propose a revised method—structure tensor-based two-step online (ST-TSO) geometry calibration algorithm—that utilizes the projection image for system calibration. Specifically, in our first step of processing, we apply the PM method for coarse estimation of parameters. The results obtained are then fine-tuned through a second step, where an iterative optimization algorithm based on the global distribution of intensities is applied. Finally, the determined parameters are entered into a modified Feldkamp, Davis, Kress (mFDK) algorithm[44] for 3D volume reconstruction.

3.1 Introduction

A general workflow is presented in Fig. 3.1. Specifically, two datasets are needed, namely a baseline intensity calibration measurement and an actual measurement. An intensity calibration measurement detects the background of the system forward X-ray projections. A study on the background image reveals the characteristics of the X-ray photon distribution and thus provides a reliable estimate of the imaging geometry origin. Actual measurements consist of an X-ray of the intended object, including possible geometry errors.

The system configuration and layout are shown in Fig. 2.1, where Fig. 2.1(b) describes the 3D mathematical geometry. For an accurate geometry description and calibration, the coordinate system used in this thesis is depicted in Fig. 3.2(a). For a 3D object, the Y-axis is along

the object rotation axis direction, and the *Z*-axis OS is along with the X-ray focus spot (X-ray source S) direction from the origin. The *X*-axis is perpendicular to the *Y*-*Z* plane, consistent with the right-hand rule. For projection coordinates, image axes U and V parallel the X and Y.

Referring to Fig. 3.2, Table 3.1 gives the definition of all geometric parameters, m, n, θ, ϕ, η, SOD, and SDD. (m, n) are object COR offsets in X and Y direction (Fig. 3.2(c) and Fig. 3.2(d)). (θ, ϕ, η) describe the object orientation caused by the *MiSpinner* mechanical error or incorrect manual placement. SOD and SDD affect the image scaling.

In practice, SOD and SDD are constants for each time scanning since the imaging equipment is stationary during measurement and the movable stage is fixed at a preset level before scanning (i.e., device information in Table 2.1). For convenience, distance SOD and SDD are transformed into pixel length by dividing the pixel resolution, $r = 0.0098$ cm/pixel, calculated from Table 2.1. Moreover, a 3D-printed stage ensures the uniformity of the front and rear parts of the *MiSpinner*, thus decreasing the roll angle ϕ error. Therefore, this paper mainly focuses on estimating the remaining four parameters (m, n, θ, η).

Table 3.1: Definition of The Geometry Parameters

m	COR offset in *X* direction
n	COR offset in *Y* direction
θ(Tilt)	rotation angle of the object along the *X*-axis
φ(Roll)	rotation angle of the object along the *Y*-axis
η(Skew)	in-plane rotation of the object along the *Z*-axis
SOD	Source to object (COR) distance
SDD	Source to detector (the Scintillation Plate) distance

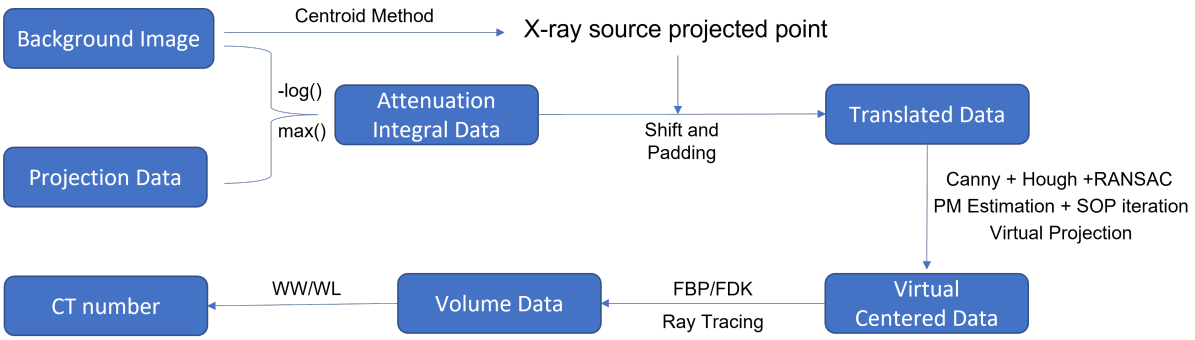


Figure 3.1: Introduction of system geometry calibration

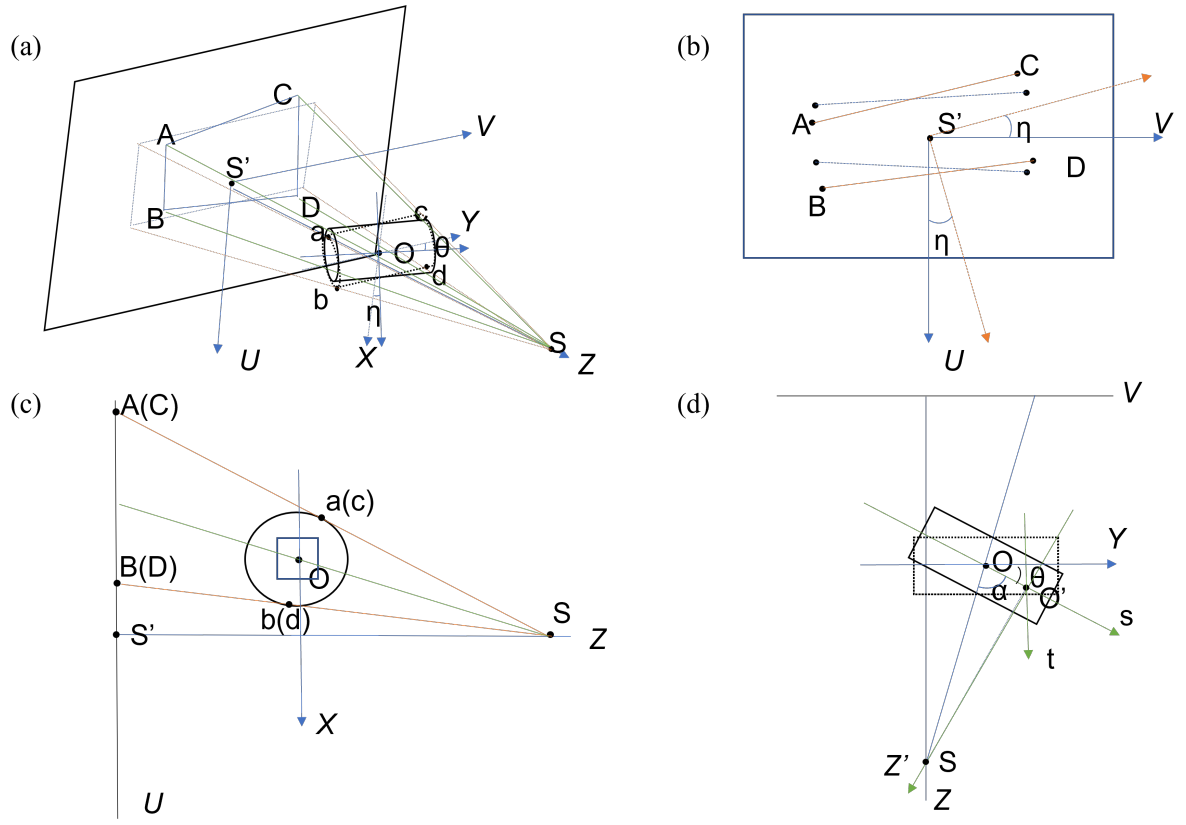


Figure 3.2: A description of imaging coordinates used in the described imaging system. *X*-, *Y*- and *Z*- axis obey the right-hand rule. Besides the coordinates, in each sub-plot, the corresponding geometric error is present as well. (a) A general 3D view of the scanning geometry. The dotted line shows the original projection of the cylinder phantom, and the solid line shows the actual projection due to the geometry errors. (b) Frontal plane of (a), the blue dotted line shows the error caused by the tilt (θ in (d)) parameter, and the orange dotted line shows the error caused by in-plane rotation (η) error. (c) Profile plane of (a), the orange line shows the vertical projection shift in *U*-direction caused by misalignment (m) in *X*-axis. (d) In the top view of (a), the orange line shows the lateral projection shift in *V*-direction caused by offset (n) in *Y*-axis. Tilt error (θ) is depicted as well.

3.2 Structure Tensor based Two-step Online Geometry Calibration Algorithm

3.2.1 System Origin Estimation

Calibration measurement reveals the photon distribution in cone-beam geometry.[45] For any pixel (i, j), its detected light intensity is

$$I = \frac{d\Omega}{4\pi} I_0 e^{-\sum a_i L_i} = \frac{dS \times D}{4\pi \times (\sqrt{D^2 + i^2 + j^2})^3} I_0 e^{-a\sqrt{D^2+i^2+j^2}}, \tag{3.1}$$

where dS is the unit area at pixel (i, j) whose origin is the X-ray source projected point, D is SDD in pixels, I_0 is raw intensity, and a is the attenuation coefficient of air. The equation indicates that the light intensity distribution is circular around the origin, which is shown in Fig. 3.3 . A raw estimation of the origin is acquired through a centroid calculation on the binarized background image. To minimize the electronic noise in scanning which is assumed to follow a Gaussian distribution, 40 background images were taken repeatedly and the averaged result was treated as the noiseless background image.

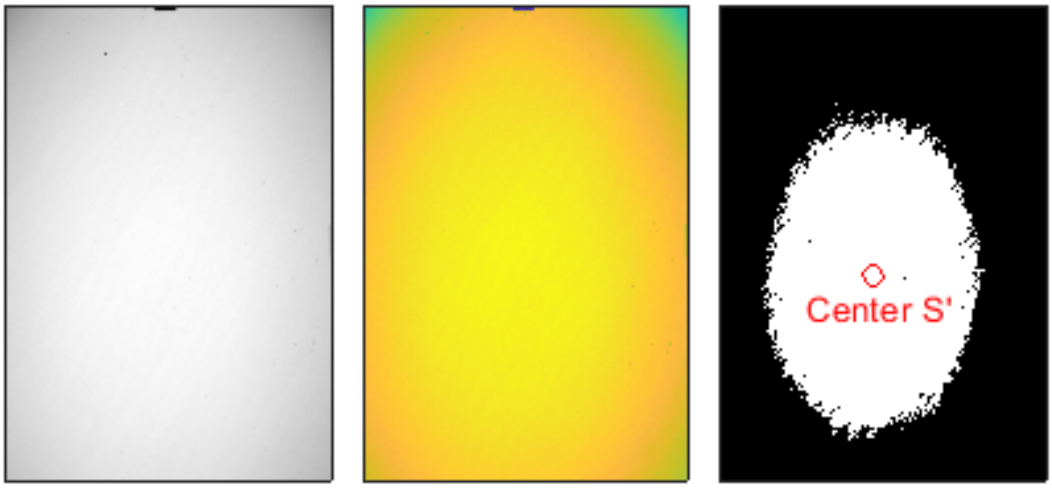


Figure 3.3: Estimation of the X-ray source projected point. (a) Background Image Ave: an average background image of 40 times image taken aiming to minimize the Poisson noise in photon propagation. (b) Pixel Value Distribution: background image intensities are transformed into color bar format to demonstrate symmetrical distribution, which appears as a light spot. (c) Binarized Background Image: estimation of the light spot center through centroid calculation by applying hard thresholding on the image (a).

3.2.2 First Step: Geometric Parameter Estimation

We model the object-rotating system as a pinhole camera model[46] described as

$$P' = MP_w = K[R, T]P_w, \tag{3.2}$$

where P_w are voxel coordinates in the world coordinate system, M is the projection matrix (PM) consisting of K, R, T: intrinsic matrix, rotation matrix, and translation vector. Each projected point P' can provide two forward projection equations. Thus, we need at least four points in a single projection to accurately estimate the seven parameters in the PM, which form a rectangle shown in Fig. 3.2(a). Equivalently, the four sides of a rectangle can provide the same rich information.

In our work, the application of the *MiSpinner* has provided a scenario for the PM method. The specimen holder, a cylindrical tube, is simultaneously projected onto the image when scanning. Its shape is almost a rectangle. The geometric information describing both sides can be obtained by the edge detection method described in the previous chapter. To figure out the forward mapping relationship (Fig. 3.2(a)), we build a local coordinate system, (s, t, z'), to assist the analysis, as Fig. 3.2(d) shows. The origin is the X-ray source point projected onto the s-axis, noted as O'. The length of $|OO'|$ is

$$|OO'| = |OS| \times \cos\alpha, \tag{3.3}$$

where

$$|OS| = \sqrt{SOD^2 + n^2}, \tag{3.4}$$

and

$$\alpha = \pi - \theta - \tan^{-1}(\frac{SOD}{n}). \tag{3.5}$$

We have points O and S (Fig. 3.2(d)) in new local coordinates equal to $(m, -|OS| \times \cos\alpha, 0)$ and $(0, 0, |OS| \times \sin\alpha)$. Then for tangent points a, b, c, d (referring to Fig. 3.2(c)), we obtain tangential equations

$$\begin{cases} (t-m)^2 + z'^2 = r^2 \\ \frac{t-m}{z'} \times \frac{t}{z'-|OS|\times\alpha} = -1 \end{cases} \tag{3.6}$$

Solutions of variables z' and t are

$$z'_1 = \frac{r^2 l_1 + mr\sqrt{m^2 + l_1^2 - r^2}}{m^2 + l_1^2}, \quad t_1 = \frac{m^2 - r^2 + z'_1 \times l_1}{m}, \tag{3.7}$$

and

$$z_2' = \frac{r^2 l_1 - mr\sqrt{m^2 + l_1^2 - r^2}}{m^2 + l_1^2}, \quad t_2 = \frac{m^2 - r^2 + z_2' \times l_1}{m}, \tag{3.8}$$

where l_1 is the length of $|O'S|$ that $|O'S| = |OS| \times \sin\alpha$ and $|OO'|$ is denoted as l_2. Hence, tangent points a, b, c, d in local coordinates are a$(t_2, -l_2 - l, z_2')$, b$(t_1, -l_2 - l, z_1')$, c$(t_2, -l_2 + l, z_2')$, and d$(t_1, -l_2 + l, z_1')$ if $m < 0$ or a$(t_1, -l_2 - l, z_1')$, b$(t_2, -l_2 - l, z_2')$, c$(t_1, -l_2 + l, z_1')$, and d$(t_2, -l_2 + l, z_2')$ if $m > 0$. l ($l > 0$) is a constant. In the following analysis, we use the $m < 0$ case for convenience.

If we translate the local coordinates to point O, new tangent coordinates will be a$(t_2 - m, -l, z_2')$, b$(t_1 - m, -l, z_1')$, c$(t_2 - m, l, z_2')$, and d$(t_1 - m, l, z_1')$. According to Eq. (3.2), we decompose the PM method into two steps. First, rotation and translation are considered:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \cos\eta & -\sin\eta & 0 \\ \sin\eta & \cos\eta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} t \\ s \\ z' \end{bmatrix} + \begin{bmatrix} m \\ n \\ \tilde{z} \end{bmatrix}, \tag{3.9}$$

where $\tilde{z} = SDD - SOD$. Substituting the coordinates of each point into the formula, we calculate their new coordinates in (x, y, z). Second, points are projected onto the detector plane, noted as A, B, C, D (Fig. 3.2(a)) by vector projection; that is, $\vec{A} = \vec{S} + \overrightarrow{SA} = \vec{S} + k \times \overrightarrow{Sa}$. Then the two edges AC and BD (Fig. 3.2(a)) have the extension relationships:

$$K_{AC} = \frac{-kX_2 \sin\eta + X_2 \cos\eta \sin\theta - n \sin\eta \sin\theta + m \sin\theta}{-kX_2 \cos\eta - X_2 \sin\eta \sin\theta - n \cos\eta \sin\theta - n \sin\theta}, \tag{3.10}$$

$$K_{BD} = \frac{kX_1 \sin\eta + X_1 \cos\eta \sin\theta - n \sin\eta \sin\theta + m \sin\theta}{kX_1 \cos\eta - X_1 \sin\eta \sin\theta - n \cos\eta \sin\theta - n \sin\theta}, \tag{3.11}$$

$$b_{AC} = \frac{X_2 \cos\theta + n \cos\theta \sin\eta + m \cos\theta \cos\eta}{-kX_2 \cos\eta - X_2 \sin\eta \sin\theta - n \cos\eta \sin\theta - n \sin\theta} \cdot (-SDD), \tag{3.12}$$

and

$$b_{BD} = \frac{X_1 \cos\theta + n \cos\theta \sin\eta + m \cos\theta \cos\eta}{kX_1 \cos\eta - X_1 \sin\eta \sin\theta - n \cos\eta \sin\theta - n \sin\theta} \cdot (-SDD), \tag{3.13}$$

where

$$X_1 = \frac{-mr^2 + rl_1\sqrt{m^2 + l_1^2 - r^2}}{m^2 + l_1^2}, \tag{3.14}$$

$$X_2 = \frac{-mr^2 - rl_1\sqrt{m^2 + l_1^2 - r^2}}{m^2 + l_1^2}, \tag{3.15}$$

$$k = \frac{\sqrt{m^2 + l_1^2 - r^2}}{r}, \tag{3.16}$$

and

$$l_1 = |OS| \sin\alpha = n \cdot \sin\theta + SOD \cdot \cos\theta. \tag{3.17}$$

K_{AC} and K_{BD} are slopes of two lines, and b_{AC} and b_{BD} are two intercepts. SDD, SOD, and r are known constants where r is the cylindrical tube radius equal to 29 mm. This extension frees us from finding specific geometric relation points (e.g., $a - A$ pair in Fig. 3.2(a)) on each projection.

3.2.3 Constrained Condition

Theoretically, the above equations can be solved analytically. However, due to the parallel nature of the detected edges, under some special conditions, our proposed line conditions may encounter a limitation. That is, in Eqs. (3.10)-(3.13), when θ and η are equal to zero, variable n will be eliminated from the formula, leading to n being unsolvable. Therefore, we propose adding an extra condition to assist in solving the problem.

Specifically, for long-object scanning, which refers to the projection of objects (including the projection of conical tubes) covering the field of view (FOV) along the *Y*-direction (Fig. 3.2(a)), the value of n that predicts the offset in the *Y*-direction will be equal to 0 when we migrate the projection image centered at the system origin. And, as we described before, the system origin can be estimated through the centroid coordinates of the background image. Considering the complexity of finding an analytical solution, we seek to find its numerical

solution by minimizing the least squares cost function as follows:

$$F = (K_{AC} - RHS_1)^2 + (K_{BD} - RHS_2)^2 + (b_{AC} - RHS_3)^2 + (b_{BD} - RHS_4)^2, \quad (3.18)$$

where RHS_i is the corresponding right-hand side part in Eqs. (3.10)-(3.13). A gradient-free algorithm, the Nelder-Mead simplex method [47], is employed to find the optimized value of the remaining m, θ, η parameters starting from zero initialization. Redundant equations with an extra system origin coordinate condition ensure that variables converge to near-optimal solutions.

For short-object scanning, to overcome the limitation in n estimation, an extra condition, the image structure tensor, is added [48]. The structure tensor measures the geometry of image structures in the neighborhood of each point, and its definition is shown below:

$$S_K u(x) = K * [\bigtriangledown u^T \bigtriangledown u](x), \quad (3.19)$$

where K stands for a Gaussian filter, $\bigtriangledown u$ stands for image gradient, and $*$ is convolution. In the image domain, it is a 2×2 matrix for each pixel. Through the eigendecomposition of the structure tensor, acquired eigenvectors define the orientation of maximum and minimum variation of image u, and the square root of the corresponding eigenvalues measures these variations.[48] Typically, when both eigenvalues are relatively small, it means a small variation in all directions, indicating a homogeneous region. When one value is large and the other is small, it means strong variation exists but only in a dominant direction, indicating an edge area. When both values are large, it means strong variation in all directions, implying a corner in the image. Therefore, in short-object scanning, we combine the structure tensor with the edge detection technique to help detect the object vertices on projections (refer to Fig. 3.2(a) A,B,C,D points). The average of the point coordinates in the *Y*-direction multiplied by a scaling factor SOD/SDD yields a good estimate of the n parameter, and the remaining parameters can be obtained numerically from Eq. (3.18) as well.

3.2.4 Second Step: Parameter Fine-Tuning

For a non-convex optimization problem, the Nelder-Mead simplex algorithm sometimes will fall into a local minimum trap. Hence, in our second step, parameter fine-tuning is performed. Instead of using local image information, the second step utilizes the SOP property to analyze the full intensity distribution [22].

$$SOP(u,v) = \int R^{\alpha'}[\mu(r,\alpha,z)]d\alpha', \tag{3.20}$$

where $R^{\alpha'}[.]$ represents the Radon transform of the attenuation coefficients μ at rotation angle α'. The SOP property has been proven to be symmetric about the rotation axis in a perfect imaging geometry in a full scan[49]. Thus, ideally, the difference between the upper and lower half of the SOP along the rotation axis equals zero, as the equation shows:

$$Err(ideal) = \sum\sum[SOP(u,v) - SOP(2u_c - u, v)]^2 = 0, \tag{3.21}$$

where (u,v) are coordinates in the detector plane and u_c is the rotation axis position. A small error in the geometry will lead to the system being asymmetric and thus increase the difference.

Accordingly, to verify the correctness of the estimated parameters, we first resample the original projections to obtain corresponding virtual projections (V and V' in Fig. 3.4) and then test the SOP Err difference on them. Note that in Fig. 3.4(b), the existence of the tilt angle θ modifies the SOD and SDD into SOD' and SDD' in the virtual projection CBCT system.

$$SOD' = SOD \cdot \cos\theta - n \cdot \sin\theta, \tag{3.22}$$

and

$$SDD' = SOD' \cdot SDD/SOD. \tag{3.23}$$

Traditionally, parameters are considered independent of each other [26]. We fix the geometry error in the order of $(\eta, n, \theta, m, \varphi)$. First of all, in-plane rotation angle η is corrected by the corresponding reverse rotation. Then n and θ are fixed by linear interpolation (Fig. 3.4(b)):

$$v = \frac{SDD' \cdot (V_0 + v' \cos\theta)}{SDD' - v' \sin\theta}, \tag{3.24}$$

and

$$u = \frac{SDD' \cdot u'}{SDD' - v' \sin\theta}, \tag{3.25}$$

where u' and v' are the coordinates in virtual projections and u and v are the coordinates in the original projection. The origin of the CBCT virtual projection coordinates is point V_0 (Fig. 3.4(b)) which is the projection of the object center. Finally, m and φ are fixed by

$$\frac{(u - U_0)\cos\varphi}{u'} = \frac{(u - U_0)\sin\varphi + \sqrt{U_0^2 + SDD'^2}}{\sqrt{U_0^2 + SDD'^2}} = \frac{v}{v'}, \tag{3.26}$$

where U_0 is the offset of the object center in the detector plane (Fig. 3.4(a)).

In summary, the second-step iteration includes three steps: prediction of spatial parameters, creation of the corresponding virtual projection, and calculation and assessment of the SOP Err. These steps are executed repeatedly until the stopping criterion is satisfied.

To avoid the local minimum trap, people use heuristic algorithms such as simulated annealing (SA) [50]. In this work, to accelerate the convergence process, we apply a modified Evolutionary Annealing Simplex (EAS) algorithm [51], which combines the advantage of the SA algorithm that gets rid of the local minimum and the advantage of the Nelder-Mead simplex method that converges faster.

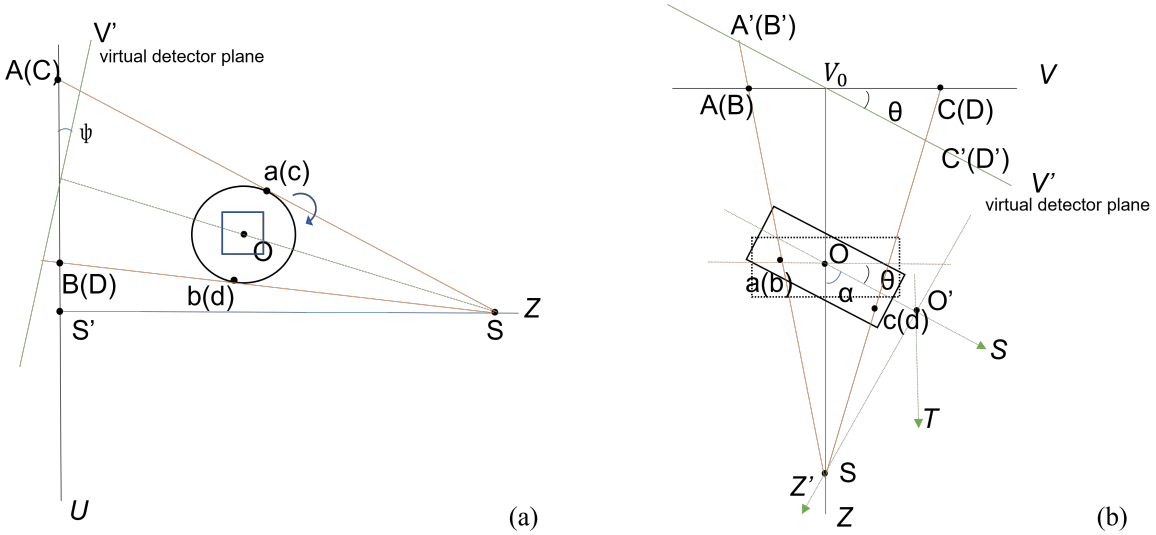


Figure 3.4: Virtual projections re-sampled from the original projections: the original detector plane is in black where projected points from the object a,b,c,d are noted as A,B,C,D. The virtual detector plane is green where projected points are A',B',C',D'. (a) virtual detector plane in *X*-*Z* plane to fix the geometric error φ and m, where artificial roll angle φ is created by m. (b) virtual detector plane in *Y*-*Z* plane is used to fix the geometric error θ and n; in this sketch, $n = 0$.

3.3 Experimental Results

3.3.1 Digital Phantom Simulation

A digital simulation was conducted to support our statement that direct PM calibration is not able to accurately estimate all geometric parameters in our research. We created a digital cylinder phantom to approximate the cylindrical tube in scanning and a cube inside to simulate the specimen. The size of the cylinder was 256×256×400 in height×width×length in pixels, and the size of the cube was 128×128×128 outside and 64×64×64 inside. SDD was set at 2000 pixels, and SOD was 1500 pixels in the simulation. Digital phantom simulations were performed in both noisy and noiseless cases. For the noisy case, we added both Poisson noise and Gaussian noise as follows:

$$I = Poisson(I_0 e^{-q_i}) + Gauss(0, \sigma_e^2), \tag{3.27}$$

where I_0 represents the intensity of the incident X-ray source, q_i is the line integral of the phantom volumes corresponding to pixel i on projection image, and σ_e^2 is the background electronic noise variance. In this simulation, we set I_0 to 10,000 and σ_e^2 to 10.

Six cases were simulated and the prediction results are presented in Table 3.2. For each case, we first obtained the forward projection with geometric errors according to the preset parameters and then used the calibration method to detect these parameters to verify the accuracy of the method. Specifically, Case 1 is a standard case in which each parameter equals zero, Cases 2–5 have only one parameter modified in the simulation, and Case 6 is the combination of all parameters. From Table 3.2, we can observe that the direct PM method shows obvious defects in the parameter n prediction in most cases, and the maximum error is 251.74 pixels. On the contrary, our proposed ST-TSO method shows better performance. In the noiseless simulation, the maximum errors for each parameter θ, η, m and n are $1.82°$, $0.055°$, 0.62 pixel and 6.24 pixels, respectively. In the noisy simulation the maximum errors have increased to $2.22°$, $0.055°$, 2.95 pixels, and 6.37 pixels, respectively, but are still within the margin of error proposed by Xi *at al.* [23]. To better view the effect of geometric error in the reconstructed volume, Fig. 3.5 shows the same phantom slice before and after calibration in each case. The results show that the proposed ST-TSO method can successfully fix the geometric error in both noisy and noiseless cases, verifying the robustness of the proposed method.

3.3.2 Fundamental Study

The Hounsfield Unit (HU) scale is a linear transformation of the original linear attenuation coefficient measurement [52]. Specifically, the radio density of distilled water at standard pressure and temperature (STP) is defined as 0 HU, while the radio density of air at STP is defined as -1000 HU. In a voxel with average linear attenuation coefficient μ , the corresponding HU value is therefore given by:

$$HU = 1000 \times \frac{\mu - \mu_{water}}{\mu_{water} - \mu_{air}}, \tag{3.28}$$

where μ_{water} and μ_{air} are respectively the linear attenuation coefficients of water and air.

Table 3.2: Calculated Results of Varying Misalignment

	θ (°)	η (°)	m (pixel)	n (pixel)
Case 1	0	0	0	0
PM method	0	0	0.5567	0
ST-TSO method	0.0018	0.0013	0.5578	0.0004
@ noisy case	0.4284	-0.0052	0.6100	-0.4927
Case 2	5	0	0	0
PM method	6.9240	-0.0023	0.9661	-87.4890
ST-TSO method	6.8158	-0.0001	0.6170	-0.1762
@ noisy case	6.6581	-0.0022	0.6155	-1.2840
Case 3	0	5	0	0
PM method	0.2378	4.9146	22.3509	-251.7403
ST-TSO method	0.9769	4.9907	0.4550	1.4841
@ noisy case	0.5886	5.0011	0.5659	0.1915
Case 4	0	0	-100	0
PM method	-0.0241	-0.0037	-98.7257	26.3856
ST-TSO method	0.2168	0.0113	-99.3883	-0.9203
@ noisy case	0.5424	0.0298	-99.3832	-0.3726
Case 5	0	0	0	-100
PM method	0	0	0.5567	0
ST-TSO method	-0.3349	-0.0004	0.6171	-99.9070
@ noisy case	0.1874	0.0052	0.5973	-93.7637
Case 6	5	5	-100	-100
PM method	5.2766	4.8963	-84.3994	-291.7787
ST-TSO method	6.319994	5.049118	-99.9043	-93.6365
@ noisy case	5.261392	4.958395	-99.6007	-98.0518

A change of one HU represents a change of 0.1% of the attenuation coefficient of water since the attenuation coefficient of air is nearly zero. Moreover, the linear attenuation coefficient is related to the incident X-ray source energy. Thus in this thesis, a fundamental study on determination of the specific water and air linear attenuation coefficients, which are close to our imaging environment, is performed.

An air test is shown in Fig. 3.6 where we measured the conical tube projection data without any sample included. Fig. 3.6(c) shows the reconstruction where each pixel represents the linear attenuation coefficient after post-log calculation following Beer's Law. Uneven distribution around the tube wall indicates the existence of the tilt angle error in the system scanning geometry. Fig. 3.6(d) shows the intensity distribution of the reconstructed image, which approximates a Gaussian distribution, and the mean value equals 0.00067/mm. Note that we

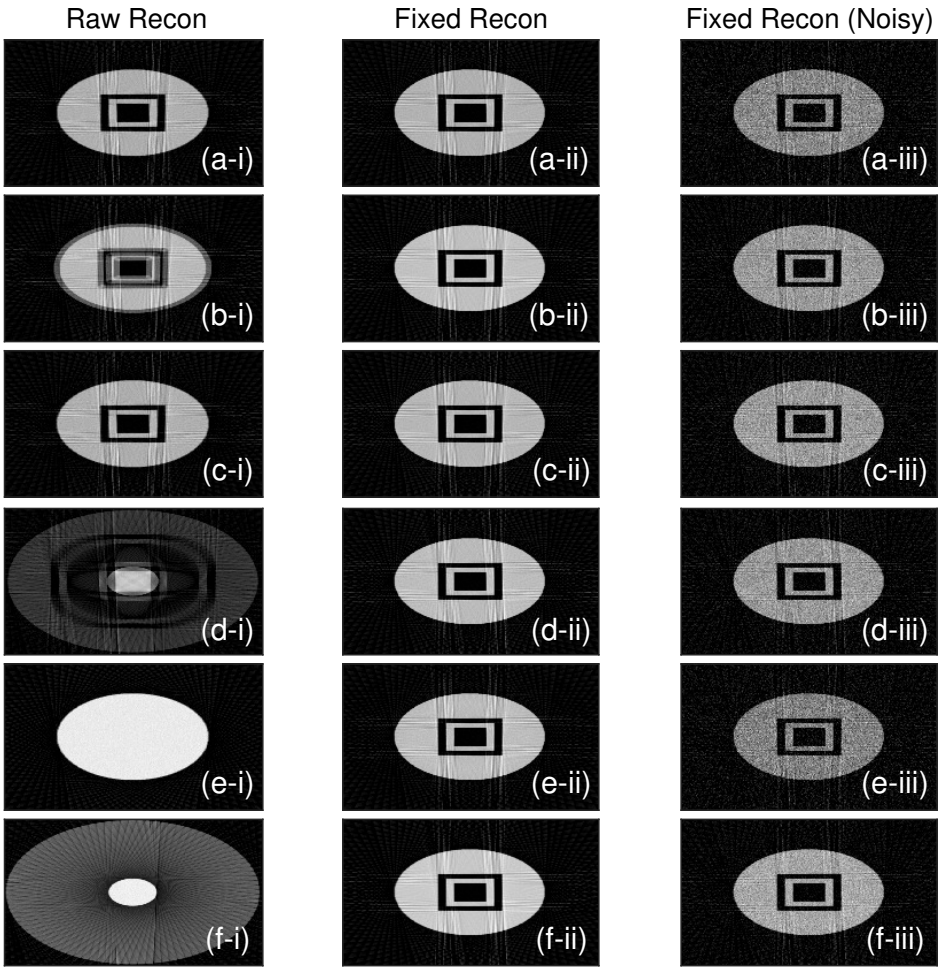


Figure 3.5: Same reconstructed slices using different sets of geometrical parameters (θ, η, m and n). Columns (i), (ii) and (iii) are the results for Raw Recon showing reconstruction without calibration, Fixed Recon showing reconstruction after the proposed ST-TSO method calibration on noiseless projection data and Fixed Recon (Noisy) showing reconstruction after the calibration on noisy data, respectively. Rows (a)-(f) represent different geometry parameter cases: (a) standard case that all parameters equal zero, (b) $\eta = 5$, (c) $\theta = 5$, (d) $m = -100$, (e) $n = -100$ and (f) $\theta = 5, \eta = 5, m = -100, n = -100$.

multiply the reconstruction with a mask image, whose value equals 1 in the ROI and 0 outside, for air attenuation coefficients estimation, thus leading to the large spike in Fig. 3.6(d). Similarly, a water test is presented in Fig. 3.7 and the mean value of the water attenuation coefficient is 0.005/mm.

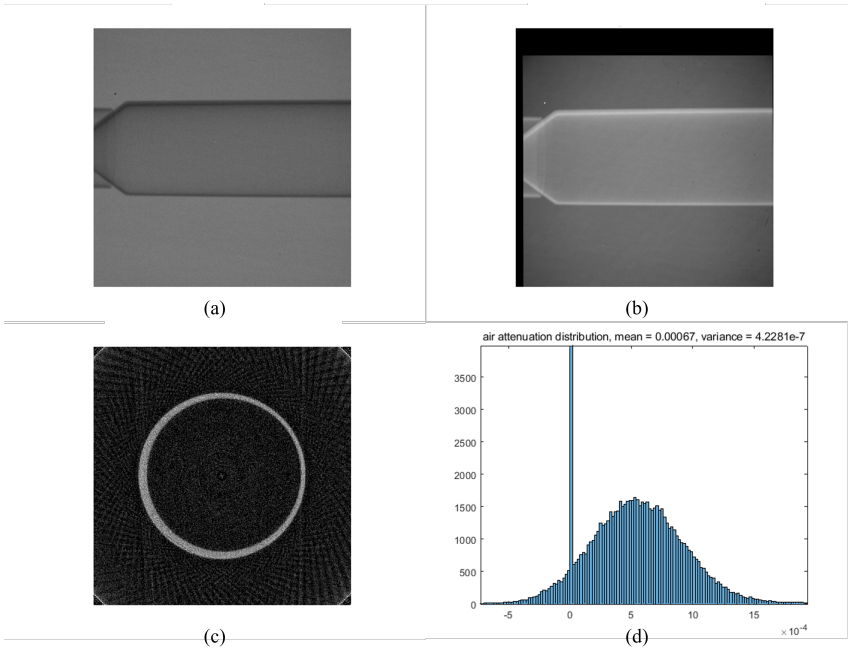


Figure 3.6: Air linear attenuation coefficient test. (a) is empty conical tube projection, (b) is corresponding after-log projection data, (c) is FDK reconstruction result, and (d) is corresponding intensity distribution, the large spike arises from the background region outside the ROI.

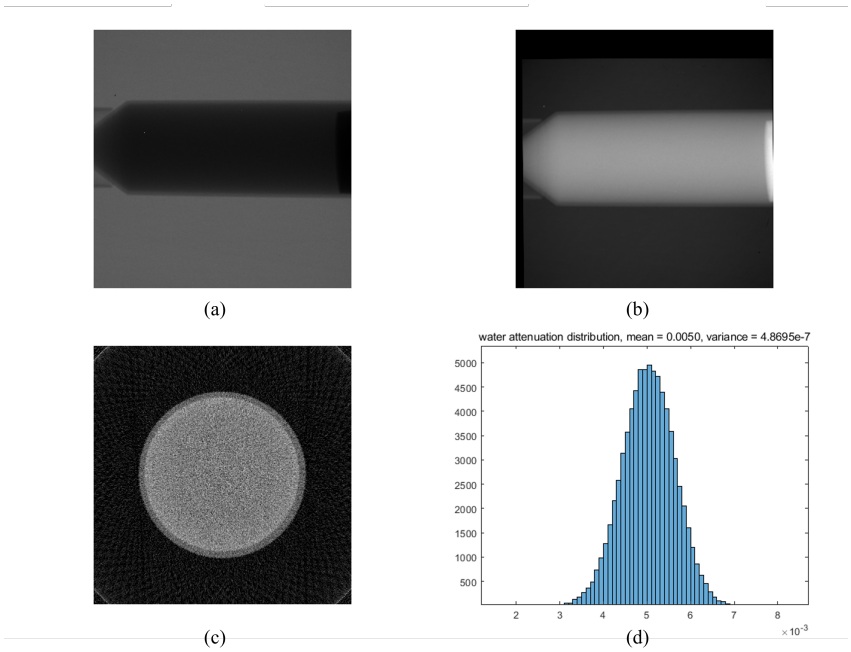


Figure 3.7: Test on water linear attenuation coefficient. Similar to air test, (a) is raw projection data, (b) is corresponding after-log projection with geometry calibration, (c) is FDK result, and (d) is its histogram

3.3.3 CT Phantom Test and Analysis

Finally, a qualitative comparison of the CT reconstruction in different processing cases is demonstrated in Fig. 3.8. CT slices (Figs. 3.8(a), (e), and (i)) in the first column show that the reconstruction without calibration completely ruins the phantom structure in all regions. CT slices (Figs. 3.8(b), (f), and (j)) in the second column reveal that the global variance-based online calibration algorithm (noted as the GV method) proposed by Dong *et al.* [33] can only recover the structures owing to thickness (Figs. 3.8(b) and (f)) but does not work in the thin case (Fig. 3.8(j)). This is mainly because of the assumption made by Dong *et al.* that approximated the 3D volume into multi-2D slices. The filtered back propagation (FBP) method cannot reconstruct a thin slice due to the lost image information. Furthermore, structure deformation (Fig. 3.8(f) density phantom) is also in evidence. The third column presents the results after FDK reconstruction on projections where only two dominant parameters, vertical offset m and in-plane rotation angle η, are fixed by our ST-TSO method. The fourth column displays the results fully calibrated by the proposed method. We can visually observe that the last two column results outperformed the first two column results.

To assess the performance numerically, a further comparison on bar pattern slices was conducted, shown in Fig. 3.9. We cropped the silicon regions in Figs. 3.9(b), (c), and (d), and two numerical analysis methods—full width at half maximum (FWHM) method and module transfer function (MTF)—were applied. FWHM measures the separation between two individual point sources, and MTF measures the relative contrast in the image. MTF is defined as the absolute value of the optical transfer function, which is calculated as the discrete Fourier transform of the line spread function. Its values indicate how much of the object's contrast is captured in the image as a function of spatial frequency. In theory, when drawing a line (the red one) crossing through the region, its cross-sectional curve depicted in Figs. 3.9(a), (d), and (g) would drop rapidly at the edge side like a step function. However, the line profiles varied gradually due to the line spread function (LSF) effect. Given the spatial resolution at position B, 0.0073 cm/pixel, the FWHM of each plot was 0.1703 mm, 0.1824 mm and 0.1745mm (Figs. 3.9(b), (e), and (h)) respectively, which was consistent with our observation on CT bar

phantoms that 150 μm resolution could be recognized. The MTF analysis provided similar results (Figs. 3.9(c), (f) and (i)). All three MTF curves have a 10% drop-off frequency of around 0.13 mm. Moreover, compared with Fig. 3.9(f) and Fig. 3.9(i), we can observe that Fig. 3.9(i) provides a smoother standard MTF curve (the blue one), illustrating that the proposed method of fixing all geometric parameters can provide better contrast details than fixing only two parameters.

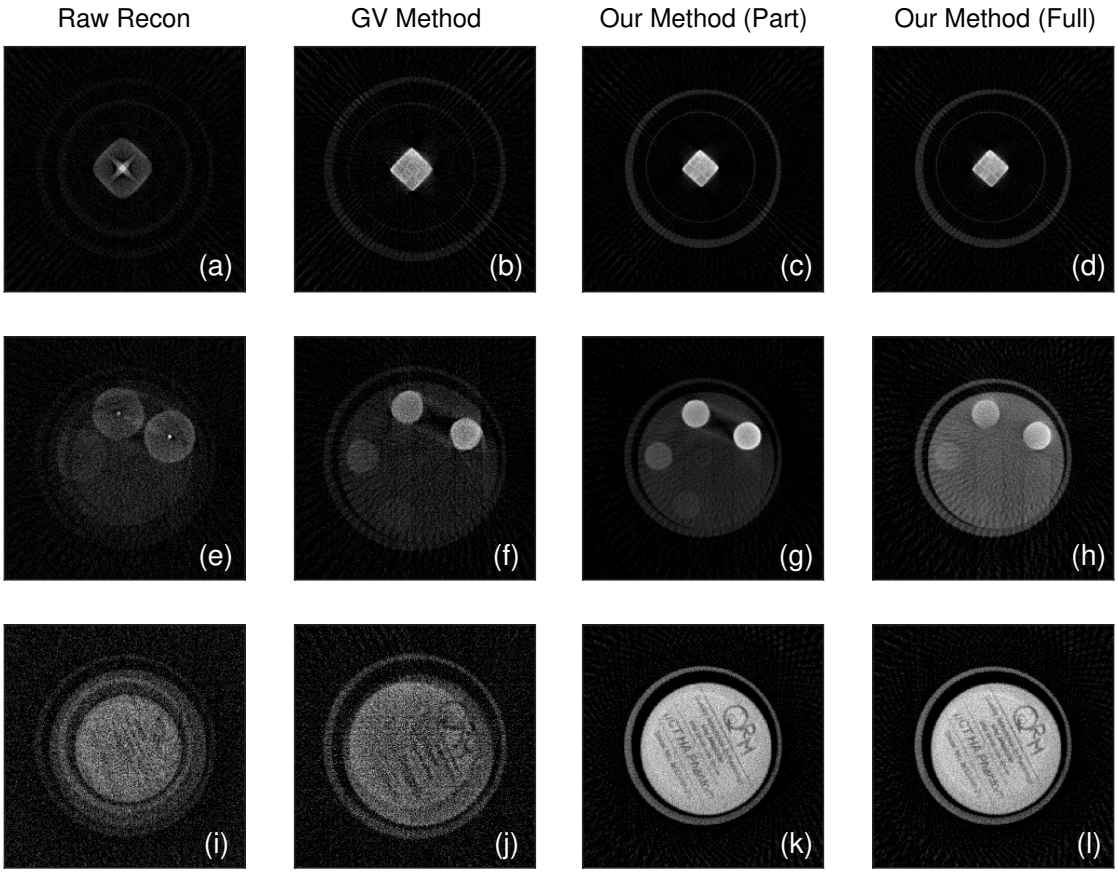


Figure 3.8: Qualitative comparison of reconstruction results on different projection calibration cases. The first column (a, e, and i) is FDK reconstruction on scanned projections without any calibration, noted as Raw Recon. The second column (b, f, and j) follows the method proposed by Dong *et al.*[33] modifying the geometry and then performs fan-FBP reconstruction, noted as GV Method. The third column (c, g, and k) refers to the FDK on projections calibrated by the proposed ST-TSO method but with only two dominant parameters fixed,[23] vertical offset and in-plane rotation angle, noted as Our Method (Part). The last column (d, h, and l) is the FDK on projections where all parameters are fully corrected, noted as Our Method (Full). Specifically, (a) - (d) are CT slices of QRM phantom, (e) - (h) are CT slices of GmbH phantom and (i) - (l) are texture slices.

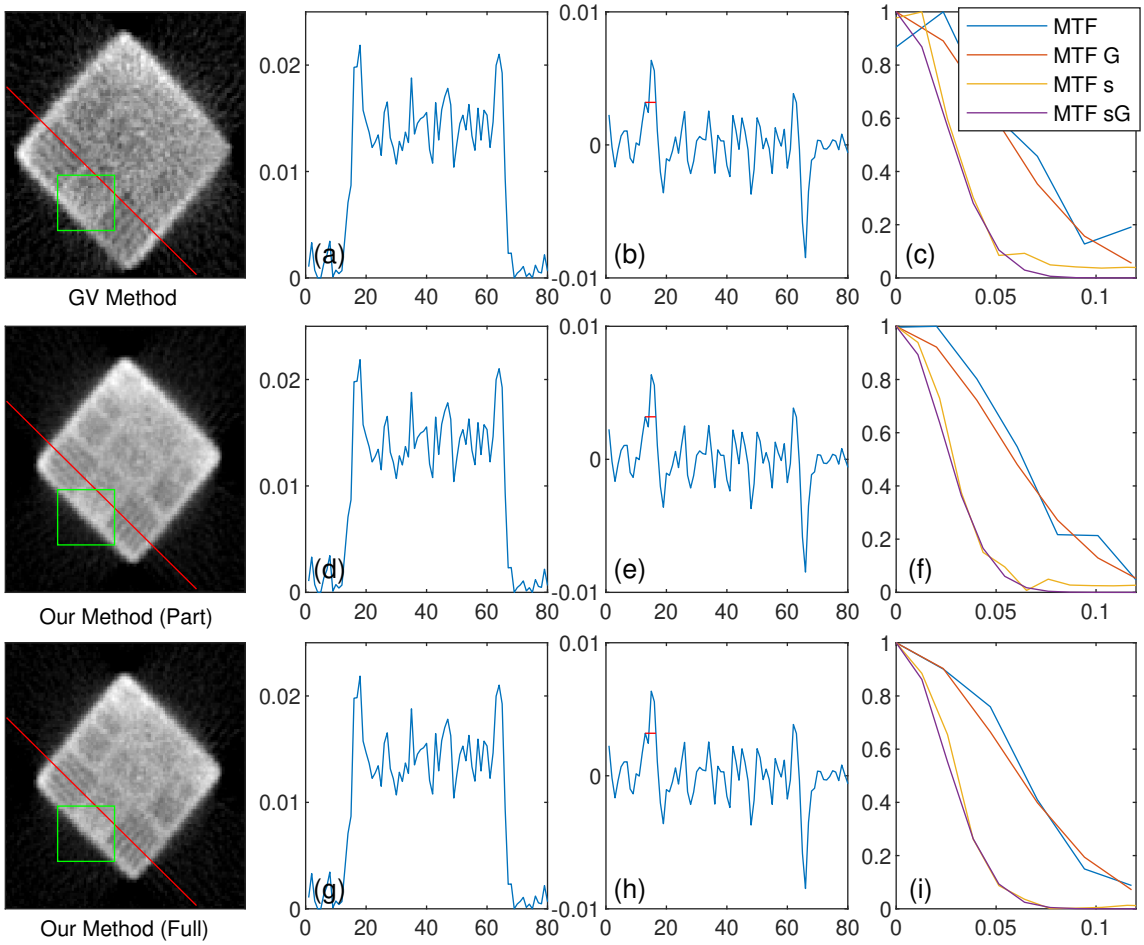


Figure 3.9: Quantitative comparison of silicon chip slice in Fig. 3.8. The first row, a zoom-in of Fig. 3.8(b), noted as GV Method. The second row, a zoom-in of Fig. 3.8(c), noted as Our Method (Part). The third row, a zoom-in of Fig. 3.8(d), noted as Our Method (Full). Two quantitative assessment methods, FWHM and MTF, were applied. (a), (d) and (g) are line profiles of the red line across the three slices, where the horizontal axis indicates the pixel index and the vertical axis indicates intensity. (b), (e) and (h) are corresponding gradient curves and the red cross-line region are their FWHM values. (c), (f) and (i) are normalized MTFs applied on the green rectangle region in the slices, where the horizontal axis is cycles/mm. In the legend, MTF is the standard MTF function, MTF G is MTF on Gaussian approximation of line spread function (LSF), MTF s indicates the smoothed MTF function, and MTF sG indicates the MTF on smoothed Gaussian approximation of LSF.

3.3.4 Ex Vivo Mice Volume Rendering

To evaluate our strategy in a practical situation, the CSTO method is applied on a mouse scanning dataset for a test. Fig. 3.10 shows an example of the 3D volume rendered from a euthanized mouse. The acquisition and reconstruction approach is the same as that of the CT phantom. It provides a unique opportunity to examine the bone structure of a biological sample.

A complete view of the rendered 3D volume is shown in Fig. 3.10(A) using 3D Slicer [53], and selected 2D slices are presented in Fig. 3.10(B). Much like our phantom data, slices reconstructed after the ST-TSO method calibration are superior and possess better spatial resolution in all regions compared with the reference GV method. Despite these improvements, it is difficult to judge which method is better without a reference image, since neither shows obvious structural errors in the 2D case. However, in 3D volume rendering, the GV-method result shows some defects in the overall structure; that is, rough bone structures and burrs in Fig. 3.10(A-i), indicating that the reference method has some flaws in accuracy. 3D volume rendering tests the performance of the calibration algorithm in the axial direction. A smooth and continuous 3D volume profile, Fig. 3.10(A-ii), confirms the accuracy of the proposed calibration algorithm in practice. In addition, the 3D volume also provides another perspective for us to observe some experimental phenomena. For example, in Fig. 3.10(C), we can clearly observe the spinal injury of the mouse through the volume data. The process was repeated for three additional animals to demonstrate the reliability of our methodology.

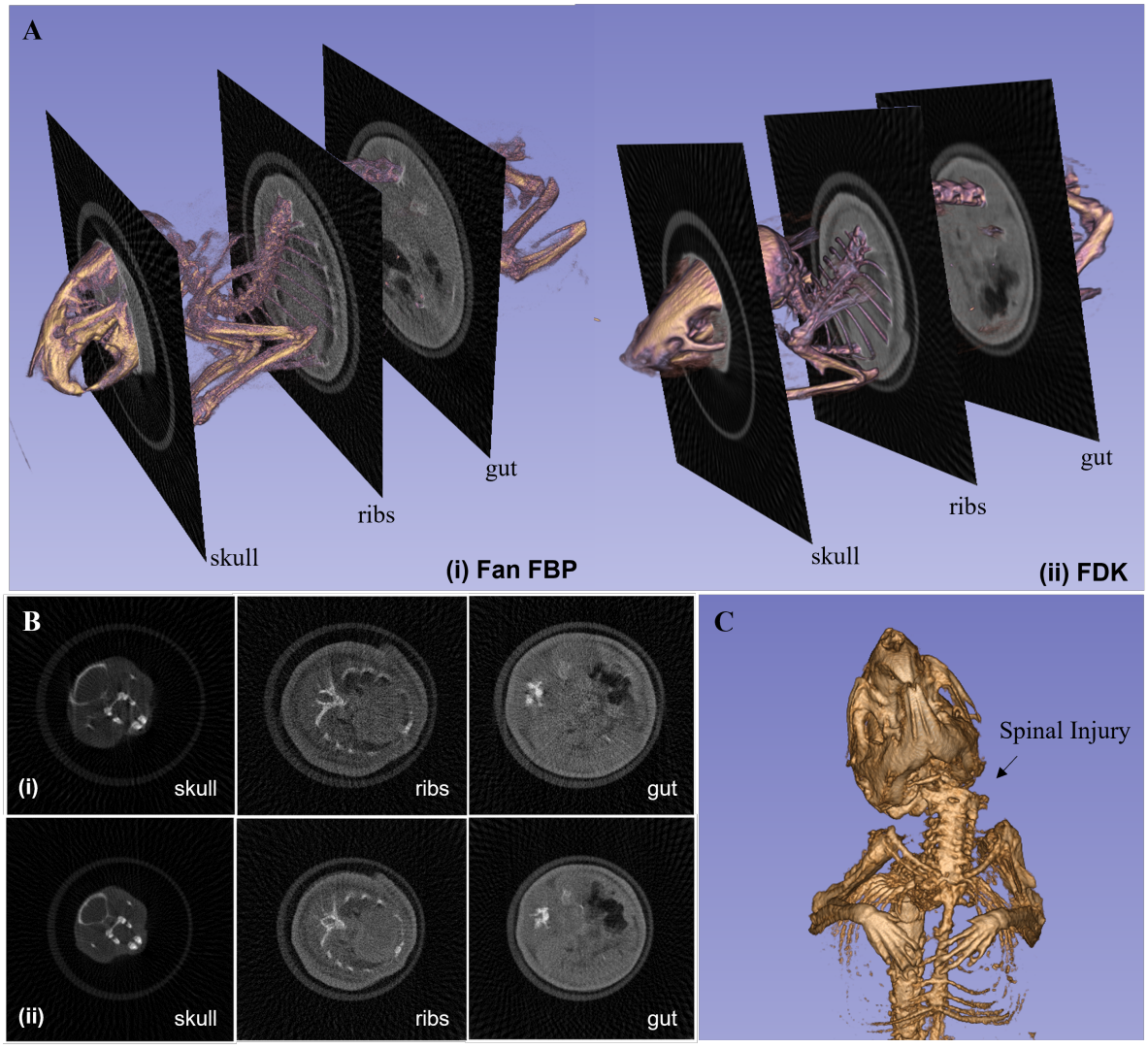


Figure 3.10: Comparison of pseudo-CT images generated from same mouse data after two calibration strategies, the GV method with fan FBP reconstruction and the ST-TSO method with FDK reconstruction. (A) 3D volume rendered in 3D Slicer given reconstructed data, (i) is the result from the GV method, and (ii) is from the ST-STO method. (B) A presentation of selected slices in (A) corresponding to the skull, ribs, and intestine regions. (C) An example of applying 3D volume rendering in spinal injury detection

3.3.5 Two-Step Processing Necessity Analysis

Although experiments verified the effectiveness of the ST-TSO method, the complete framework reached a complexity level that raised the question of whether each step actually makes a difference. For this reason, we carried out a study where we analyzed the impact of each step in isolation.

In this section, both simulation and experiment were performed to verify the necessity of two-step processing. As we stated before, cost function (Eq. (3.18)) in first-step processing may drop to a local minimum. In Fig. 3.11, we simulated three cases and each plot represents the cost function value distribution in relation to angle parameters. It is hard to depict Eq. (3.18), since it is a function of four variables. To better visualize the cost function, in Fig. 3.11, we set the two variables m and n constant and the angle variables vary from -10 degree to 10. We can observe that in Fig. 3.11(a) and (b), plots show a convex property along the η side and appear smooth at the bottom region along the θ direction. In Fig. 3.11(c), there is a ridge along the θ direction at the bottom where a local minimum exists on both sides. Therefore, it is reasonable to add one more step in our proposed method to strengthen its accuracy and robustness.

In addition, a practical experimental result also supports our statement. Fig. 3.12 presents a comparison of two processing approaches, the first-step-only ST-TSO method and the complete two-step ST-TSO method. For the CT phantom case, there is no obvious difference between two sagittal views. However, for mouse data, we could observe that the reconstructed volume with the first method (Fig. 3.12(b)) appears blurry, while the second one (Fig. 3.12(d)) maintains greater resolution. Moreover, we can still observe that a mismatch structure error exists in Fig. 3.12(b), indicating that the first-step-only ST-TSO method cannot guarantee the accuracy. In Fig. 3.12(d), the two-step ST-TSO method successfully fixes the problem.

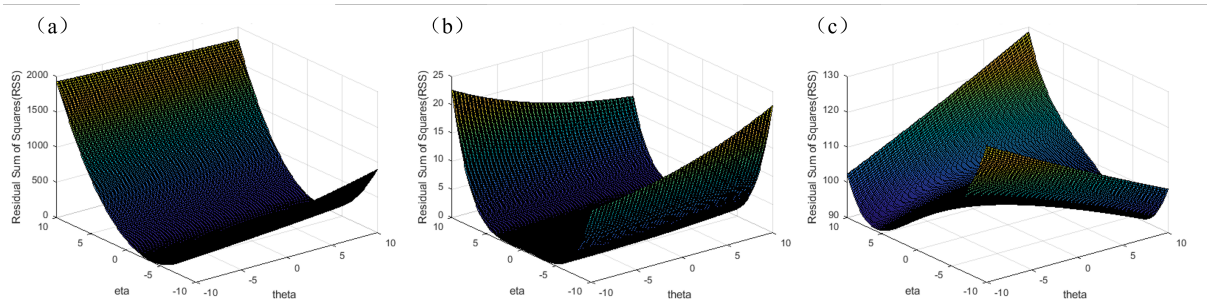


Figure 3.11: Digital phantom simulation of the cost function value distribution in relation to rotation angles. For each case, ground truth value of θ, η, m, and n are pre-determined, which means projection image information are fixed as well. Distribution plots describe the Eq. (3.18) values when θ, η vary from -10 to 10. (a) is case θ = -3, η = -3, m = -100, n = -100, (b) is case θ = -3, η = -3, m = 0, n = -100 and (c) is case θ = -3, η = -3, m = -100, n = 0.

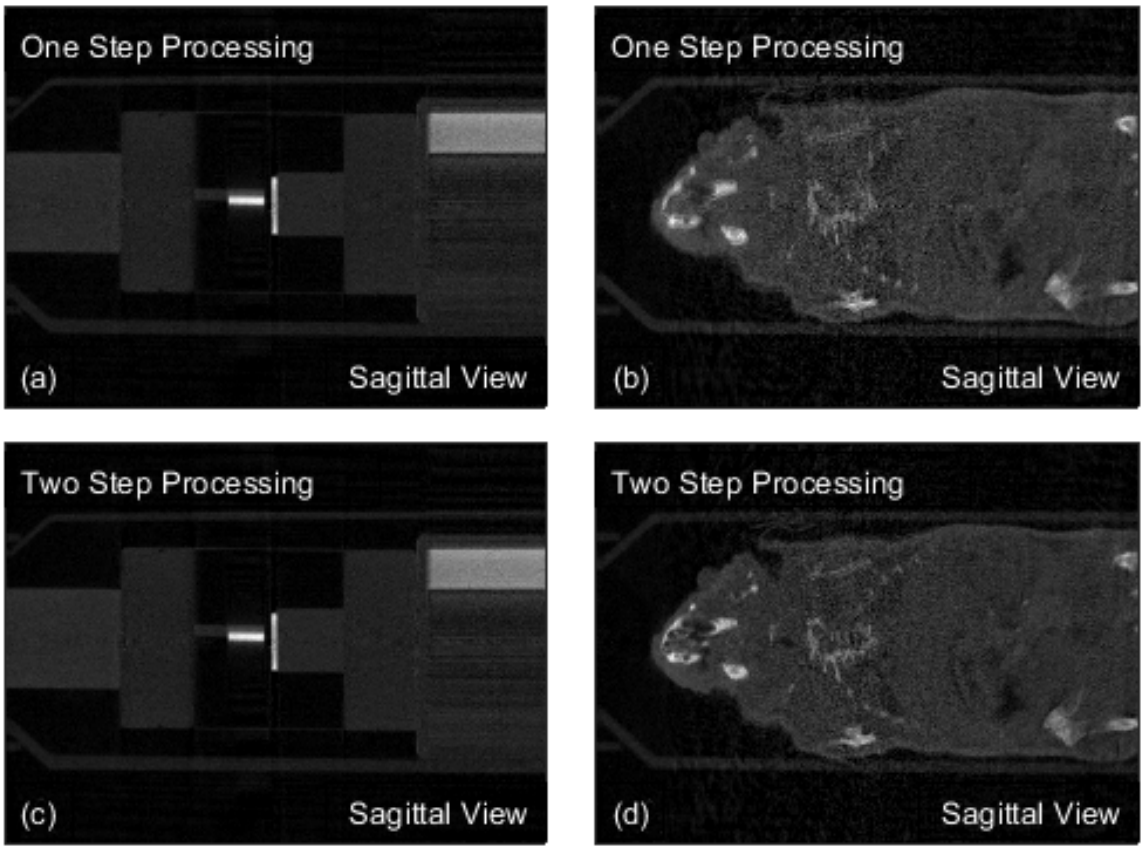


Figure 3.12: Demonstration of two-step processing necessity in geometry correction. Images (a) and (b) are sagittal views of reconstructed 3D volume with a first-step only ST-TSO calibration algorithm, and images (c) and (d) are volume data reconstructed after a complete two-step ST-TSO method calibration

.

3.3.6 EAS algorithm analysis

Our EAS algorithm is an extension from the classic heuristic SA algorithm that aims to accelerate the second-step iteration. For each loop, the classic SA algorithm randomly selected a few candidate points within the search range for global minimum value updating. Instead, in the EAS algorithm, we replaced the random selection with the controlled selection, Nelder-Mead algorithm based search strategy, to accelerate the convergence and avoid resource waste. A comparison of the EAS algorithm and the conventional SA algorithm is shown in Table 3.3. Both of them initialize with the same parameter, and results have shown that the convergence of the EAS algorithm to similar SOP Err is four times faster than the SA method. This result indicates that our modified EAS algorithm can successfully accelerate the algorithm.

Table 3.3: Comparison of the EAS algorithm and the SA algorithm

	SA algorithm				EAS algorithm			
	tilt(θ)	slant(η)	m	n	tilt(θ)	slant(η)	m	n
start point	-2.13668	-0.08348	-77.3422	0	-2.13668	-0.08348	-77.3422	0
end point	-0.92153	0.125145	-76.3399	-0.7103	-1.2408	0.062757	-76.3598	1.044854
SOP Err	14.0791				14.0341			
iterations	110(20 tests for each loop)				200			
time	15354s				3689s			

3.4 Conclusion and Discussion

In this chapter, we have proposed a novel structure-tensor-based two-step (ST-TSO) online calibration method for our CBCT imaging prototype design. Four parameters can be estimated simultaneously: the offset perpendicular to the rotation axis (vertical offset), the offset along the rotation axis (lateral offset), the tilt angle, and the skew angle. Our first-step processing extends the traditional projection matrix (PM) method to line conditions. Combined with image structure tensor information, it is able to make an initial prediction of geometric parameters quickly. Our second-step processing fine-tunes the parameters iteratively by fully utilizing projection image intensities.

In the past, an intuitively appealing approach to calibrating the geometric parameters online was to utilize the reconstructed images. Cost functions on image entropy, sharpness, and global variance have been well designed in the past decades. Unfortunately, a general iterative calibration process includes continuous reprojection and reconstruction, which is time-consuming in 3D volume rendering. Instead, our fine-tuning method that involves iterations only computes the resampling and the cost function in the projection image domain, which can greatly shorten the optimization time and reduce the memory requirements in the iteration.

Our proposed ST-TSO method based on the actual situation consists of two-step processing and an additional condition, which creates some complexity. To justify this added complexity, subsequent digital phantom simulation was performed to verify the necessity of each step in the method—for instance, the addition of a structure tensor in assisting parameter prediction. Extended line conditions have provided sufficient information in predicting unknowns. However, the edge-parallel nature of detection leads to the direct PM method showing some

drawbacks in practice. As a solution, the addition of a structure tensor has shown tremendous performance improvement in simulation. Moreover, a test on actual data verifies that the two-step processing framework can ensure that the reconstruction results are closer to the real situation, which improves the robustness of the algorithm. Furthermore, our simplified version of ST-TSO method in estimating one or two variables is attached in Appendix B

In small-animal imaging, the acquisition of endpoint anatomical data has the potential to provide valuable insights that could allow for increased scientific rigor in small-animal imaging studies and thereby reduce overall animal usage. The add-on device, the *MiSpinner*, was initially designed for bioluminescence imaging. It can be placed within popular small-animal imaging systems to aid in 360° image collection. We showed here that minor alterations in the way *MiSpinner* data are processed can be used to provide CT reconstructions. By exploiting the characteristics of the *MiSpinner* insert, our analysis has revealed that distortion of boundaries can reflect geometric errors in the actual scanning. This finding helps us avoid the need for dedicated phantoms in system calibration and applies the PM method here.

With the help of edge detection techniques, the performance in the CT phantom test has verified that our method outperforms the reference method in correcting geometry errors. The reconstructed image structure has no obvious deformation, and the contrast is well maintained. To show that this is a reproducible result, we imaged three additional animal carcasses through the acquisition of 96 intervals and applied the same method. Our approach yields similar high-quality results. A limitation is the threshold variability between samples and our lack of electron density reference standards for any given scan. Having these references will allow us to use this approach to interrogate, for instance, changes in bone density seen between animal cohorts[54].

Chapter 4

Structure Tensor Total Directional Variation Based Projection Image Denoising and Synthesis

Projections from our retrofit system generally suffer from low-dose noise effects and sparse sampling in the angular domain, leading to streak artifacts in the CT reconstruction. This chapter proposes a novel method for noise removal and projection synthesis. It is based on an adversarial based semi-supervised learning network, named Proj-SynNet, wherein the detected low-dose noisy projections are used for network training.

4.1 Introduction

Two main types of artifacts arise in the retrofit imaging system, low-dose radiation noise [55] and sparse sampling artifacts [56]. Low-dose radiation—1-3 mGy average dosage in our system compared with typical 100 mGy dosage in pre-clinical *in vivo* CT imaging [57]—inevitably causes photon fluctuation from the mean during propagation [58]. Furthermore, use of a scintillator that converts X-rays to visible light leads to photon scattering in pre-clinical imaging, which degrades the projection image resolution [59] and increases the electronic background noise [58, 60]. Likewise, sparse sampling, e.g., taking projections every 3.75 degrees in a full scan, introduces streak artifacts into the reconstructed image [56].

To improve the reconstructed image quality, numerous advanced methods have been proposed [58, 61–71]. Typical solutions are model-based algorithms that simulate the photon propagation according to statistical properties, e.g., the penalized weighted least-square (PWLS) model [61–63, 66] based on a Gaussian noise assumption and the Rudin-Osher-Fatemi (ROF)

model [69, 72, 73] based on a Poisson noise distribution or comprehensive model [58] that approximates the photon fluctuation as a Poisson distribution and electronic background noise as Gaussian distribution. Generally, there are three ways to model the problem:

1. image-based iterative reconstruction, which yields the desired CT image from the post-log sinogram data by iterative optimization. This approach constructs the forward model according to the scanning geometry and usually incorporates a CT image prior distribution term to improve algorithm performance, e.g., using the Markov random field (MRF) hypothesis which is characterized by the Gibbs distribution [66, 74] or total variation (TV) or its variants based on compressed sensing (CS) theory [64, 65, 68].

2. projection-based pre-processing method, which aims to recover the pre-log projection data from a noisy measurement and then reconstructs the image [58, 70, 71]. This approach constructs the forward model according to the Beer-Lambert law and combines with the projection image intrinsic characteristics such as edge-preserving Huber penalty [75] or sinogram piece-wise linear assumption [58] as a penalty term.

3. image-based post-processing method, which directly generates the desired CT image from its noisy reconstruction [69, 73, 74]. This approach originates from an image denosing perspective that regards the noisy image as being obtained by corrupting a noiseless one. Moreover, some post-processing methods directly apply denoising algorithms, e.g., BM3D [76] and non-local means [77].

Despite the promise of these methods, questions still exist in practical applications:

1. For the model-based iterative method, the forward model design is normally assumed to be a standard geometry while in practical scanning, acquired projection data often possess geometric errors.

2. For projection-based pre-processing method, the model alone cannot solve the sparse-sampling problem.

3. For image-based post-processing method, it is hard to handle complex textures due to insufficient prior knowledge about the reconstructed image.

Other than those model-driven methods, emerging data-driven methods, such as deep learning (DL) neural networks, have also gained increasing attention in CT image reconstruction [78–90]. Application of DL has covered every step in CT imaging, including preprocessing [88–90], reconstruction [85–87] and post-processing [82–84]. These proposed learning algorithms are typically supervised, requiring pairs of low-dose or sparse reconstructed CT images and the corresponding high-dose ground truth images for network training. The key to the model success lies in the use of sufficient training data. In clinical research, the acquisition of ground truth data may be accessed through professional equipment and a large number of scans. However, in pre-clinial research, ground truth image insufficiency is an inescapable problem, especially in retrofit studies of existing imaging systems. Instead, a large number of unlabeled low-dose noisy projections are easier to obtain.

Therefore, it is critical to utilize the unlabeled low-dose projections for DL training, which applies to unsupervised or semi-supervised learning. The unsupervised DL methods fully utilize the unlabeled data for data-driven decision-making, while the semi-supervised methods approximate the deep transfer learning using a few supervised image pairs and massive unlabeled data for network training. Recently, Liang *et al.* proposed a model-based unsupervised DL network that consists of a projection block, a domain transform block, and an image block for low-dose CT denoising [91]. Zeng *et al.* proposed a noise-generating-mechanism(NGM)-driven unsupervised DL network for sinogram data restoration [92]. Meng *et al.* presented a semi-supervised learned sinogram restoration network (SLSR-Net) for low-dose CT reconstruction [93]. Although many DL regimes have been designed and verified impressive performance in image quality enhancement, few of them focus on unsupervised or semi-supervised DL networks in solving sparse sampling problem.

In the past, sparse sampling was resolved by sinogram interpolation in the 2D case. Esmaeil *et al.* interpolated the missing data along the angle direction with cubic-spline interpolation [94]. Zeng *et al.* proposed a displacement function interpolation method which aims to simulate intensity variation along the angle direction [95]. Li *et al.* decomposed the sinogram value into a series of sinusoid-like curves and interpolated the data by estimating the intensity

change through tensor matrix eigenvector decomposition [96]. Lately, in projection domain, algorithms that aim to maintain a data-spatial relationship, i.e., inter-slice consistency, have also been proposed through synthesizing intermediate projections to increase view angles [97, 98].

In this research, inspired by the development of video frame synthesis in DL [99], we present a structure tensor (ST) based systematic projection processing framework that aims to solve the low-dose sparse sampling problem, which consists of two processing blocks, a model-based denoising block and a semi-supervised image synthesis block. Specifically, in the model-based denoising block, we consider the noise generation statistical model under maximum a posterior probability (MAP) framework with a structure tensor based total directional variation (ST-TDV) term. The regularization term, which approximates the directional TV [100, 101], is able to reduce the stair artifact caused by the standard TV kernels [48, 102]. In the semi-supervised image synthesis block, we synthesize extra high-quality projections with an adversarial based convolutional neural network (CNN). Different from the unsupervised/self-supervised DL that treats the denoised image as ground truth data, in this work, we believe access to a small number of high-dose images, which refers to the pre-clinical CT scans from clinical veterinary scanner, can achieve better performance in network training. Papyan *et al.* has shown that the forward pass of the CNN is tightly connected to the multi-layer convolutional sparse coding (ML-CSC), which guarantees uniqueness and sparsity of the high latent feature maps [103]. Thus U-net structure CNN is applied to capture the image geometry feature and then transferred into a sampler generator for intermediate projection image synthesis. Finally, the synthesized images are added to the original projection image sequence for CT image reconstruction.

To the best of our knowledge, this is the first attempt to use semi-supervised DL to solve the low-dose sparse sampling problem. Compared with traditional directional interpolation algorithms, a well-trained network can significantly reduce the computation time for image synthesis. Numerical simulation has shown that the proposed network can yield competitive results, and tests on real data have shown its ability to improve the image quality.

4.2 Iterative Reconstruction Technique

Apart from the FBP/FDK algorithm, iterative reconstruction (IR) is an alternative solution to obtain a CT image. Generally, the forward Radon transform can be described as

$$g = Hf + N \tag{4.1}$$

where f stands for the integral of attenuation coefficients along the X-ray path inferred from the observed projection data, H is a weighted matrix representing the forward Radon transform, f is a vector of attenuation coefficients representing 2D images re-arranged into a 1D vector and N stands for the noise occurred during measurement. Moreover, in this research, we ignore the variability of the image object f and consider it to be deterministic.

Incomplete projections, or sparse sampling, will lead to the matrix H being ill-conditioned due to the matrix H having more columns than rows. For instance, given the size of H is $m \times n$, m may be far smaller than n. Moreover, note that the additive noise N does not limit our analysis, since it refers to the deviation of an instance of projection data from its expectation. Thus, to solve this problem, we use the iteration technique to transform the equations into an optimization problem and find the solution.

Algebraic Reconstruction Technique (ART) is one of the early iteration methods applied in CT to restore the image. Its basic idea is to consider each element of projection data g_i as a projection of an n-variable into an m-length hyperplane. Then project the initial guess of the variables onto different projections, and update the difference by iteration until it meets certain condition. The algorithm is based on the Kaczmarz method [104] for solving the linear system $g = Hf$. Its accuracy is highly affected by the number of projections.

Compared with FBP that relies on idealized mathematical models, ART-based algorithms model the acquisition geometry numerically, which better represents the actual situation. Later

developed SART and SIRT algorithms accelerate the iteration by simultaneous processing using a *projection-based iteration* [105]. Furthermore, the geometry model may suffer from non-idealities, e.g., system blur, geometry error, etc. Iterative methods incorporating these models are often called model-based iterative reconstruction (MBIR).

4.2.1 Model-Based Image Reconstruction

Generally, MBIR originates from a MAP approach based on Bayes Theorem [106], i.e.,

$$\tilde{f} = \arg\max_{f} P(f|g) = \arg\max_{f} P(g|f)P(f), \tag{4.2}$$

where $P(f|g)$ is the posterior, $P(g|f)$ is the likelihood function, and $P(f)$ is the image prior. The forward model and noise model are described by the likelihood function in statistics, called the data fidelity term. And the image prior is denoted as the regularization term. For simplicity, in this research, we assume X-rays are monochromatic. To ensure that the symbols are consistent in context, we change the projection data f to Y and measured data g to P as in the previous section.

4.2.2 Data Fidelity Term

Previous analysis of repeated measurements from the sample phantom has revealed that low-dose projection data follows approximately a Gaussian distribution [61, 66], whose variance is associated with the data sample mean, expressed as

$$\sigma_i^2 = \frac{1}{I_0} e^{\bar{q}_i} [1 + \frac{1}{I_0} e^{\bar{q}_i} (\sigma_e^2 - 1.25)], \tag{4.3}$$

where $\bar{q}_i$ is the mean, σ_i^2 is the variance of the repeated measurements at detector bin i, and σ_e^2 is the variance of the background noise. Further analysis also indicates that there is no statistical correlation between the noise and the projection views [61]. These observation with the MAP

framework lead to the well known weighted least squares (WLS) model [66], which is

$$P(P|Y) = \prod_{i=1}^{N} p(P_i, Y_i) = \prod_i \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(P_i - Y_i)^2}{2\sigma_i^2}}, \tag{4.4}$$

and its negative logarithm is

$$\phi(Y) = \frac{1}{2}(P - Y)^T \Sigma^{-1}(P - Y) + \frac{1}{2}\ln(2\pi\Sigma), \tag{4.5}$$

or

$$\phi(Y) = \frac{1}{2}(P - Hx)^T \Sigma^{-1}(P - Hx) + \frac{1}{2}\ln(2\pi\Sigma), \tag{4.6}$$

where $\phi(Y)$ denotes the cost function and the covariance matrix Σ is diagonal with the *i*-th entry σ_i^2.

Although WLS possesses advantage in computation, it fails to accurately use the Poisson distribution characteristic in phantom transmission. In this research, our noise generation model incorporates both Poisson noise and Gaussian noise as [61, 67]

$$I = Poisson(I_0 e^{-Y_i}) + Gauss(0, \sigma^2), \tag{4.7}$$

where I_0 represents incident X-ray source intensity, Y_i is the line integral of phantom volumes corresponding to pixel i on the projection image, and σ^2 is the background electronic noise variance. Specifically, a full expression of the noise generation statistical model is

$$for\, each\, bin\, i, \begin{cases} Poisson\, Q_i \sim P(Q_i|Y_i) = \frac{Y_i^{Q_i}}{Q_i!} e^{-Y_i} \\ Gaussian\, \varepsilon_i \sim N(0, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\varepsilon_i^2}{2\sigma^2}} \end{cases}, \tag{4.8}$$

where Q is intermediate-state projection data corrupted by Poisson noise. Then the measured data P is

$$P(P|Y) = P(P|Q)P(Q|Y) = \prod_{n=1}^{N} \frac{(I_0 e^{-Y_i})^{Q_i}}{Q_i!} e^{-I_0 e^{-Y_i}} \times \frac{1}{(\sqrt{2\pi\sigma^2})^N} e^{-\frac{\|P-Q\|_2^2}{2\sigma^2}} \tag{4.9}$$

4.2.3 Regularization/Penalty Term

In image processing, to further improving the IR accuracy, a regularization term, e.g., Tikhonov operator [107], is introduced to refine the image texture. A standard form in IR with a regularization term is expressed as

$$f_\lambda = \arg\min_f \{\|Hf - g\|_2^2 + \lambda \|Lf\|_2^2\} \tag{4.10}$$

where λ is a weighting scalar and Lf is the regularization term, and f and g are defined as in Eq. (4.1).

Similarly, in MBIR, we incorporate the image prior as a regularization term. For instance, in the PWLS model, an MRF is used to model the sinogram prior distribution, i.e.,

$$P(Y) = \frac{1}{Z} e^{-\frac{\alpha}{2} Y^T M Y}, \tag{4.11}$$

where Z is a constant, α is a weighting parameter, and M is an $N \times N$ square matrix possessing a highpass property. Then combining Eq. (4.5) and Eq. (4.11), we got corresponding PWLS model; that is,

$$\phi(Y) = \frac{1}{2}(P - Y)^T \Sigma^{-1} (P - Y) + \frac{1}{2} \ln(2\pi\Sigma) + \frac{1}{2} Y^T M Y, \tag{4.12}$$

and

$$\tilde{Y} = \arg\min_{Y \geqslant 0} \phi(Y), \tag{4.13}$$

where $\tilde{Y}$ is the desired sinogram data.

The image prior is crucial to successful recovery of the CT sinogram data. By utilizing prior information, MBIR methods have shown to be superior in noise-induced artifact suppression as compared to the traditional FBP method. However, minimization of the MRF penalty term, which approximates a quadratic l_2-norm, will lead to too much spatial regularization (smoothing) compared to the original signal. Besides, in the sinogram domain, quadratic design in Eq. (4.12) ignores the intrinsic image properties, i.e., the different dimensions represent

different physical or geometric meanings, which leads to certain kinds of blur in reconstructions [61, 70]. Hence, a non-quadratic penalty term is applied to preserve the high-frequency components, e.g, total variation (TV) kernel.

CS theory has proved that accurate reconstruction of sparse signals can be realized under certain constraint conditions, signal sparsity and restricted isometry property (RIP), or incoherent sampling [108, 109]. In CT image reconstruction, considering that attenuation occurring in soft tissue is similar to a piecewise constant, its gradient image meets the sparsity condition and its limited projections at different angles meet the incoherent sampling condition in CS theory. Therefore, TV minimization can be used as a penalty term in CT iterative reconstruction. A common cost function can be described as [110]

$$\tilde{f}_\alpha = \arg\min\{\|Hf - g\|_2^2 + \alpha \|f\|_{TV}\}, \tag{4.14}$$

where

$$\|f_{s,t}\|_{TV} = \sum_{s,t} \sqrt{(f_{s,t} - f_{s-1,t})^2 + (f_{s,t} - f_{s,t-1})^2}, \tag{4.15}$$

f, g, H definition are aligned with Eq. (4.1). TV penalty uses the l_1-norm to enforce the sparsity constraint, where the formulation is also known as the basis pursuit denoising (BPD) problem or the least absolute shrinkage and selection operator (LASSO) problem in other fields. By assuming those projections (hyperplanes) are convex sets, the projection onto convex sets (POCS) method, which aims to find the solution through moving f toward the intersection of all hyperplanes, is applied [68, 110, 111]. Another simpler realization is to apply the steepest decent algorithm with Huber approximation on TV term, which is

$$f^{n+1} = f^n + \lambda[H^T(g - Hf) - \alpha(\bigtriangledown \|f\|_{TV})], \tag{4.16}$$

where

$$\|f\|_{TV} \approx \varphi_\tau(x) = \begin{cases} |x| - \frac{T}{2} & |x| > T \\ \frac{1}{2\tau}x^2 & |x| \leqslant T \end{cases}. \tag{4.17}$$

The approximation guarantees the existence of a gradient within the domain.

Although TV has shown good performance in edge preservation, in soft tissue regions the TV term tends to introduce extra artifacts; that is, the region is segmented due to the existence of small steps around the low-contrast part, called staircase artifacts [112, 113]. To solve this problem, the subgradient must be defined properly using a higher-degree total variation (HDTV), such as using the total generalized variation (TGV) kernel [67, 114]. Some algorithms use spatially-dependent edge and texture orientations in the form of total directional variation (TDV) to preserve the edge information [100, 101, 115].

Total Generalized Variation (TGV): In order to eliminate the staircase artifacts from TV-based algorithms, Bredies *et al.* proposed the TGV term, whose definition is [114]

$$TGV_{\alpha}^{k}(u) = \sup\{\int_{\Omega} u\text{div}^{k} v dx | v \in C_{c}^{k}(\Omega, Sym^{k}(R^{d})), \left\|\text{div}^{l} v\right\|_{\infty} \leqslant \alpha_{l}\}, \tag{4.18}$$

where $l = 1, 2, 3, ..., k-1$ and k is an order of TGV, $\alpha = (\alpha_0, \alpha_1, ..., \alpha_k - 1)$ represents the positive weights for TGV. Its primal form is

$$TGV_{\alpha}^{k}(u) = \inf_{u_l \in C_c^{k-1}} \sum_{l=1}^{k} \alpha_{k-l} \left\|\varepsilon(u_{l-1}) - u_l\right\|_1, \ \varepsilon(u_{l-1}) = \frac{1}{2}(\nabla u_{l-1} + \nabla u_{l-1}^{T}) \tag{4.19}$$

When $k = 1$, TGV coincides with the TV term. When $k = 2$, it becomes

$$TGV_{\alpha}^{2}(u) = sup\{\int u\text{div}^{2} v dx | v \in C_{c}^{2}(\Omega, S^{d\times d}), \left\|v\right\|_{\infty} \leqslant \alpha_0, \left\|\text{div} v\right\|_{\infty} \leqslant \alpha_1\}, \tag{4.20}$$

where

$$\begin{aligned} \left\|v\right\|_{\infty} &= sup(\sum_{i=1}^{d} |v_{ii}(x)|^2 + 2\sum_{i<j} |v_{ij}(x)|^2)^{1/2} \\ \left\|\text{div}\, v\right\|_{\infty} &= sup(\sum_{i=1}^{d} \left|\sum_{j=1}^{d} \frac{\partial v_{ij}}{\partial x_j}\right|^2)^{1/2}. \end{aligned} \tag{4.21}$$

Likewise, its second-order primal form can be written as

$$TGV_{\alpha}^{2}(u) = \min_{w} \alpha_1 \int_{\Omega} |\nabla u - w| \, dx + \alpha_0 \int_{\Omega} |\varepsilon(w)| \, dx. \tag{4.22}$$

Here, $\varepsilon(w) = \frac{1}{2}(\nabla w + \nabla w^T)$, x is the index, u is the image, v is the dual variable of the TV term, and $\varepsilon(w)$ is a matrix-valued Radon measure. The second-order TGV provides a way to balance the first- and second-derivative terms via the weights α_0 and α_1. Specifically, through the second-order TGV definition, we can observe that in smooth regions, where $w \approx \nabla u$, $\min TGV_\alpha^2(u) \approx \min \|\nabla^2 u\|$; in edge regions, where $\|\nabla^2 u\| > \|\nabla u\|$, $\min TGV_\alpha^2(u) \approx \min \|\nabla u\|$.

A classic application of the TGV term in MBIR methods is the PWLS-TGV algorithm, which is

$$\phi(f) = \arg\min_{f \geqslant 0}(g - Hf)^T \Sigma^{-1}(g - Hf) + \beta TGV_\alpha^2(f). \tag{4.23}$$

By introducing the auxiliary variable μ, we can reformulate the function as

$$\begin{aligned} &\arg\min_{f,\mu}(g - H\mu)^T \Sigma^{-1}(g - H\mu) + \beta TGV_\alpha^2(f) \\ &s.t. \mu = f, \end{aligned} \tag{4.24}$$

and rewrite in Augmented Euler-Lagrange form [116]:

$$L(\mu, f, \lambda) = \arg\min_{f,\mu,\lambda}(g - H\mu)^T \Sigma^{-1}(g - H\mu) + \lambda \|\mu - f\|^2 + \beta TGV_\alpha^2(f). \tag{4.25}$$

This problem can be decomposed into two alternating optimization steps, P1 for raw estimation and P2 for accurate optimization:

$$\begin{aligned} (P1) &: \mu = \arg\min_{\mu}(g - H\mu)^T \Sigma^{-1}(g - H\mu) + \lambda \|\mu - f\|^2 \\ (P2) &: f = \arg\min_{f} \lambda \|\mu - f\|^2 + \beta TGV_\alpha^2(f) \end{aligned} \tag{4.26}$$

In the implementation, a separable surrogate algorithm [63] is used to solve $(P1)$, and a first order primal-dual (FOPD) method [114] is appled to solve $(P2)$.

Total Directional Variation (TDV): In functional analysis, Eq. (4.14) can be treated as a general function with a perturbation. Then we can derive its corresponding Euler-Lagrange partial

differential equation (PDE):

$$2(Hf - g) - \nabla \cdot \frac{\nabla f}{|\nabla f|^2} = 0. \tag{4.27}$$

We can find that the TV kernel is an isotropic diffusion kernel that smooths the image in all directions, which is not very suitable for images with a dominant direction, e.g., sinograms. Based on this, Bayarm *et al.* [100] proposed the TDV to increase the kernel sensitivity. Recalling the TV definition Eq. (4.15), let Δ_1 and Δ_2 denote horizontal and vertical difference:

$$\Delta_1 f(s,t) = f(s,t) - f(s-1,t),\ \Delta_2 f(s,t) = f(s,t) - f(s,t-1). \tag{4.28}$$

Then, TV can be rewritten as

$$\|f\|_{TV} = \sum_{s,t} \|\Delta f(s,t)\|_2 = \sum_{s,t} \sup_{t \in B_2} \langle \Delta f(s,t), t \rangle , \tag{4.29}$$

where B_2 is the unit ball of the l_2 norm. A directional total variation can be realized by replacing B_2 with an ellipse domain $E_{\alpha,\theta}$ that is oriented along the angle θ with unit length minor axis and a major axis of length α as shown in Fig. 4.1:

$$\|f\|_{TV} = \sum_{s,t} \sup_{t \in E_{\alpha,\theta}} \langle \Delta f(s,t), t \rangle = TV_{\alpha,\theta}(f). \tag{4.30}$$

Then Eq. (4.14) becomes

$$\tilde{f}_\alpha = \arg\min\{\|Hf - g\|_2^2 + \alpha TV_{\alpha,\theta}(f)\}. \tag{4.31}$$

When direction θ aligns with the gradient vector Δf direction, TDV reaches the minimum value. Let matrix R_θ and Λ_α denote the rotation and scaling matrices,

$$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix},\ \Lambda_\alpha = \begin{bmatrix} \alpha & 0 \\ 0 & 1 \end{bmatrix}, \tag{4.32}$$

then we obtain

$$E_{\alpha,\theta} = R_\theta \Lambda_\alpha B_2, \tag{4.33}$$

and

$$\begin{aligned} TV_{\alpha,\theta}(f) &= \sup_{t \in E_{\alpha,\theta}} \langle \Delta f, t \rangle \\ &= \sup_{v \in B_2} \langle \Delta f, R_\theta \Lambda_\alpha v \rangle . \end{aligned} \tag{4.34}$$

Similar to the Huber approximation in the TV algorithm, results can be approximately acquired through gradients by $y - \lambda \Delta^T R_\theta \Lambda_\alpha v^*$, where

$$v^* = \arg\min_{v \in B_2} \left\| y - \lambda \Delta^T R_\theta \Lambda_\alpha v \right\|_2^2 . \tag{4.35}$$

Typical numerical solutions can be estimated through the fast iterative shrinkage algorithm (FISTA) [117]. Furthermore, we find that TDV is similar to the first-order TGV but multiplied by a weighting function related to derivative orientation.

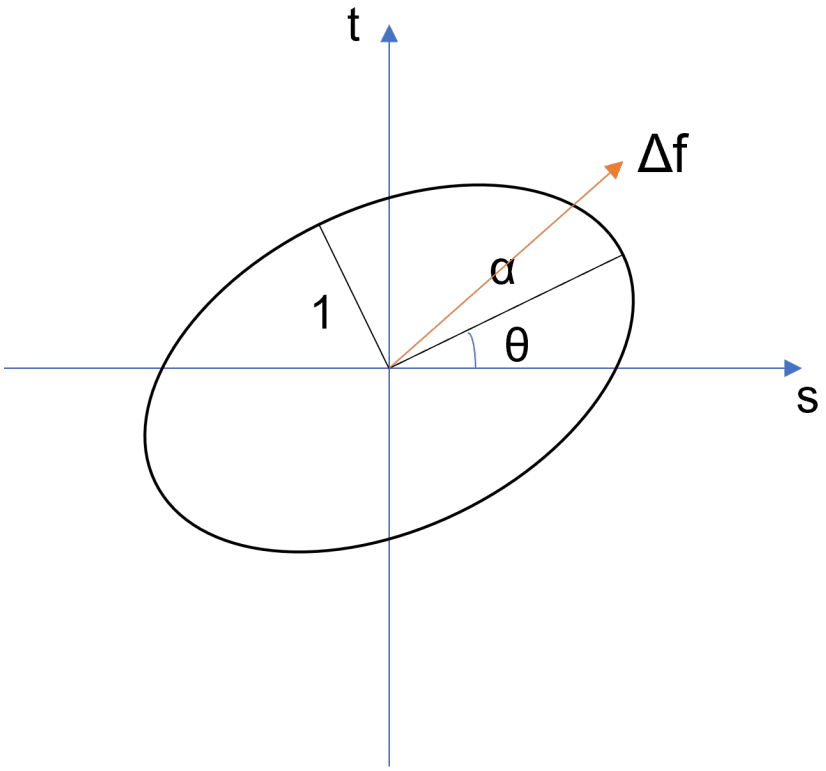


Figure 4.1: Illustration of TDV ellipse domain: Δf, which is a gradient vector, represents the dominant direction of the texture in the image, α is the ratio of the horizontal and vertical difference that larger than 1.

4.2.4 ST-TDV based MBIR Algorithm

Recalling the structure tensor definition in Section 3.2.3, we determine the image geometry variation through the eigenvector decomposition of the image structure tensor. In [48], Lefkimmiatis *et al.* proposed a structure tensor based total variation kernel, which is

$$STV(Y) = \sum_{i=1}^{N} \left\| (\sqrt{\lambda_i^+}, \sqrt{\lambda_i^-}) \right\|_1, \tag{4.36}$$

where λ_i^+ and λ_i^- are two eigenvalues of the structure tensor with eigenvector decomposition. In this research, inspired by the TDV, we propose the ST-TDV, which utilizes the corresponding eigenvectors that define the maximum and minimum variation orientation, as image prior. To better demonstrate the ST-TDV, we provide the following definition:

Definition 4.1 (Weighted Directional Derivative)**.** Let $\nabla = (\partial_1, \partial_2)^T$ be derivative, then $\nabla u = \nabla \otimes u = (\partial_1 u, \partial_2 u)^T$, and let M denote the directional matrix, i.e., its diagonal element $M_j = \Lambda_\alpha (R_\theta)^T$. If we use the vector $v = (\cos\theta, \sin\theta)^T$ and $v_\perp$ to denote the direction, then we obtain the weighted directional derivative of u:

$$M\nabla \otimes u = \Lambda_\alpha (R_\theta)^T \nabla \otimes u = \begin{pmatrix} \alpha_1 \nabla_v u \\ \alpha_2 \nabla_{v_\perp} u \end{pmatrix}, \tag{4.37}$$

where the directional derivative along a vector field z is

$$\nabla_z u(x) = \nabla u(x) \cdot z = \sum_{i=1}^{2} \partial_i u z_i. \tag{4.38}$$

Then ST-TDV becomes

$$ST - TDV(Y) = \begin{pmatrix} \nabla_{e_1} Y \\ \alpha \nabla_{e_2} Y \end{pmatrix}, \tag{4.39}$$

where e_1, e_2 are normalized eigenvectors aligned with the direction vector v and $v_\perp$ and α is a scaling matrix related to the eigenvalues, equaling M in Definition 4.1. Moreover, referring to

the TV semi-norm definition [114], ST-TDV can be re-written as

$$ST - TDV_{\beta}(Y, M) = \sup_{\Psi} \left\{ \int_{\Omega} Y \mathrm{div}_M \Psi dx | \, \|\Psi\|_{\infty} \leqslant \beta \right\}, \tag{4.40}$$

which is similar to the first-order TGV with a weighting matrix. Since the directional derivative acquired from an eigenvector is highly related to each projection pixel, we believe a pixel-wise anisotropic diffusion property can effectively mitigate the staircase artifact caused by the standard TV kernel. Thus our projection image prior design equals

$$p(Y) = e^{-\tau ST - TDV(Y)}, \tag{4.41}$$

where τ is a scaling parameter.

By combining Eq. (4.9) and Eq.(4.41), we can obtain the entire posterior distribution $p(P, Q|Y)p(Y)$. In fact, instead of maximizing its logarithm, we solve it by minimizing $-log(p(P, Q|Y)p(Y))$, that is,

$$\min_{Y,Q} \sum_{n=1}^{N} (\frac{(P_i - Q_i)^2}{2\sigma^2} - Q_i \ln I_{0i} + Q_i Y_i + \ln(Q_i!) + I_{0i} e^{-Y_i}) + \tau ST - TDV(Y). \tag{4.42}$$

As Eq. (4.25) and Eq. (4.26) show, we can decompose the objective function into two optimizing sub-problems with adding an extra auxiliary variable V:

$$\begin{aligned} (P1) &: \min_{Y,Q} \sum_{n=1}^{N} (\frac{(P_i - Q_i)^2}{2\sigma^2} - Q_i \ln I_{0i} + Q_i Y_i + \ln(Q_i!) + I_{0i} e^{-Y_i}) + \frac{\mu}{2}(Y - V)^2 \\ (P2) &: \min_{V} \frac{\mu}{2}(Y - V)^2 + \tau ST - TDV(V) \end{aligned} \tag{4.43}$$

Then we can apply an alternating optimization method to solve the equation. Specifically, for $(P1)$, we deploy the alternating direction method of multipliers (ADMM) strategy [118] as follows:

1. **Solving Q:** With the other parameter fixed, Q can be updated by solving

$$h(Q) = \min_{Q} \sum_{i=1}^{N} (\frac{(P_i - Q_i)^2}{2\sigma^2} - Q_i \ln I_{0i} + Q_i Y_i + \ln(Q_i!)), \tag{4.44}$$

which can be calculated separately as Algorithm 1 shows:

Algorithm 1 Algorithm for Updating Q

Input: Y, P, I_0 and σ

Initialize Q stored from last loop
for i = $1 : N$ **do**
  while $h(Q_i) > h(Q_i + 1)$ **do**
    Update $Q_i = Q_i + 1$
  end while
  while $h(Q_i) < h(Q_i + 1)$ **do**
    Update $Q_i = Q_i - 1$
  end while
end for

Output: $Q^* = Q$

2. **Solving Y:** With the other parameter fixed, Y can be updated by solving

$$\min_{Y} \sum_{i=1}^{N} (Q_i Y_i + I_{0i} e^{-Y_i}) + \frac{\mu}{2}(Y - V)^2. \tag{4.45}$$

A similar surrogate algorithm is performed for a numerical solution. Let Y^k denote the *k*-th step approximation, then we obtain new cost function:

$$\begin{aligned} Q(Y, Y^k) &= (Q - I_0 \odot e^{-Y_k})(Y - Y_k) + \frac{L}{2}(Y - Y^k)^2 + \frac{\mu}{2}(Y - V)^2 \\ &= \frac{L}{2}\left[Y - Y^k + \frac{1}{L}(Q - I_0 \odot e^{-Y_k})\right]^2 + \frac{\mu}{2}(Y - V)^2, \end{aligned} \tag{4.46}$$

where $\odot$ denotes element-wise multiplication, L is the Lipschitz constant to ensure continuity, and μ is a positive scalar. A closed-form solution is

$$\tilde{Y}^k = \frac{\mu V + L Y^k - (Q - I_0 \odot e^{-Y_k})}{L + \mu}. \tag{4.47}$$

Then entire iteration in the FISTA framework can be described as Algorithm 2 shows.

Algorithm 2 Algorithm for Updating Y

Input: Y^0, V, I_0 and μ

1: Initialize $Y^k = Y^0$ and $t^k = 1$
2: **for** k = $1 : M$ **do**
3: Update $\tilde{Y}^k = \arg\min_Y Q(Y, Y^k)$ with Eq. (4.43)
4: Update $t^{k+1} = \frac{1+\sqrt{1+4(t^k)^2}}{2}$
5: Update $Y^{k+1} = \tilde{Y}^k + \frac{t^k-1}{t^{k+1}}(\tilde{Y}^k - \tilde{Y}^{k-1})$
6: **end for**

Output: $Y^* = Y^k$

In (P2), to numerically solve the Eq. (4.40), we first rewrite the image prior numerically as in [101, 114]:

$$ST - TDV_\alpha(V, M) = \sup \langle V, \text{div}_M \Psi \rangle \, , \, \|\text{div}_v \Psi\|_\infty \leqslant \alpha. \tag{4.48}$$

Its corresponding primal format is

$$ST - TDV_\alpha(V, M) = \inf_{z_0=V, z_q=0, j=0,\ldots,q} \sum_{j=1}^{q} \alpha_{q-j} \left\| K_j z_{j-1} - z_j \right\|_1 , \; K_j = M_j \nabla, \tag{4.49}$$

where z_j is the image higher-order derivative component. Then (P2) is transformed into a nonlinear optimization problem, to which we apply the FOPD algorithm [119] as TGV method did.

first order primal dual (FOPD): Given a nonlinear optimization problem:

$$\min_x F(k(x)) + G(x), \tag{4.50}$$

The FOPD algorithm aims to solve the primal-dual form of the problem:

$$\min_{x \in X} \max_{y \in Y} \langle Kx, y \rangle + G(x) - F^*(y), \tag{4.51}$$

which is also known as the generic saddle-point problem. The general solution is

Algorithm 3 FOPD algorithm

1: Initialization: choose $\tau, \sigma > 0$, $\theta \in [0, 1]$ and set $\bar{x}^0 = x^0$
2: Iterations $(n \geqslant 0)$: Update x^n, y^n, $\bar{x}^n$ as follows:

$$\begin{cases} y^{n+1} = prox_{\sigma,F^*}(y^n + \sigma K \bar{x}^n) \\ x^{n+1} = prox_{\tau,G}(x^n - \tau K^* y^{n+1}) \\ \bar{x}^{n+1} = x^{n+1} + \theta(x^{n+1} - x^n) \end{cases} \tag{4.52}$$

Referring to the FOPD algorithm, we can re-write the (P2) problem as

$$V^* = \arg\min \sum_{j=1}^{q} \alpha_{q-j} \left\| K_j z_{j-1} - z_j \right\|_1 + \frac{\mu}{2\tau} \left\| V - Y \right\|_2^2, \tag{4.53}$$

and its primal-dual format is

$$\min_z \max_w = \left\langle \sum_{j=1}^{q} K_j z_{j-1} - z_j, w_j \right\rangle - \sum_{j=1}^{q} I_{\left\{ \|\cdot\|_\infty \leqslant \alpha_{q-j} \right\}}(w_j) + \frac{\mu}{2\tau} \left\| V - Y \right\|_2^2, \tag{4.54}$$

where $I_{\left\{ \|\cdot\|_\infty \leqslant \alpha_{q-j} \right\}}(w_j)$ is an indicating function that equals 0 within the feasible domain and ∞ outside. To solve Eq. (4.54), we apply the proximal gradient descent method [115, 119, 120]. Based on the assumption that variable z_j is independent, we extend the Eq. (4.54) as follows

$$\min_z \max_w = \left\langle \sum_{j=1}^{q} K_j z_{j-1} - z_j, w_j \right\rangle - \sum_{j=1}^{q} I_{\left\{ \|\cdot\|_\infty \leqslant \alpha_{q-j} \right\}}(w_j) + \frac{\mu}{2\tau} \left\| \begin{bmatrix} V \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{q \times 1} - \begin{bmatrix} Y \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{q \times 1} \right\|_2^2, \tag{4.55}$$

then it can be divided into multiple small problems, i.e, for each j,

$$\sum_{j=1}^{q} \left[\langle K_j z_{j-1} - z_j, w_j \rangle - I_{\left\{ \|\cdot\|_\infty \leqslant \alpha_{q-j} \right\}}(w_j) + \frac{\mu}{2\tau} \left\| V_j - Y_j \right\|_2^2 \right]. \tag{4.56}$$

Then Algorithm 3 becomes: $Prox_{\sigma F^*}$**:** Given the indicating function I, the proximal problem can be transformed into a projection problem that projects onto the convex sets $w_i | \left\| w_i \right\|_\infty \leqslant$

α_{q-j}. The associated projections can be presented as follows:

$$prox_{\sigma F^*}(\tilde{w}) = \sum_{j=1}^{q} proj_{\sigma F^*}(\tilde{w}_j),\ proj_{\sigma F^*}(\tilde{w}_j) = \frac{\tilde{w}_j}{\max(1, \alpha_{q-j}^{-1} \left\|\tilde{w}_j\right\|_2)}. \tag{4.57}$$

$Prox_{\tau G}$**:** Given Eq. (4.51), the proximal problem of function G is

$$proj_{\tau,G}(\tilde{z}) = \arg\min_{z} \left\{ \frac{\mu}{2\tau} \left\| \begin{pmatrix} \tilde{z}_0 - Y \\ 0 \\ \vdots \\ 0 \end{pmatrix} \right\|_2^2 + \frac{1}{2m} \left\| \begin{pmatrix} z_0 \\ z_1 \\ \vdots \\ z_q \end{pmatrix} - \begin{pmatrix} \tilde{z}_0 \\ \tilde{z}_1 \\ \vdots \\ \tilde{z}_q \end{pmatrix} \right\|_2^2 \right\}, \tag{4.58}$$

where $z_0 = V, z_1 = V_1, ..., z_q = 0$. The minimum value can be acquired through gradient method. Therefore, we have similar iterations as follows

$$\begin{cases} w^{n+1} = prox_{\sigma,F^*}(w^n + \sigma K \bar{z}^n) \\ z^{n+1} = prox_{m,G}(z^n - \tau K^* w^{n+1}) \\ \bar{z}^{n+1} = z^{n+1} + \theta(z^{n+1} - z^n) \end{cases}, \tag{4.59}$$

where a detailed description is presented in Algorithm 4.

Algorithm 4 Algorithm for Updating V

Input: Y, α, $K = M\nabla$, τ, θ and μ

Initialize $z = [V, 0, ...0]^T$ and $w = 0$
while stopping criteria is not satisfied **do**
  for j = $1 : q-1$ **do**
    $w_j^n = w_j^n + \sigma\left(K_j \bar{z}_{j-1}^n - \bar{z}_j^n\right)$
    $w_j^{n+1} = w_j^n / \max(1, \alpha_{q-j}^{-1} \left\|\tilde{w}_j\right\|_2)$
  end for
  $w_q^n = w_q^n + \sigma K_q \bar{z}_{q-1}^n$
  $w_q^{n+1} = w_q^n / \max(1, \alpha_0^{-1} \left\|\tilde{w}_q\right\|_2)$
  for j = $2 : q$ **do**
    $z_{j-1}^n = z_{j-1}^n - mK_j^* w_j^{n+1} - w_{j-1}^{n+1}$
    $z_{j-1}^{n+1} = z_{j-1}^n$
  end for
  $z_0^n = z_0^n - mK_1^* w_1^{n+1}$
  $z_0^{n+1} = (z_0^n + Y)(1 + \mu m/\tau)^{-1}$
  $\bar{z}^{n+1} = z^{n+1} + \theta(z^{n+1} - z^n)$
end while

In summary, a general solution to our proposed MBIR method, called the MAP-ST-TDV method that combines ST-TDV prior and a MAP model, is shown in Algorithm 5.

Algorithm 5 Algorithm for solving Eq. (4.43)

Input: I_0, $P = I_0 \odot e^{-Y}$

1: Initialize μ, ρ, τ, $Y = \max(\ln(I_0) - \ln(P), 0)$
2: **while** stopping criteria is not satisfied **do**
3: Update Q by Algorithm 1
4: Update Y by Algorithm 2
5: Update V by Algorithm 4
6: Update $\mu = \rho\mu$
7: **end while**

4.3 Projection Synthesis Technique

The denoised projection image is further calibrated by our proposed geometry calibration algorithm, and then we increase the projection through image synthesis to assist in solving the sparse sampling problem. In this research, synthesis is realized through a DL network, since conventional methods are time-consuming.

4.3.1 Deep Learning Introduction

Neural networks (NN) are inspired by the inner working of human brains based on the concept of biological neurons and networks. In biology, neuron interaction with another involves complicated biochemical reactions, shown in Fig. 4.2(a). In short, a single neuron will pass a message to another once the input signal exceeds a certain threshold. Similarly, a mathematical model of this reaction is shown in Fig. 4.2(c), which is the unit module in NN. Its interaction includes two steps, accumulation and activation, which can be represented as

$$f = \sigma(w^T x + b) \tag{4.60}$$

where w is a weighting matrix, x is the input signal, b is a bias signal and $\sigma(\cdot)$ is the activation function. For instance, a widely deployed activation function is called a rectified linear unit

(ReLU) function [121]:

$$\sigma(x) = \max(x, 0) \tag{4.61}$$

Other common activation functions are the logistic sigmoid function and the hyperbolic tangent function [121]. In summary, from Eq. (4.60), we can treat the single neuron as a nonlinear nonnegative optimization function, and a NN is a function of these functions, whose output is a nonlinear response to the input data. In DL, a NN not only builds the nonlinear mapping between input and output, but also incorporates representation learning or feature learning in layer design to capture high-level latent information. One advantage of high-level representation learning is that DL is able to compensate for the lack of prior knowledge of features or unavailability of labeled data in the output.

In DL, CNN, which uses kernel function to realize sparse interaction, is the most commonly applied one in analyzing images inspired by the study of Visual Cortex [122]. It reveals that individual neurons respond to stimuli only in restricted region (known as the Receptive Field) which can tremendously reduce the storage request.

For each layer of a CNN, the signal is processed by a kernel function to capture the spatial and temporal dependencies, which is similar to a convolution of a 2D image. Compared with other NN types, CNN has shown certain advantages:

1. weights are shared for the kernel function based on the assumption that the feature has the same importance at different location, thus CNN can reflect any affine transformation and save storage on weights.

2. CNN has introduced a pooling layer to further reducing the image size, especially applying max pooling to perform noise suppression simultaneously, which is similar to a pyramid structure to extract high-level features.

The later developed generative adversarial network (GAN) framework has strengthened CNN performance, which consists of a generator and discriminator. The generator is typically a deconvolutional neural network, and the discriminator is a CNN.

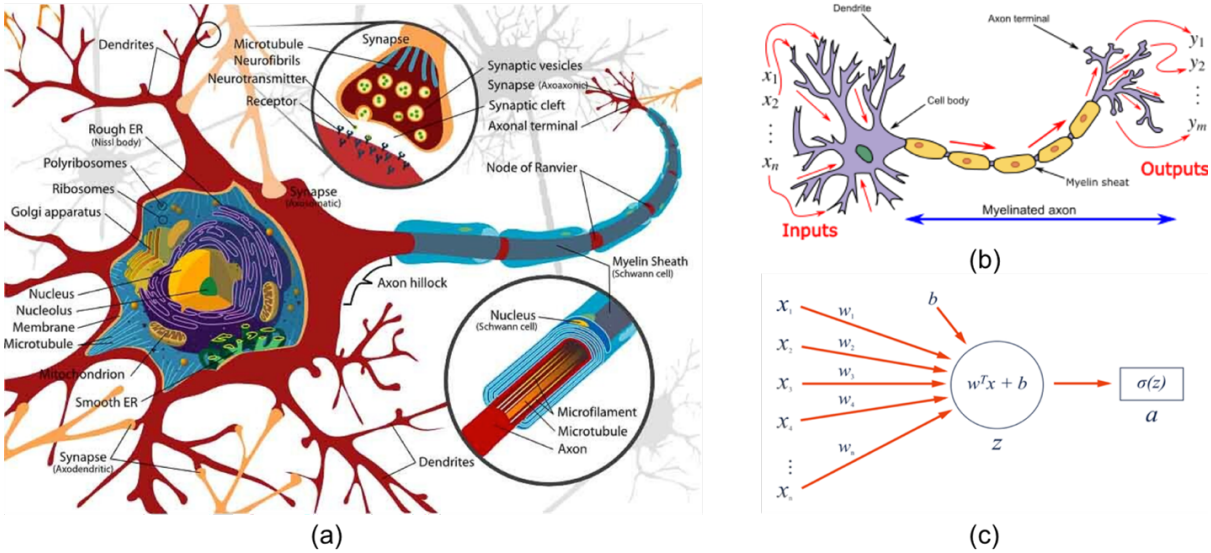


Figure 4.2: Introduction to Neural Networks. (a) biological neural network modeling brain's interaction (Wikipedia: `https://en.wikipedia.org/wiki/Artificial_neural_network`). (b) single biological neuron modeling message passing. (c) artificial neuron that simulates (b)

4.3.2 Spatial Transformer Network

A conventional image synthesis algorithm is time-consuming. In our simulation, implementation of the algorithm in [98] takes 3 hours to obtain 4 times the projection images from 96 to 384. Such inefficiency limits its application in practice. In order to accelerate the image synthesis, inspired by the spatial sampling [123] problem in video frame synthesis, we propose to use a data-driven network as a replacement regardless of the complicated mathematical background. Our proposed spatial transformer network is shown in Fig. 4.3. As the figure shows, a five-layer convolution block has been applied to encode the high-level image information and then deconvolved into a three-channel feature map—called tensor block in the context—for further processing. Besides, each channel represents an image tensor component along one dimension in space, i.e.,

$$T = (T_x, T_y, T_z) = U(x_1, x_2), \tag{4.62}$$

where U stands for the U-net autoencoder and x_1 and x_2 represent the previous and next projection image. With the assumption that the intermediate projection is linearly, equally interpolated by two adjacent projections, two sampling matrices m_1 and m_2 on the two images are

calculated once the tensor block is transferred into the Bilinear Sampler Generator and the new projection image is interpolated as the equation shows.

$$x = T_z * B(x_1, m_1) + (1 - T_z) * B(x_2, m_2),\ m_{i,j} = (i \pm \frac{T_x}{2}, j \pm \frac{T_y}{2}) \tag{4.63}$$

where m is the sampling matrix and B is a bilinear interpolation. Positive and negative calculation represents the corresponding coordinates in later and earlier images. After the matrix is obtained, a pixel-wise bilinear interpolation is applied to form an intermediate projection image.

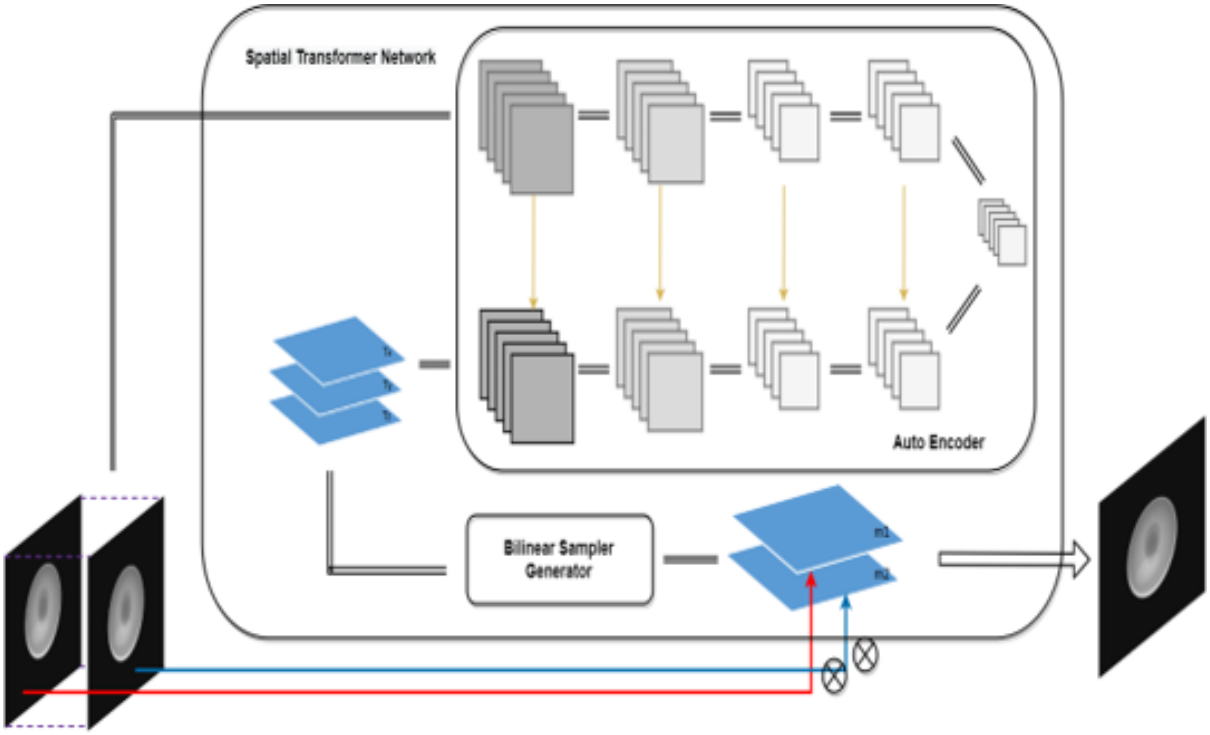


Figure 4.3: Introduction of proposed network structure. Top: Spatial Transformer Network which consists of 5 convolution layers. For each layer, the number of filters equals 64, 128, 256, 512, 1024, respectively. Image size is reduced to half by convolution pooling from one layer to another and upsampling realized by deconvolution. Input is two adjacent sparse sampled projection images and output is a 3-channel feature map recording intensity-flow information. Bilinear Sampler Generator is a coordinates production block that transforms a tensor block into two sampling matrices.

4.3.3 GAN framework

Previous work has demonstrated that adversarial learning can successfully compensate the structure loss between two domains, source image and target image. In this research, we adopt a GAN framework for our network training, Fig. 4.4, where the spatial transformer network is treated as a generator and a discriminator is added to distinguish between real images and synthetic images. In order to improve the accuracy and stability of GAN training, besides the general adversarial loss applied for discriminator training, a comprehensive loss function has been applied for generator training.

$$G^* = arg\ min_G[\lambda_1 \mathcal{L}_{GAN} + \lambda_2 \mathcal{L}_1(y, \hat{y}) + \lambda_3 \mathcal{L}_{DP}(y, \hat{y})] \tag{4.64}$$

$$\mathcal{L}_{GAN} = \mathbb{E}_{y,\hat{y}}[logD(y)] + \mathbb{E}_{x,z}[log(1 - D(G(x)))] \tag{4.65}$$

$$\mathcal{L}_1(y, \hat{y}) = \|y - \hat{y}\|_1 \tag{4.66}$$

$$\mathcal{L}_{DP}(y, \hat{y}) = \frac{1}{N} \sum_{j=1}^{N} \frac{1}{C_j H_j W_j} \|\phi_j(\hat{y}) - \phi_j(y)\|_2^2 \tag{4.67}$$

where $\mathcal{L}_{GAN}$ is adversarial loss that enables the network to synthesize images with more realistic features, $\mathcal{L}_1$ is content loss that enables an image to share a similar global structure, and $\mathcal{L}_{DP}$ is discriminator perceptual loss[79] that strengthens training stability, ϕ_j is the activation layer in the discriminator. Parameters $\lambda_1 = 1$, $\lambda_2 = 50$ and $\lambda_3 = 20$ are used to balance the losses here. The number of training epochs is set to 200, the learning rate obtained from 5-fold cross-validation is 0.0001, and the batch size is set to 4 due to the image size. For comparison, two learning-based sinogram interpolation algorithms, U-net [88] and GAN based sinogram inpainting network (SIN) [79], are applied as well.

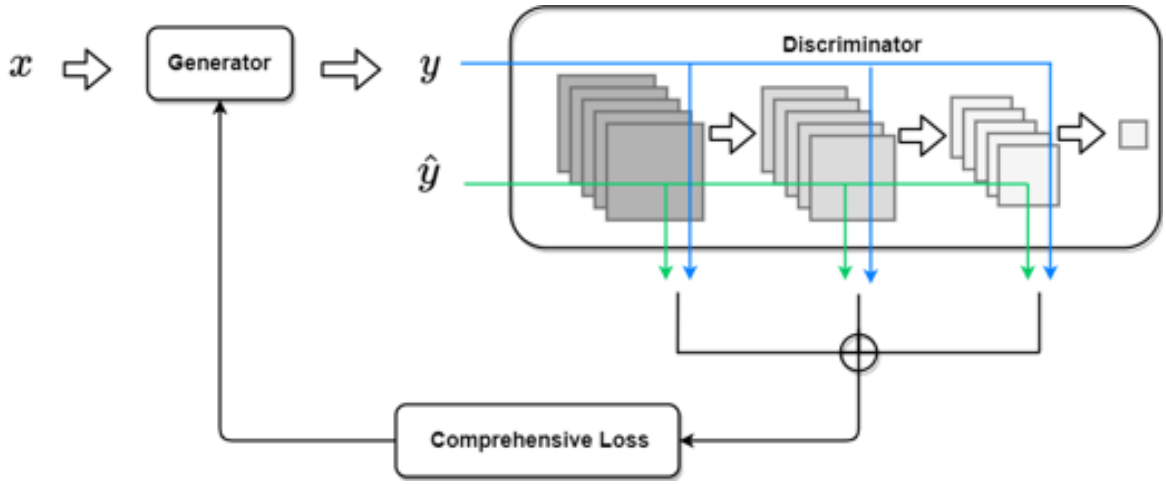


Figure 4.4: GAN framework: x is input multi-projections, y is output of the spatial transformer generator, $\hat{y}$ is ground truth projection image. Discriminator consists of three convolution blocks that include one convolution layer, one batch normalization layer, and one ReLU activation layer for each block.

4.3.4 Extra: Deep Transfer Learning

Deep transfer learning (DTL) [124] is another framework that aims to solve the insufficient training data problem. Wide application of U-net in medical imaging has proved its ability to effectively extract the high-dimensional features and establish links between input images. However, the requirement of a large number of training images has limited its application. Therefore, DTL has been proposed to improve the reconstructed image quality with reduced computational cost while maintaining accuracy. The detailed workflow is shown in Fig.4.5.

We first collected a publicly available large-scale dataset M, which is also mouse data but scanned from another device (denoted as device A), for initial training. The major challenge is that the publicly available dataset M is based on Device A, whose noise pattern is different from ours. It is infeasible to directly use a model that is trained based on Device A to our device. However, we are able to obtain small-scale phantom images that are collected from our own device, and we use the TV based reconstruction images as *true label* (denoted as dataset P). We choose phantom data since the TV based reconstructions are more ideal, and we use dataset P for the further task of fine-tuning weights. The datasets details are shown in Table 4.1.

In summary, we propose to train a reconstruction deep model using publicly available large-scale dataset M (denoted as Supervised Task A) to learn the reconstruction mapping of

a mouse (possibly biased to Device A). The pre-trained model cannot be directly applied to our device, since the model is not exposed to our device and has no prior knowledge about the unique configurations of our device. Thus, we propose to fine tune the pre-trained model using dataset P (denoted as Supervised Task B), and we hypothesize that dataset P contains knowledge about our own device. Once fine-tuned on dataset P, the specific knowledge regarding our device can be fed to the original reconstruction model.

As shown in Fig 4.6, in Fine-Tuning U-net structure, the max pooling layer has been replaced by an average pooling layer given the consideration that an average pooling can smooth the flattened region from a noise pattern in a way that is closer to the distribution of the ideal image. Also, the average of four adjacent pixels can utilize more image content information than max pooling since limited training data is provided in task B.

Table 4.1: Datasets Descriptions

Dataset	Device	Input	True Label
M	A	simulated mice sparse CT	ground truth mice CT
P	Ours	sparse CT on CT phantom	CT phantom sparse CT + TV
Target	Ours	reconstructed mice sparse CT	target restored CT images

Further improvement is realized by cycle training, CycleGAN, which frees us from the paired data requirement [125, 126]. The framework is shown in Fig. 4.7, which includes two U-nets. This design enables us to avoid the limitation of DTL; that is, we use the MBIR result as the true label in DTL fine tuning. Instead, it utilizes two types of images, input FDK result and target clinical CTs, for network training. A well-trained network is able to transform the reconstruction image contrast to the target image.

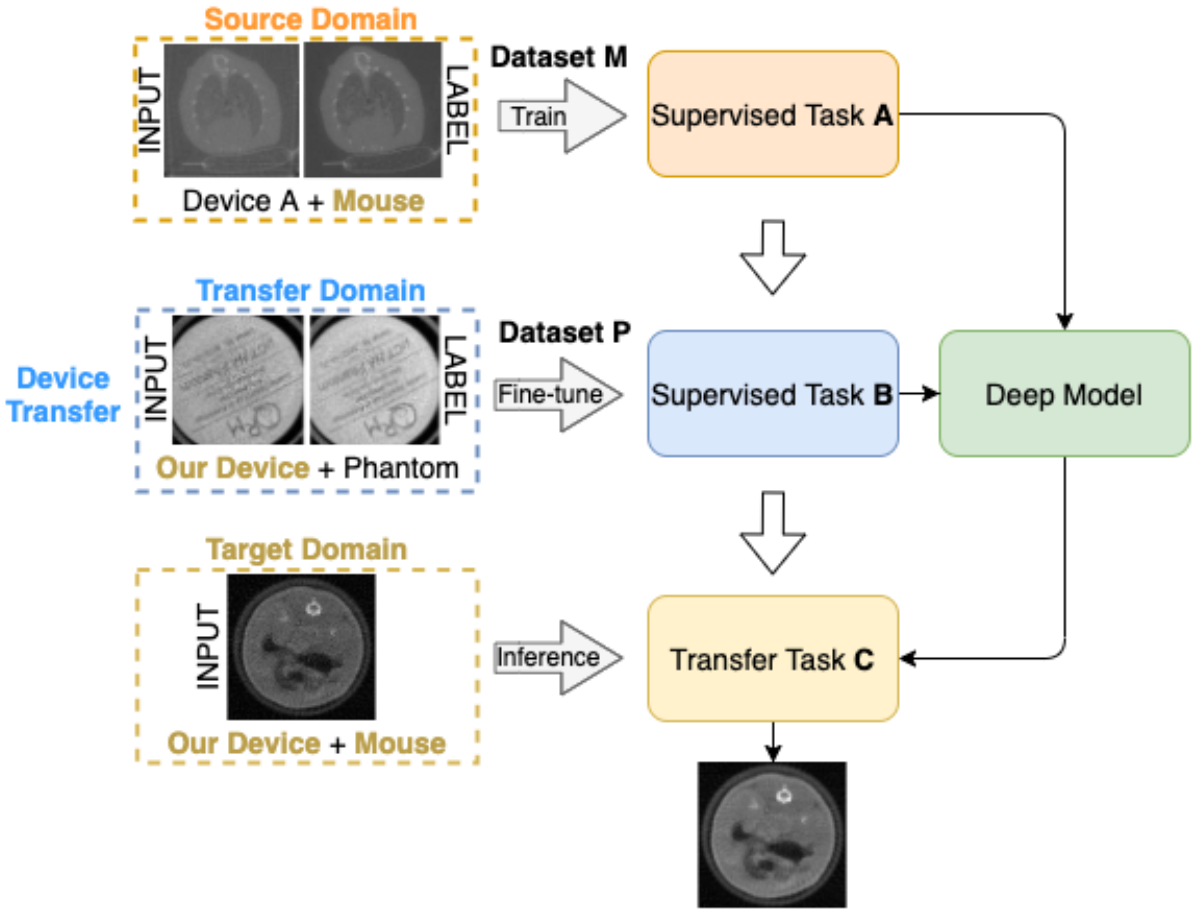


Figure 4.5: Deep tranfer learning framework: Supervised task A is to learn the mapping of image constructions possibly with device bias. Supervised task B is to neutralize the device bias and transfer to our own device. Our final task C is to reconstruct mouse images scanned by our own device.

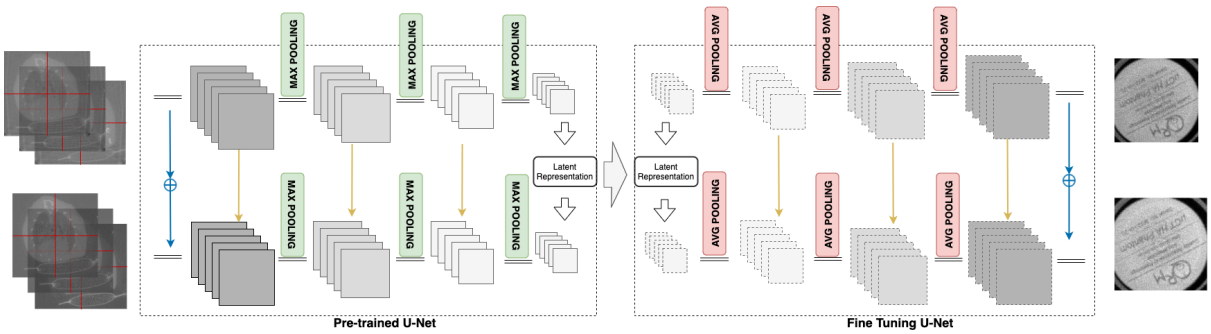


Figure 4.6: Transfer Model Structures. A) Model at left is the pre-trained model on Dataset M with max-pooling intertwined; B) Model at right is the fine-tuning procedure on Dataset P with average-pooling intertwined.

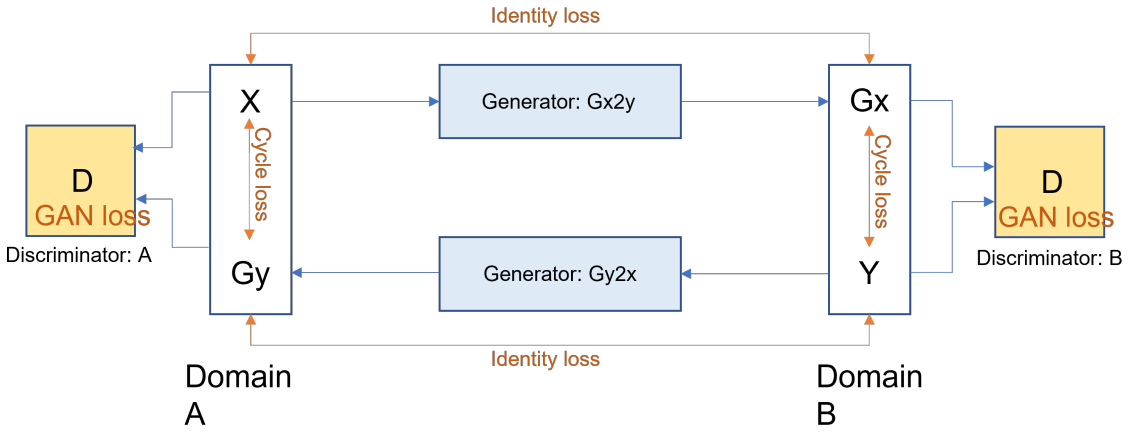


Figure 4.7: Illustration of CycleGAN framework, which includes two generators and two discriminators. Each generator includes a 5-layer U-net, and each discriminator includes a 3-convolution block as Fig. 4.4 shows.

4.3.5 Data Preparation

In this investigation, the PM Public Micro-CT Image Database was used for network training. It consists of an image set of longitudinal micro-CT scans of four live mice and three post-mortem mice, which were made and uploaded by the VIA group and the Weiss lab. To avoid the possibility of motion and breathing artifacts caused by live mice, we only selected the three dead mouse images for training.

In image synthesis network training, the training dataset consisted of 97 projections produced for each selected mouse volume data, together with five other mouse projection datasets scanned from the *MiSpinner*. Every three projections formed a training unit where head and tail projections were used as network input and the mid projection was used as the ground truth image for discriminator input in GAN training.

For the comparison test, 4 times sampling projection data, 384 projections for each mouse were used as ground truth images. The original 97 projections were converted to 384 projections by linear interpolation. Then the interpolated projections and corresponding ground-truth images were fed into the discriminator for training.

In DTL training, for supervised task A, three dead mouse images provided a total of 2566 low-dose micro-CT slices of the mouse thorax region. Each slice and its corresponding FDK sparse reconstruction were produced by the ASTRA toolbox [127]. We also divided the images

into smaller size patches, 256×256, before feeding them into the network to increase the capacity of the training set and to ensure the distribution of weights among neural networks were independent from the image pattern.

For supervised task B, instances-based deep transfer learning means that partial instances in the source domain can be utilized by the target domain with appropriate weight. Thus TV iterative results of the CT phantom image were labeled as images with "appropriate weight" to compose a distribution in the target domain and then together with their FDK results used for model transfer learning in task B.

4.4 Experimental Results

4.4.1 MAP-ST-TDV Denoising Study

To verify the robustness and feasibility of the proposed MAP-ST-TDV method, two simulations were conducted: a 2D Lungman chest phantom simulation and a 3D Shepp-Logan phantom simulation. In the 2D simulation, for simplicity, we applied the parallel scanning mode in projection generation, i.e., a total of 97 projections evenly distributed around 360 degrees were generated. The phantom image size was 512× 512 in pixels. In the 3D simulation, similar to the Chapter 3 simulation, we set SOD equal to 1500, SDD equal to 2000, and the phantom volume size to 512× 512 × 512. Besides, to simulate the low-dose radiation case, we set the incident X-ray intensity to 1×10^4, which is 1/100 of other simulation [67] and added the same Poisson and Gaussian noise into projections as Eq. (4.7) described.

Lungman chest phantom simulation is presented in Fig. 4.8, which includes a comparison of our algorithm with other state-of-the-art MBIR methods. The original phantom image is displayed in Fig. 4.8(a), and the corresponding sparse reconstruction is shown in Fig. 4.8(b) and Fig. 4.8(c) with and without noise. It can be seen that the FBP result is clearly corrupted by sparse sampling and noise. A comparison of different MBIR methods, PWLS, PWLS-TV, PWLS-TGV, MAP-TV, MAP-TV2 is displayed in Figs.. 4.8(d-h). To ensure fairness in the comparison, we set the main loop in each algorithm to 40 iterations and the inner loop when relevant, e.g., FOPD in TGV and ST-TDV, to 100 iterations. Generally, we can observe

that MAP-TV2 and MAP-ST-TDV results outperformed other results, where they can both effectively suppress noise-induced artifacts and recover the phantom structure. Meanwhile, a residual image, the difference between the reconstructed image and the original phantom image, is presented in Fig. 4.9 to support our observation, where we find that PWLS, MAP-TV2 and MAP-ST-TDV can better preserve the edges with fewer boundary detail loss. Furthermore, the line profile at pixel position $x = 170$ and y from 110 to 140, shown in Fig. 4.8(a) with a red mark, is displayed in Fig. 4.10. We can see that our MAP-ST-TDV result overall is closer to the phantom data than other methods. Specifically in high value region, where we marked with red rectangle, we can view that MAP-ST-TDV method maintains similar shape and closer values than other methods.

A quantitative comparison is shown in Table 4.2. The comparison uses three measurements: peak signal-to-noise ratio (PSNR), structure similarity index measure (SSIM), and mean absolute error (MAE) for all competing methods. The larger PSNR and SSIM are, the closer the target image is to the reference one. On the contrary, the smaller the MAE is, the better the reconstructed image is. From the table, it is clear that our proposed algorithm outperforms other competing methods in terms of all these evaluations.

Table 4.2: PSNR, SSIM and MAE measurements of the Lungman Phantom reconstructed by all MBIR methods

	nFBP	PWLS	PWLS-TV	PWLS-TGV	MAP-TV	MAP-TV2	MAP-ST-TDV
PSNR	19.40	19.55	23.47	21.67	22.54	23.11	**23.62**
SSIM	0.2038	0.2081	0.4439	0.4561	0.3318	0.3975	**0.4562**
MAE	0.0753	0.074	0.0417	0.0414	0.0499	0.0442	**0.0405**

Next we extended the simulation into 3D volume case, i.e., Shepp-Logan phantom, and the simulation result is presented in Fig. 4.11. In this case, instead of processing the sinogram which is a 2D image along the scanning angle distribution, we sequential denoised the object projection image directly with three MBIR methods, PWLS-TV, MAP-TV and MAP-ST-TDV algorithm, based on the piece-wise constant assumption on the natural projection image. Similarly, numerical analysis, PSNR, SSIM, MAE and MSE in Table 4.3, have shown that our proposed method can achieve the best performance in image denoising.

Table 4.3: PSNR, SSIM, MAE and MSE measurements of the Shepp-Logan Phantom reconstructed by PWLS-TV, MAP-TV, MAP-ST-TDV methods

	nFBP	PWLS-TV	MAP-TV	MAP-ST-TDV
PSNR	22.95	23.98	23.91	**24.64**
SSIM	0.2061	0.03357	0.2713	**0.4223**
MAE	0.0469	0.0354	0.0389	**0.0303**
MSE	0.0051	0.0040	0.0040	**0.0034**

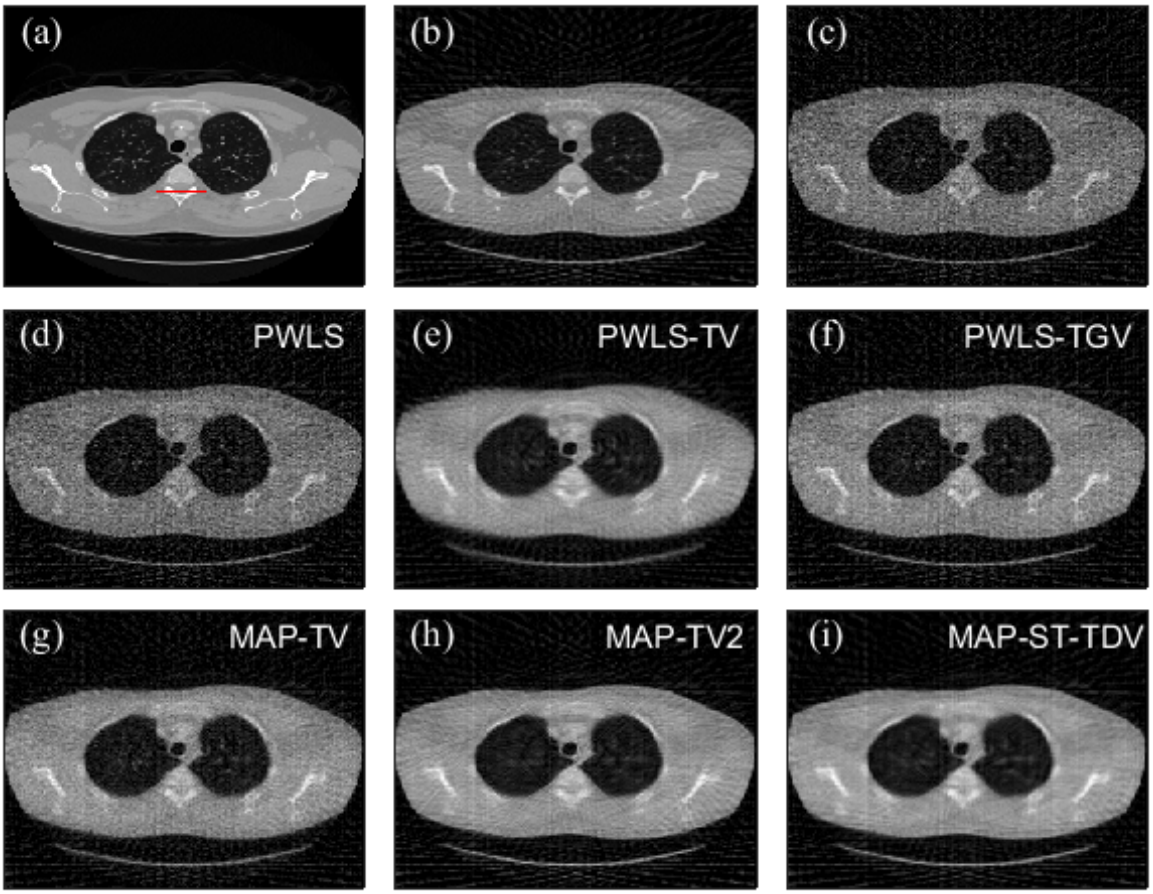


Figure 4.8: Comparison of different MBIR methods on LungMan chest phantom image: (a) Original phantom image, (b) low-dose 97-projection sparse reconstruction, (c) sparse reconstruction with noise, (d) sinogram denoising with PWLS method and corresponding reconstruction, (e) PWLS-TV result, (f) PWLS-TGV result, (g) MAP-TV result, (h) MAP-TV2 result, (i) Our proposed MAP-ST-TDV method result.

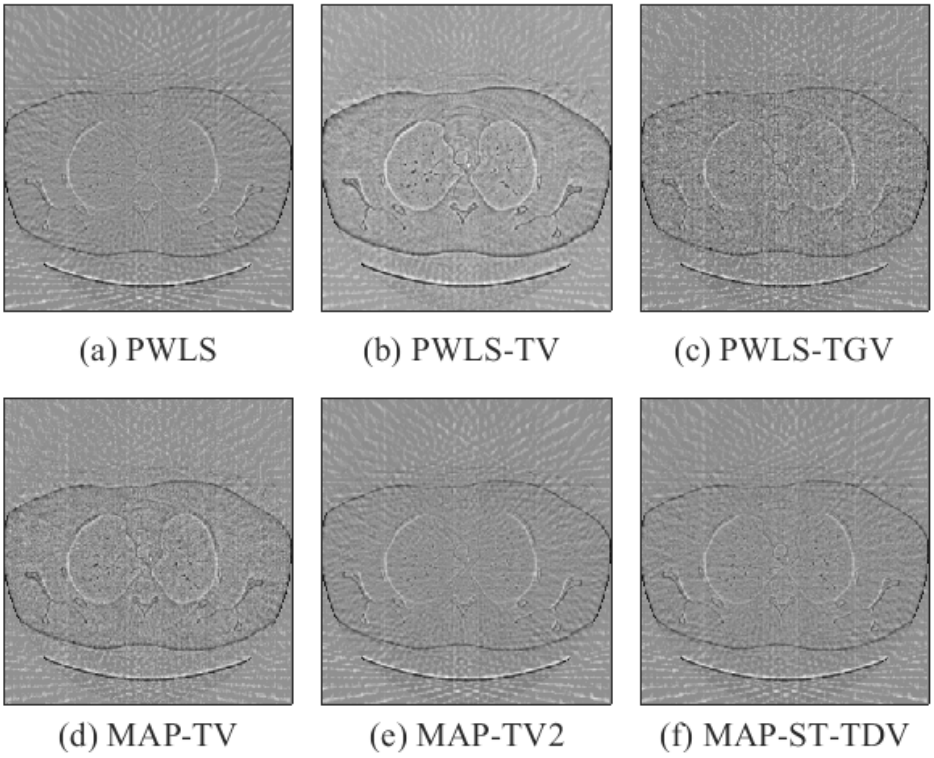


Figure 4.9: (a) - (f) are residual images of the result reconstructed by PWLS, PWLS-TV, PWLS-TGV, MAP-TV, and MAP-TV2.

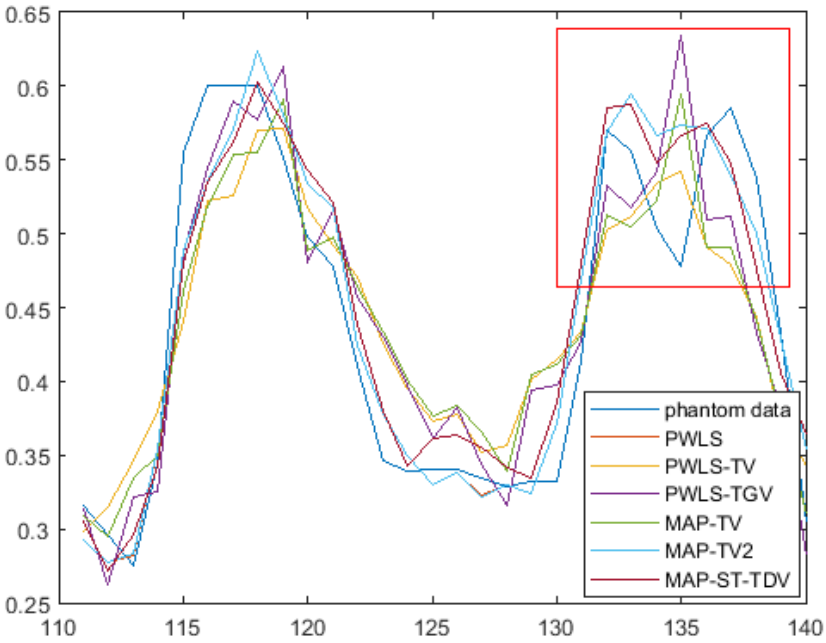


Figure 4.10: Line profile of the red line marked in phantom image Fig. 4.8(a) and recovery results of the six MBIR methods.

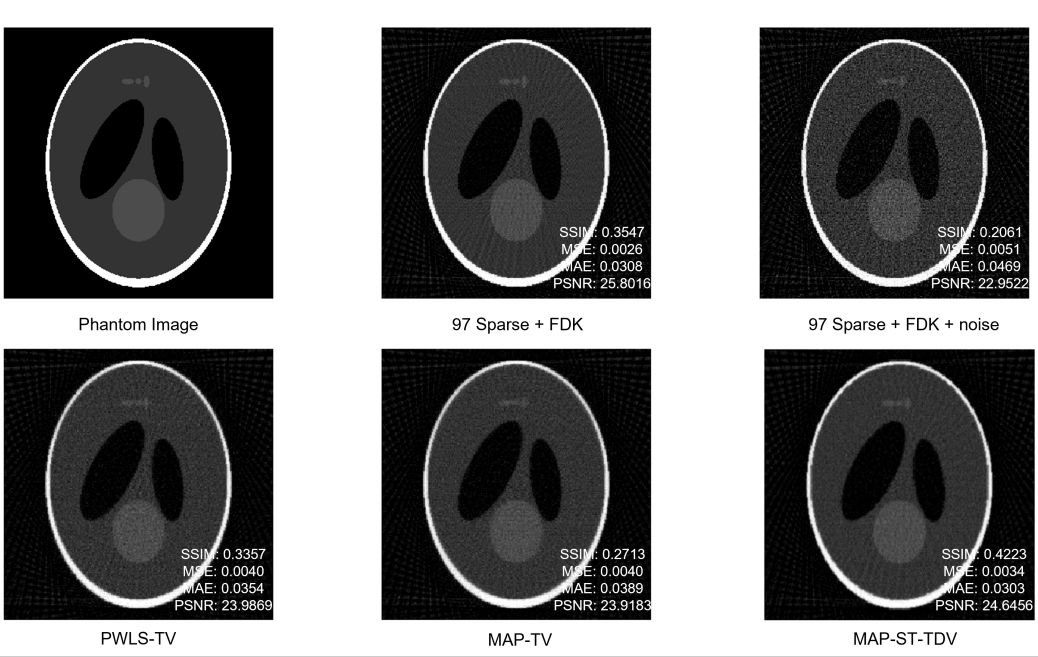


Figure 4.11: ST-TDV method test on 3D Shepp-Logan phantom: (a) original phantom image, (b) 97-projection sparse reconstruction, (c) sparse reconstruction with noise, (d)-(f) are FDK result after PWLS-TV, MAP-TV, and MAP-ST-TDV denoising.

4.4.2 Image Synthesis Study

Further study was done to verify the feasibility of the proposed network, which we call Proj-syn-net. A simulation on the 3D Shepp-Logan phantom was conducted, and the result is shown in Fig. 4.13. To ensure the alignment of the units in CBCT simulation, the source to detector/scintillation plate distance and source to isocenter distance were converted into pixel length, 5106 and 4382 respectively. Two numerical analysis methods, SSIM and PSNR, were applied for image quality assessment. In this case, a mix of open source data, PM public data, and some of the real-world denoised projection were used for semi-supervised learning. A comprehensive loss, Eq. (4.64) to Eq. (4.67), was applied and Adam optimizer was used for acceleration.

A five-fold cross validation was applied in the network training, and Fig. 4.12 shows the change of learning curve with respect to different learning rates. Determination of hyper-parameters is crucial to the network performance, especially the learning rate. A large learning rate will lead to the network overfitting and collapsing, as shown in Fig. 4.12(d), where training

loss and validation loss do not converge. Furthermore, we can observe that $lr = 0.0001$ generates the least MSE—0.2771—in network training. Accordingly, we applied the same learning rate to other learning-based methods to ensure fairness.

A comparison of all three DL methods is shown in Fig. 4.13(a). We can see that all of them can successfully reconstruct the phantom structure meanwhile effectively reducing the streak artifacts. More specifically, our proposed network, (a-f), improves the SSIM from 0.3143 to 0.6202, landing in the interval between the two comparison methods, (a-d) and (a-e). Even though PSNR has shown a reduction, line profiles (b) and (c) along the center column have shown similar intensity distribution of the three nets.

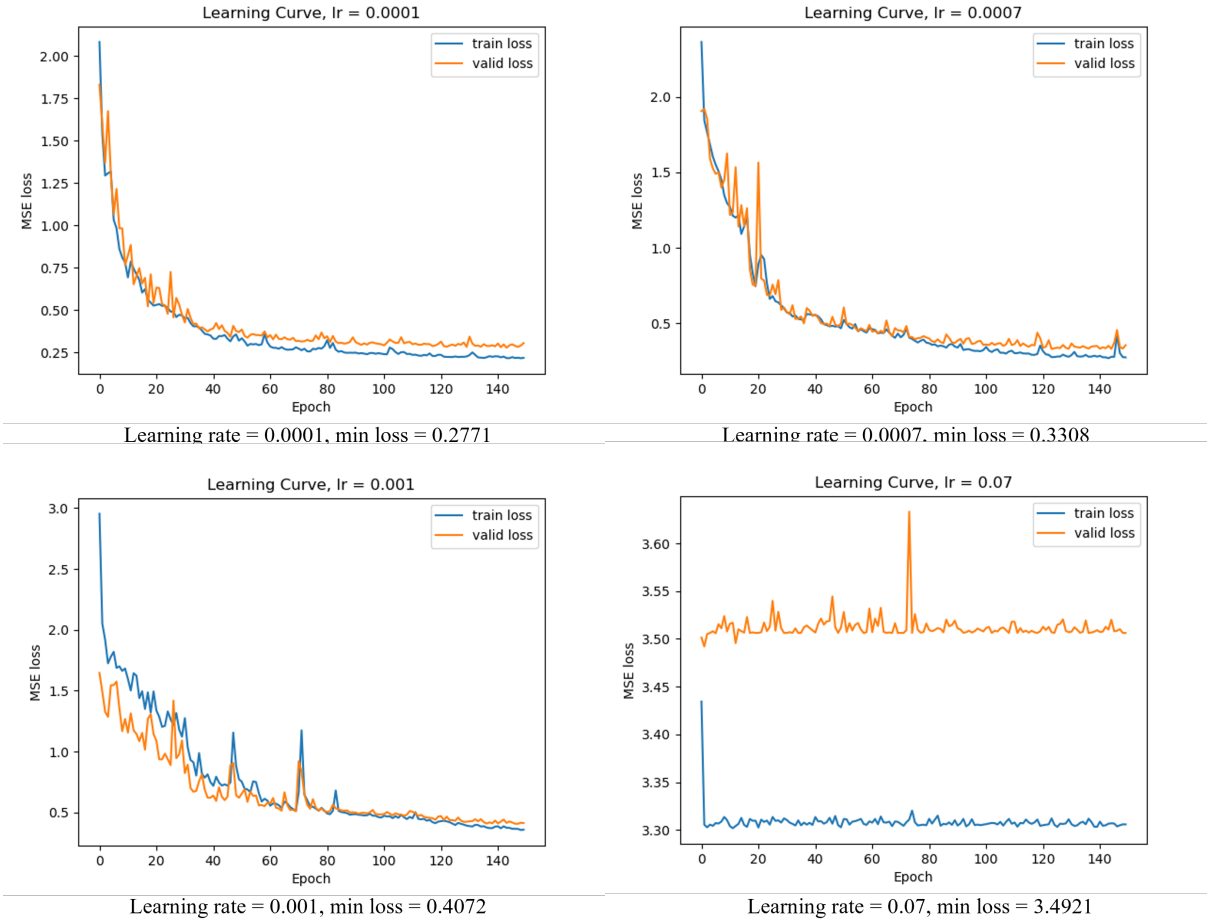
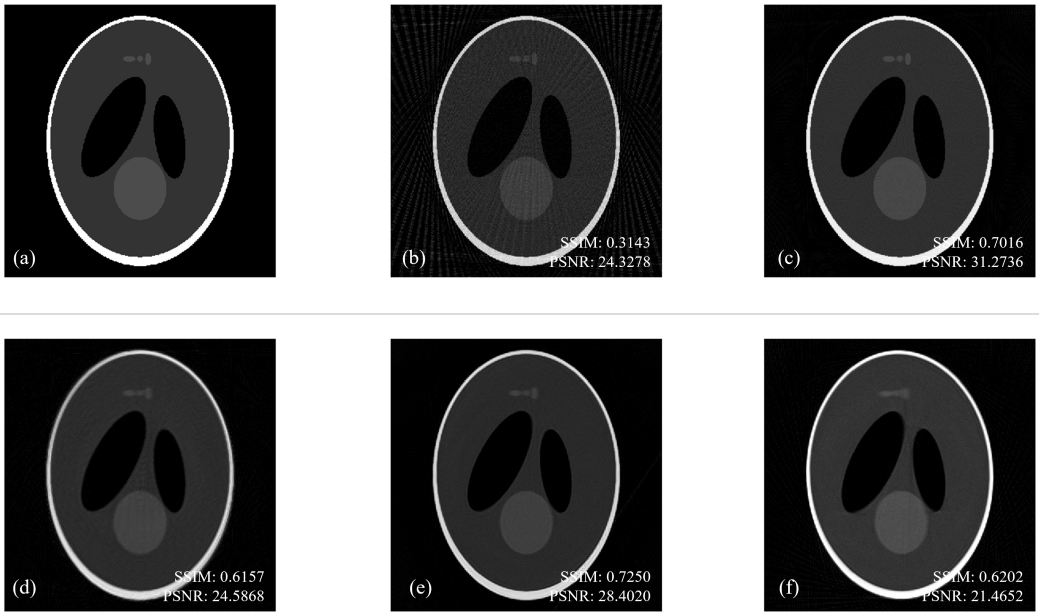
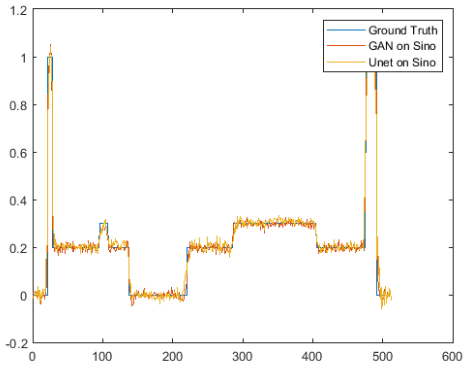


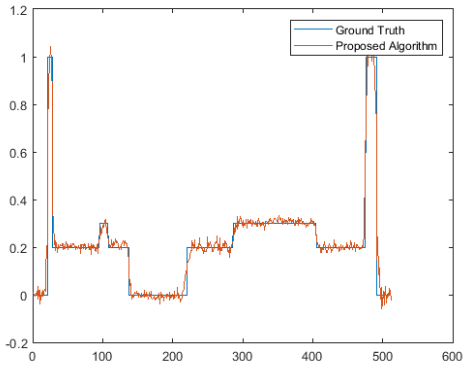
Figure 4.12: Illustration of learning rate effect one the network training. Minimum MSE loss is acquired at 0.2771, 0.3308, 0.4072, 3.4921 with respect to learning rate equals 0.0001, 0.0007, 0.001 and 0.07, separately.



(a) Shepp-Logan phantom simulation



(b) line profile of center column with U-net and GAN



(c) line profile of center column with proposed method

Figure 4.13: Illustration of proposed network performance in phantom simulation: (a) simulation of 3D S-L phantom under different situations: (a-a) phantom ground truth image, (a-b) reconstruction under 97 sparse-angle sampling, (a-c) result under 384 sampling, (a-d) reconstruction from the sinogram extended by trained U-net, (a-e) reconstruction from the sinogram extended by trained GAN, (a-f) reconstruction from the proposed network; (b) line profile of the center column from the result (a-d) and (a-e); (c) line profile of the center column from the result (a-f).

A qualitative test on real data further comparing the network performance and the result is shown in Fig. 4.14. Extra competing methods—classic TV-based MBIR method and structure tensor based image synthesis method, also known as directional image interpolation (DI)—are included for comparison. The rectangle region has revealed that both DI method result,

Fig. 4.14(c) and the U-net interpolation method result, Fig. 4.14(d), introduce extra ringing artifacts into reconstruction, while the GAN result, Fig. 4.14(e) oversmooths the text region. Only the proposed method result, Fig. 4.14(f), preserve the texture details well and give rise to the competing result with the MBIR result. Furthermore, we can see that the Proj-syn-net result possesses better contrast than Fig. 4.14(b), where quantum noise and the dark region in the center are effectively removed. Therefore, we can conclude that our proposed network can successfully improve the image quality in terms of the existing limited projection data.

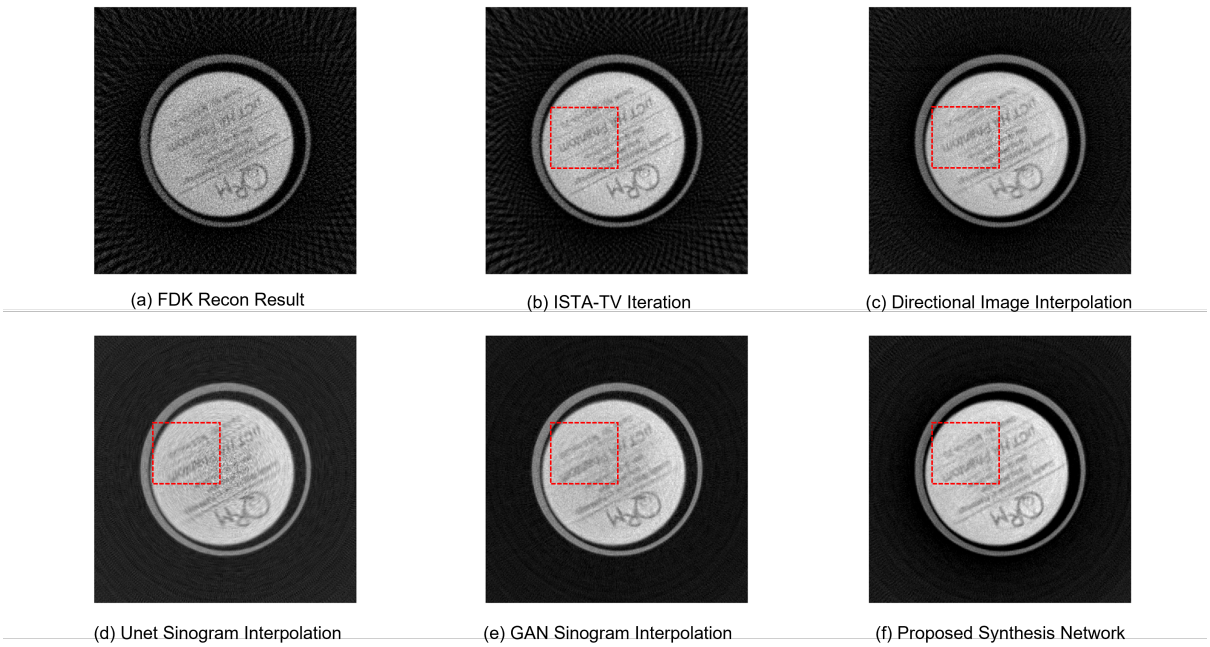


Figure 4.14: Reconstruction of QRM phantom in real-world scanning with the MiSpinner: (a) FDK result from 97 angles scanned, (b) Post-processing with classic MBIR method: ISTA-TV, (c) FDK result from 388 synthetic projections interpolated by classic structure tensor based image synthesis method, (d) FDK after U-net sinogram interpolation, (e) FDK after GAN sinogram interpolation, (d) FDK after Proj-syn-net projection synthesis.

4.4.3 DTL Learning

In supervised task A, the sparse reconstruction images of PM public data were fed into the network and the output compared with the corresponding ground truth image. As previously mentioned, we divided the image into 256×256 smaller patches. Some regions overlap since image size was not divisible by 256, so that the training data capacity reached a total of 10,266 images. 9293 of them are randomly assigned to a training dataset, and the remaining 973 are

assigned to a validation dataset. The ratio of the norm of the error between two images and the norm of the ground truth image was computed as the loss function, and the Adam optimizer was used to optimize the loss function. After feeding all the training data into the network, the network weights were updated with the objective to minimize the loss function. The process was repeated until the loss could not be further optimized, roughly about 130 epochs. The learning rate was set to be 0.01 initially. Trained weights were stored and preloaded into supervised task B network for transfer learning.

In supervised task B, the input was the measured CT phantom FDK result, and the output was compared with the corresponding TV IR result. In contrast to task A, we cropped the image to 256×256 and only 850 images were obtained for task B training. In order to get a reliable and stable model, we employed a 5-fold cross validation for network weights fine tuning, which accounts for 70% for training, 20% for validation, and 10% for testing of the whole dataset respectively. Mean square error was used as loss function, and the Adam optimizer was still used for optimization. The validation evaluation curve is shown in Fig. 4.15. The average validation SNR increased from 35.362 to 35.798 dB. Test evaluation shows that SNR increased from 33.1417 to 33.3750 dB.

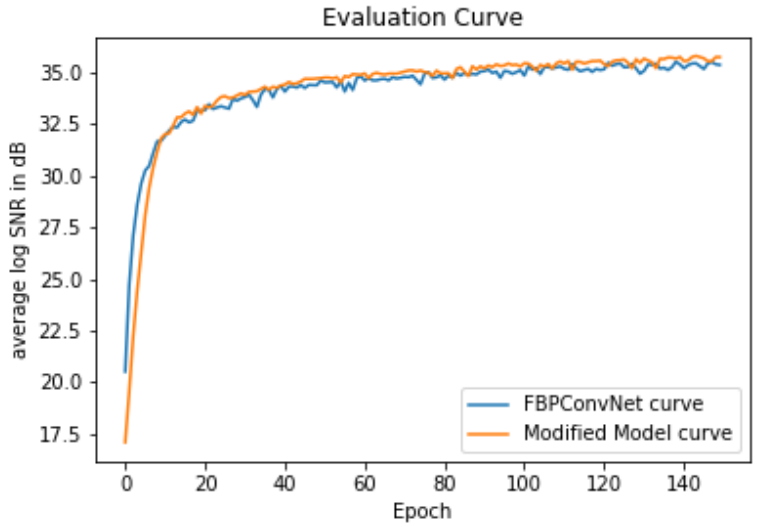


Figure 4.15: Signal-to-noise ratio estimation on FBPConvNet/U-net model and modified model in dB

Once measured projection data finished calibration, FDK reconstruction was conducted and then fed into the trained neural network for prediction. SIRT, CGLS, and TV IR methods were tested as well. A qualitative comparison of different algorithms on the same slice of the same mouse data is shown in Fig. 4.16. A rectangular area within the center region is zoomed in for detailed observation. For all iteration algorithms, their results restored the image features and maintained good contrast but also left the streak artifacts. For FBPConvNet, relevant results were oversmoothed and lost some detailed contrast in the soft tissue region. A good result was obtained through joint processing of the iteration algorithm and the trained network, shown in Fig. 4.16(f,i,l). Among these results, TV based iteration combined with network after DTL showed a good balance between feature contrast maintenance and artifact removal. The experimental results qualitatively verify the feasibility of the proposed DTL method and reveal that image restoration can potentially be significantly improved through a combination processing of a conventional iterative algorithm and a trained neural network.

In our study, the MBIR-related algorithm was realized through MATLAB, and the network-related realization was implemented with Python based on PyTorch. The training was done on a Google cloud workstation with 25GB RAM, and an online GPU was used to accelerate the training process.

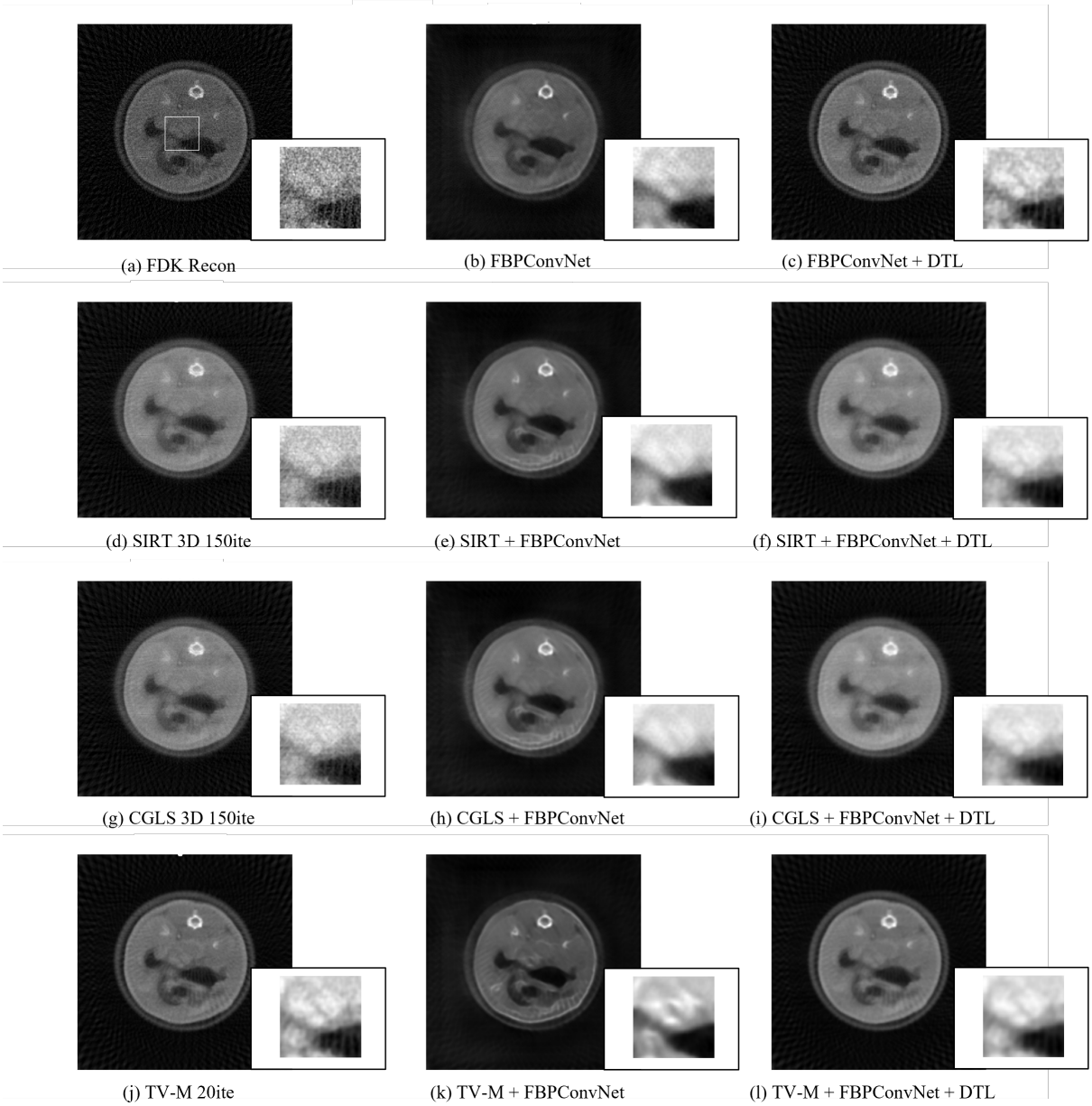


Figure 4.16: Comparison of various algorithms on mouse cross-sectional image restoration under sparse sampling. (a) FDK reconstruction on mouse data, which is also the raw input in iterative algorithms and trained network. The white rectangle region for each image is zoomed in for qualitative comparison. (b) output from the FBPConvNet trained by open-source data, which shows blur in the region. (c) output from the FBPConvNet + DTL, where artifacts remain. (d) result (a) after 150 SIRT iteration, while the contrast is worse than (c). (e) FBPConvNet output given input (d), where blur still exists. (f) FBPConvNet + DTL output after (d) that is better than (e) in contrast. (g) (a) after 150 CGLS iterations similar to (d). (h) result (g) corresponding FBPConvNet output. (i) result (g) corresponding FBPConvNet + DTL output that looks better than (f). (j) result (a) after 20 TV-M iteration, where artifacts remain but are an improvement over (d) and (g). (k) result (j) corresponding to FBPConvNet output, which shows an oversmoothed result. (l) result (j) corresponding to FBPConvNet + DTL output, which shows a balance between feature restoration and artifact region smoothing.

4.4.4 CycleGAN Exploring

Fig. 4.17 shows our network cycling training as Fig. 4.7 described. In this test, the input source domain images A are FDK results from actual scanning. The target domain images B are mouse micro-CT slices from outside the vet lab. A total of 6 mouse datasets, 3 from A and 3 from B, are used for network training. The image size equals 512×512 in pixels, and we set the network batch size to 2, epochs to 200, and learning rate to 0.0002. As Fig. 4.17 shows, the corresponding output A' of A not only possesses the same texture information but also maintains good contrast as well as B does.

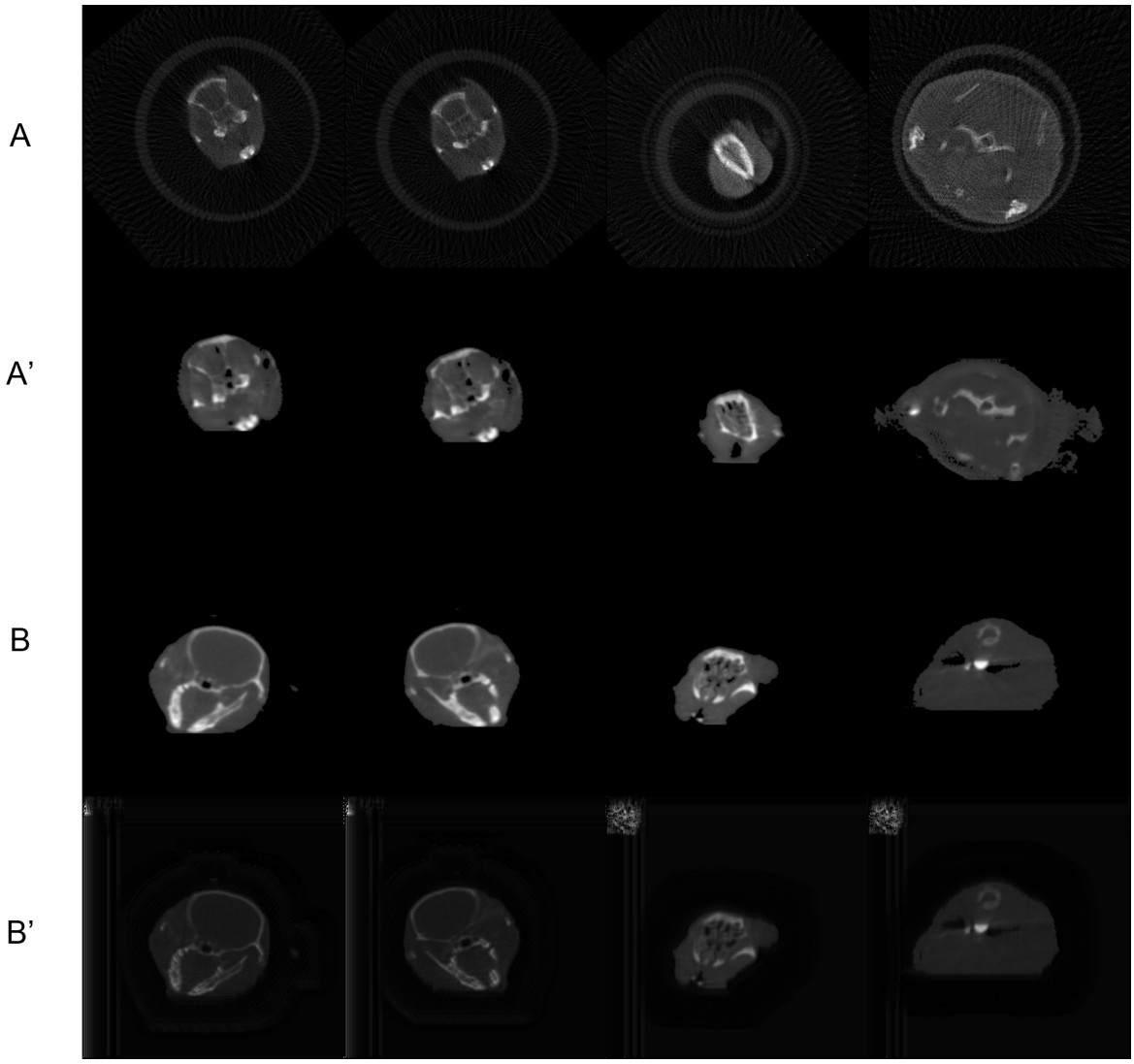


Figure 4.17: Reconstruction image quality enhancement by CycleGAN network under self-supervised learning. A is the source domain image, and B is the target domain. A' is the output of A in the target domain and B' is the output of B in the source domain.

The proposed method has shown some flaws in accuracy. We can see in the last two column results that it has introduced some extra image contents. However, it has shown potential to improve the reconstruction quality to a level that supervised network and MBIR methods cannot achieve in the absence of a more informed prior, such as clinical paired CT slices or accurate organ attenuation coefficient information.

4.5 Conclusion and Discussion

In this chapter, we have proposed a novel structure-tensor-based MBIR model, MAP-ST-TDV, for projection image pre-processing. Specifically, our model follows the MAP framework, which accounts for the photon propagation error and electronic background noise simultaneously, to fully utilize the projection data statistical properties for data denoising. Furthermore, we construct a new regulation term, ST-TDV, to better describe the image prior under low-dose imaging. Given that the structure tensor is capable of measuring the image geometry, we apply the eigenvector decomposition to find the direction in which the image intensity varies most. This enables our model to conduct an anisotropic diffusion processing on projection data to effectively reduce the staircase artifacts and maintain the high-frequency detail. An effective alternating algorithm, including ADMM and FOPD methods, was developed to solve the proposed model, whose performance has been demonstrated through a series of simulations in both 2D and 3D cases.

Moreover, we have proposed an adversarial learning-based spatial transformer network, termed Proj-syn-net, for addressing an issue in sparse image reconstruction without ground truth, which is common in new device development or in pre-clinical research. Our proposed network consists of a generator, which includes a 5-layer spatial transformer network and a bilinear sampler generator, and a discriminator including 3-layer convolution block. The Proj-syn-net is designed to improve reconstruction image quality by synthesizing extra intermediate projections, which is an approximation of the state-of-the-art DI method but with much less time required. In addition, this design enables a mixed training dataset, which combines the densely sampled phantom projections and sparsely sampled real scans. Such semi-supervised learning can effectively reduce the impact of device gap. Our experimental results have shown

that the proposed network can successfully recover the most image structure compared with other networks that possess sufficient ground truth imagery.

Furthermore, we have explored a DTL based neural network to improve reconstructed image quality given the lack of corresponding ground truth imagery. Our proposed method has revealed a possibility of cross-device information utilization and experimental results have demonstrated that a combination of an iterative algorithm and a DTL network can better restore the image contents compared to several baseline methods. Further work will be focused beyond the post-processing method. We expect to further improve the DTL performance by incorporating a pre-processing method, e,g, sinogram interpolation, or some other sparse coding algorithm which is less sensitve to the noise pattern.

Chapter 5

Future Work

Chapters of this dissertation span the development of a new imaging system (IVIS with the *MiSpinner*), scanning geometry calibration (multi-level thresholding algorithm and ST-TOP online method), and image quality enhancement (MAP-ST-TDV algorithm for projection data denoising and Proj-syn-net for projection synthesis). Each of these contributions has helped our retrofit imaging system acquire better CT-like reconstruction images.

One of the most difficult matters for applying the multi-level thresholding algorithm in Chapter 2 is the accurate segmentation of the ROI and background region. This process is especially difficult for the low-dose case, where reconstructed linear attenuation coefficients are close to the background values. Since the publication of our intensity-based segmentation algorithm, DL-based NN methods have emerged that could provide accurate segmentation, which has shown potential application in our study [128, 129].

The ST-TSO calibration method introduced in Chapter 3 has been thoroughly analyzed under various conditions. For instance, some investigation of the effect of detector saturation has validated that our proposed algorithm can successfully predict the geometry parameters and reconstruct the image correctly in the face of saturation. However, a few limits are worth improving. First, ST-TSO can at most estimate five parameters simultaneously due to an intrinsic property of the projection data. That is, in our small-animal research, edge detection can only provide one-dimensional boundary distribution information. Lack of vertical prior information prevents us from estimating the remaining parameters. Further design of the imaging system, e.g., specimen holder with specific and simple mark, is worth investigating. In addition, our PM-based part, which realizes 2D-3D transformation, is based on the default hypothesis that

each projection is imaged on the focal plane without distortion. However, in practice, it will be often violated. It is worthwhile to perform a study on X-ray focal spots to further improve the algorithm accuracy [130]. Furthermore, a study on defocusing can provide another perspective for image depth information estimation, which can be used for SDD prediction.

Chapter 4 mainly focuses on the optimization of the reconstructed image quality. We have designed a sophisticated MBIR method for projection image denoising and a network for compensating sparse sampling. Although experiments verify the effectiveness of each method, the complete framework is complex. At the time of writing this dissertation, efforts have been made to simplify the complete processing, e.g., comprehensive Neural Network. Moreover, results have shown that the Proj-syn-method still falls short in SNR estimation with a classic NN. The error is mainly concentrated in the high-frequency edge area where we believe it is caused by the bilateral linear interpolation error or some remaining errors in the tensor block. Further work will be focused beyond the assumption that an interpolated projection is linearly interpolated by adjacent projections. We expect to further improve the spatial transformer network performance by incorporating 3D convolution kernels and a self-attention mechanism.

Bibliography

[1] HK Genant, K Engelke, and S Prevrhal. "Advanced CT bone imaging in osteoporosis". In: *Rheumatology* 47.suppl_4 (2008), pp. iv9–iv16.

[2] Roy D Altman et al. "Radiographic assessment of progression in osteoarthritis". In: *Arthritis & Rheumatism: Official Journal of the American College of Rheumatology* 30.11 (1987), pp. 1214–1225.

[3] Keith W Muir et al. "Imaging of acute stroke". In: *The Lancet Neurology* 5.9 (2006), pp. 755–768.

[4] Thierry F Vandamme. "Use of rodents as models of human diseases". In: *Journal of pharmacy & bioallied sciences* 6.1 (2014), p. 2.

[5] Lee W Goldman. "Principles of CT: radiation dose and image quality". In: *Journal of nuclear medicine technology* 35.4 (2007), pp. 213–225.

[6] Manuela C Felix et al. "Image-guided radiotherapy using a modified industrial micro-CT for preclinical applications". In: *PLoS One* 10.5 (2015), e0126246.

[7] Istvan Szanda et al. "National Electrical Manufacturers Association NU-4 performance evaluation of the PET component of the NanoPET/CT preclinical PET/CT scanner". In: *Journal of nuclear medicine* 52.11 (2011), pp. 1741–1747.

[8] Nicolas Aide et al. "High throughput static and dynamic small animal imaging using clinical PET/CT: potential preclinical applications". In: *European journal of nuclear medicine and molecular imaging* 37.5 (2010), pp. 991–1001.

[9] Benjamin L Franc et al. "Small-animal SPECT and SPECT/CT: important tools for preclinical investigation". In: *Journal of nuclear medicine* 49.10 (2008), pp. 1651–1663.

[10] Frank Verhaegen, Patrick Granton, and Erik Tryggestad. “Small animal radiotherapy research platforms”. In: *Physics in Medicine & Biology* 56.12 (2011), R55.

[11] F Edward Boas, Dominik Fleischmann, et al. “CT artifacts: causes and reduction techniques”. In: *Imaging Med* 4.2 (2012), pp. 229–240.

[12] Mehmet Bilgen. “Feasibility and merits of performing preclinical imaging on clinical radiology and nuclear medicine systems”. In: *International journal of molecular imaging* 2013 (2013).

[13] DP Clark and CT Badea. “Advances in micro-CT imaging of small animals”. In: *Physica Medica* 88 (2021), pp. 175–192.

[14] Che-Wei Liao et al. “Self-assembled micro-computed tomography for dental education”. In: *PloS one* 13.12 (2018), e0209698.

[15] David W Holdsworth and Michael M Thornton. “Micro-CT in small animal and specimen imaging”. In: *Trends in Biotechnology* 20.8 (2002), S34–S39.

[16] Donald F Swinehart. “The beer-lambert law”. In: *Journal of chemical education* 39.7 (1962), p. 333.

[17] Sigurdur Helgason and S Helgason. *The radon transform*. Vol. 2. Springer, 1980.

[18] M Grass, Th Köhler, and R Proksa. “3D cone-beam CT reconstruction for circular trajectories”. In: *Physics in Medicine & Biology* 45.2 (2000), p. 329.

[19] Lee A Feldkamp, Lloyd C Davis, and James W Kress. “Practical cone-beam algorithm”. In: *Josa a* 1.6 (1984), pp. 612–619.

[20] Dennis L Parker. “Optimal short scan convolution reconstruction for fan beam CT”. In: *Medical physics* 9.2 (1982), pp. 254–257.

[21] Anne Rougee et al. “Geometrical calibration for 3D x-ray imaging”. In: *Medical Imaging 1993: Image Capture, Formatting, and Display*. Vol. 1897. SPIE. 1993, pp. 161–169.

[22] Yuanzheng Meng, Hui Gong, and Xiaoquan Yang. “Online geometric calibration of cone-beam computed tomography for arbitrary imaging objects”. In: *IEEE transactions on medical imaging* 32.2 (2012), pp. 278–288.

[23] Dongxing Xi et al. “The study of reconstruction image quality resulting from geometric error in micro-CT system”. In: *2010 4th international conference on bioinformatics and biomedical engineering*. IEEE. 2010, pp. 1–4.

[24] I Vidal-Migallon et al. “Simulation of mechanical misalignments in a cone-beam micro-CT system”. In: *2008 IEEE Nuclear Science Symposium Conference Record*. IEEE. 2008, pp. 5007–5009.

[25] Youngbin Cho et al. “Accurate technique for complete geometric calibration of cone-beam computed tomography systems”. In: *Medical physics* 32.4 (2005), pp. 968–983.

[26] Hongkai Yang, Kejun Kang, and Yuxiang Xing. “Geometry calibration method for a cone-beam CT system”. In: *Medical physics* 44.5 (2017), pp. 1692–1706.

[27] Yi Sun et al. “A calibration method for misaligned scanner geometry in cone-beam computed tomography”. In: *Ndt & E International* 39.6 (2006), pp. 499–513.

[28] S Ouadah et al. “Self-calibration of cone-beam CT geometry using 3D–2D image registration”. In: *Physics in Medicine & Biology* 61.7 (2016), p. 2613.

[29] Tilman Donath, Felix Beckmann, and Andreas Schreyer. “Automated determination of the center of rotation in tomography data”. In: *JOSA A* 23.5 (2006), pp. 1048–1057.

[30] Shouping Zhu et al. “Automated motion correction for in vivo optical projection tomography”. In: *IEEE transactions on medical imaging* 31.7 (2012), pp. 1358–1371.

[31] Hengyong Yu and Ge Wang. “Data consistency based rigid motion artifact reduction in fan-beam CT”. In: *IEEE Transactions on Medical Imaging* 26.2 (2007), pp. 249–260.

[32] Nghia T Vo et al. “Reliable method for calculating the center of rotation in parallel-beam tomography”. In: *Optics express* 22.16 (2014), pp. 19078–19086.

[33] Di Dong et al. “Automated recovery of the center of rotation in optical projection tomography in the presence of scattering”. In: *IEEE journal of biomedical and health informatics* 17.1 (2012), pp. 198–204.

[34] Y Kyriakou et al. “Simultaneous misalignment correction for approximate circular cone-beam computed tomography”. In: *Physics in Medicine & Biology* 53.22 (2008), p. 6267.

[35] Yixing Huang et al. “Restoration of missing data in limited angle tomography based on Helgason–Ludwig consistency conditions”. In: *Biomedical Physics & Engineering Express* 3.3 (2017), p. 035015.

[36] Yudell L Luke. *Integrals of Bessel functions*. Courier Corporation, 2014.

[37] Paul R Edholm, Robert M Lewitt, and Bernt Lindholm. “Novel properties of the Fourier decomposition of the sinogram”. In: *Physics and Engineering of Computerized Multidimensional Imaging and Processing*. Vol. 671. SPIE. 1986, pp. 8–18.

[38] WG Hawkins, PK Leichner, and N-C Yang. “The circular harmonic transform for SPECT reconstruction and boundary conditions on the Fourier transform of the sinogram”. In: *IEEE transactions on medical imaging* 7.2 (1988), pp. 135–138.

[39] John Canny. “A computational approach to edge detection”. In: *IEEE Transactions on pattern analysis and machine intelligence* 6 (1986), pp. 679–698.

[40] GT Shrivakshan and Chandramouli Chandrasekar. “A comparison of various edge detection techniques used in image processing”. In: *International Journal of Computer Science Issues (IJCSI)* 9.5 (2012), p. 269.

[41] George Stockman and Linda G Shapiro. *Computer vision*. Prentice Hall PTR, 2001.

[42] Konstantinos G Derpanis. “Overview of the RANSAC Algorithm”. In: *Image Rochester NY* 4.1 (2010), pp. 2–3.

[43] Johnathon R Walls et al. “Correction of artefacts in optical projection tomography”. In: *Physics in Medicine & Biology* 50.19 (2005), p. 4645.

[44] Chumin Zhao et al. “Cone-beam imaging with tilted rotation axis: Method and performance evaluation”. In: *Medical physics* 47.8 (2020), pp. 3305–3320.

[45] Richard Gordon, Gabor T Herman, and Steven A Johnson. “Image reconstruction from projections”. In: *Scientific American* 233.4 (1975), pp. 56–71.

[46] Mohammad Abedin-Nasab. *Handbook of robotic and image-guided surgery*. Elsevier, 2019.

[47] Jeffrey C Lagarias et al. “Convergence properties of the Nelder–Mead simplex method in low dimensions”. In: *SIAM Journal on optimization* 9.1 (1998), pp. 112–147.

[48] Stamatios Lefkimmiatis et al. “Structure tensor total variation”. In: *SIAM Journal on Imaging Sciences* 8.2 (2015), pp. 1090–1122.

[49] Xiaoquan Yang et al. “Abnormal pixel detection using sum-of-projections symmetry in cone beam computed tomography”. In: *Optics express* 20.10 (2012), pp. 11014–11030.

[50] Peter JM Van Laarhoven and Emile HL Aarts. “Simulated annealing”. In: *Simulated annealing: Theory and applications*. Springer, 1987, pp. 7–15.

[51] Andreas Efstratiadis and Demetris Koutsoyiannis. “An evolutionary annealing-simplex algorithm for global optimisation of water resource systems”. In: *Proceedings of the Fifth International Conference on Hydroinformatics*. 2002, pp. 1423–1428.

[52] David R Dance. “Diagnostic Radiology Physics A handbook for teachers and students”. In: (2014).

[53] Andriy Fedorov et al. “3D Slicer as an image computing platform for the Quantitative Imaging Network”. In: *Magnetic resonance imaging* 30.9 (2012), pp. 1323–1341.

[54] Elizabeth B Gausden et al. “Opportunistic use of CT imaging for osteoporosis screening and bone density assessment: a qualitative systematic review”. In: *Jbjs* 99.18 (2017), pp. 1580–1590.

[55] So Jung Lee et al. “A prospective comparison of standard-dose CT enterography and 50% reduced-dose CT enterography with and without noise reduction for evaluating Crohn disease”. In: *American Journal of Roentgenology* 197.1 (2011), pp. 50–57.

[56] Sajid Abbas et al. “Effects of sparse sampling schemes on image quality in low-dose CT”. In: *Medical physics* 40.11 (2013), p. 111915.

[57] Cesar Molinos et al. “Low-dose imaging in a new preclinical total-body PET/CT scanner”. In: *Frontiers in medicine* 6 (2019), p. 88.

[58] Qi Xie et al. “Robust low-dose CT sinogram preprocessing via exploiting noise-generating mechanism”. In: *IEEE transactions on medical imaging* 36.12 (2017), pp. 2487–2498.

[59] Muralidhar Mupparapu et al. “Development and application of a novel nanophosphor scintillator for a low-dose, high-resolution digital X-ray imaging system”. In: *International Congress Series*. Vol. 1281. Elsevier. 2005, pp. 1256–1261.

[60] Ruola Ning, Xiangyang Tang, and David Conover. “X-ray scatter correction algorithm for cone beam CT imaging”. In: *Medical physics* 31.5 (2004), pp. 1195–1202.

[61] Jing Wang et al. “Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose X-ray computed tomography”. In: *IEEE transactions on medical imaging* 25.10 (2006), pp. 1272–1283.

[62] Xuehang Zheng et al. “PWLS-ULTRA: An efficient clustering and learning-based approach for low-dose 3D CT image reconstruction”. In: *IEEE transactions on medical imaging* 37.6 (2018), pp. 1498–1510.

[63] Jeffrey A Fessler. “Penalized weighted least-squares image reconstruction for positron emission tomography”. In: *IEEE transactions on medical imaging* 13.2 (1994), pp. 290–300.

[64] Yan Liu et al. “Adaptive-weighted total variation minimization for sparse data toward low-dose x-ray computed tomography image reconstruction”. In: *Physics in Medicine & Biology* 57.23 (2012), p. 7923.

[65] Yi Liu et al. “Median prior constrained TV algorithm for sparse view low-dose CT reconstruction”. In: *Computers in biology and medicine* 60 (2015), pp. 117–131.

[66] Tianfang Li et al. “Nonlinear sinogram smoothing for low-dose X-ray CT”. In: *IEEE Transactions on Nuclear Science* 51.5 (2004), pp. 2505–2513.

[67] Shanzhou Niu et al. “Sparse-view x-ray CT reconstruction via total generalized variation regularization”. In: *Physics in Medicine & Biology* 59.12 (2014), p. 2997.

[68] Emil Y Sidky et al. “A constrained, total-variation minimization algorithm for low-intensity x-ray CT”. In: *Medical physics* 38.S1 (2011), S117–S125.

[69] Triet Le, Rick Chartrand, and Thomas J Asaki. “A variational approach to reconstructing images corrupted by Poisson noise”. In: *Journal of mathematical imaging and vision* 27.3 (2007), pp. 257–263.

[70] Patrick J La Riviere. “Penalized-likelihood sinogram smoothing for low-dose CT”. In: *Medical physics* 32.6Part1 (2005), pp. 1676–1683.

[71] Patrick J La Rivière, Junguo Bian, and Phillip A Vargas. “Penalized-likelihood sinogram restoration for computed tomography”. In: *IEEE transactions on medical imaging* 25.8 (2006), pp. 1022–1036.

[72] Leonid I Rudin, Stanley Osher, and Emad Fatemi. “Nonlinear total variation based noise removal algorithms”. In: *Physica D: nonlinear phenomena* 60.1-4 (1992), pp. 259–268.

[73] Dang Thanh, Prasath Surya, et al. “A review on CT and X-ray images denoising methods”. In: *Informatica* 43.2 (2019).

[74] Julian Besag. “On the statistical analysis of dirty pictures”. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 48.3 (1986), pp. 259–279.

[75] Kevin J Little and Patrick J La Rivière. “Sinogram restoration in computed tomography with a non-quadratic, edge-preserving penalty”. In: *2011 IEEE Nuclear Science Symposium Conference Record*. IEEE. 2011, pp. 2534–2536.

[76] Aram Danielyan, Vladimir Katkovnik, and Karen Egiazarian. “BM3D frames and variational image deblurring”. In: *IEEE Transactions on image processing* 21.4 (2011), pp. 1715–1728.

[77] Hao Zhang et al. “Applications of nonlocal means algorithm in low-dose X-ray CT image processing and reconstruction: A review”. In: *Medical physics* 44.3 (2017), pp. 1168–1185.

[78] Michael T. McCann Kyong HwanJin. “Deep Convolutional Neural Network for Inverse Problems in Imaging”. In: *IEEE Transactions on Image Processing* 26(9) (2017), pp. 4509–4522. DOI: 10.1109/TIP.2017.2713099.

[79] Haoyu Wei et al. “2-Step Sparse-View CT Reconstruction with a Domain-Specific Perceptual Network”. In: *arXiv preprint arXiv:2012.04743* (2020).

[80] Shipeng Xie and Tao Yang. “Artifact removal in sparse-angle CT based on feature fusion residual network”. In: *IEEE Transactions on Radiation and Plasma Medical Sciences* 5.2 (2020), pp. 261–271.

[81] Zongwei Zhou et al. “Unet++: A nested u-net architecture for medical image segmentation”. In: *Deep learning in medical image analysis and multimodal learning for clinical decision support*. Springer, 2018, pp. 3–11.

[82] Hu Chen et al. “Low-dose CT with a residual encoder-decoder convolutional neural network”. In: *IEEE transactions on medical imaging* 36.12 (2017), pp. 2524–2535.

[83] Zhicheng Zhang et al. “A sparse-view CT reconstruction method based on combination of DenseNet and deconvolution”. In: *IEEE transactions on medical imaging* 37.6 (2018), pp. 1407–1417.

[84] Hu Chen et al. “LEARN: Learned experts’ assessment-based reconstruction network for sparse-data CT”. In: *IEEE transactions on medical imaging* 37.6 (2018), pp. 1333–1347.

[85] Xiaogang Yang et al. “Tomographic reconstruction with a generative adversarial network”. In: *Journal of synchrotron radiation* 27.2 (2020), pp. 486–493.

[86] Tobias Würfl et al. “Deep learning computed tomography”. In: *International conference on medical image computing and computer-assisted intervention*. Springer. 2016, pp. 432–440.

[87] Tobias Würfl et al. “Deep learning computed tomography: Learning projection-domain weights from image domain in limited angle problems”. In: *IEEE transactions on medical imaging* 37.6 (2018), pp. 1454–1463.

[88] Xu Dong, Swapnil Vekhande, and Guohua Cao. “Sinogram interpolation for sparse-view micro-CT with deep learning neural network”. In: *Medical Imaging 2019: Physics of Medical Imaging*. Vol. 10948. International Society for Optics and Photonics. 2019, 109482O.

[89] Ziheng Li et al. “A sinogram inpainting method based on generative adversarial network for limited-angle computed tomography”. In: *15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*. Vol. 11072. International Society for Optics and Photonics. 2019, p. 1107220.

[90] Huizhuo Yuan, Jinzhu Jia, and Zhanxing Zhu. “Sipid: A deep learning framework for sinogram interpolation and image denoising in low-dose CT reconstruction”. In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE. 2018, pp. 1521–1524.

[91] Kaichao Liang et al. “A Model-Based Unsupervised Deep Learning Method for Low-Dose CT Reconstruction”. In: *IEEE Access* 8 (2020), pp. 159260–159273.

[92] Dong Zeng et al. “Noise-Generating-Mechanism-Driven Unsupervised Learning for Low-Dose CT Sinogram Recovery”. In: *IEEE Transactions on Radiation and Plasma Medical Sciences* (2021).

[93] Mingqiang Meng et al. “Semi-supervised learned sinogram restoration network for low-dose CT image reconstruction”. In: *Medical Imaging 2020: Physics of Medical Imaging*. Vol. 11312. International Society for Optics and Photonics. 2020, 113120B.

[94] Esmaeil Enjilela et al. “Cubic-Spline Interpolation for Sparse-View CT Image Reconstruction With Filtered Backprojection in Dynamic Myocardial Perfusion Imaging”. In: *Tomography* 5.3 (2019), pp. 300–307.

[95] Gengsheng L Zeng. “Sparse-view tomography via displacement function interpolation”. In: *Visual computing for industry, biomedicine, and art* 2.1 (2019), pp. 1–10.

[96] Yinsheng Li et al. “Strategy of computed tomography sinogram inpainting based on sinusoid-like curve decomposition and eigenvector-guided interpolation”. In: *JOSA A* 29.1 (2012), pp. 153–163.

[97] Hua Zhang, Matthijs Kruis, and Jan-Jakob Sonke. “Directional sinogram interpolation for motion weighted 4D cone-beam CT reconstruction”. In: *Physics in Medicine & Biology* 62.6 (2017), p. 2254.

[98] Matthias Bertram et al. “Directional view interpolation for compensation of sparse angular sampling in cone-beam CT”. In: *IEEE transactions on medical imaging* 28.7 (2009), pp. 1011–1022.

[99] Ziwei Liu et al. “Video frame synthesis using deep voxel flow”. In: *Proceedings of the IEEE International Conference on Computer Vision*. 2017, pp. 4463–4471.

[100] Ilker Bayram and Mustafa E Kamasak. “A directional total variation”. In: *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*. IEEE. 2012, pp. 265–269.

[101] Simone Parisotto et al. “Higher-order total directional variation: Imaging applications”. In: *SIAM Journal on Imaging Sciences* 13.4 (2020), pp. 2063–2104.

[102] Virginia Estellers, Stefano Soatto, and Xavier Bresson. “Adaptive regularization with the structure tensor”. In: *IEEE Transactions on Image Processing* 24.6 (2015), pp. 1777–1790.

[103] Vardan Papyan, Yaniv Romano, and Michael Elad. “Convolutional neural networks analyzed via convolutional sparse coding”. In: *The Journal of Machine Learning Research* 18.1 (2017), pp. 2887–2938.

[104] Yuejie Chi and Yue M Lu. “Kaczmarz method for solving quadratic equations”. In: *IEEE Signal Processing Letters* 23.9 (2016), pp. 1183–1187.

[105] Wei Xu and Klaus Mueller. “Learning effective parameter settings for iterative CT reconstruction algorithms”. In: *Workshop on High Performance Image Reconstruction (HPIR)*. Citeseer. 2009.

[106] James Joyce. “Bayes’ theorem”. In: (2003).

[107] Gene H Golub, Per Christian Hansen, and Dianne P O’Leary. “Tikhonov regularization and total least squares”. In: *SIAM journal on matrix analysis and applications* 21.1 (1999), pp. 185–194.

[108] Yonina C Eldar and Gitta Kutyniok. *Compressed sensing: theory and applications*. Cambridge university press, 2012.

[109] Yaakov Tsaig and David L Donoho. “Extensions of compressed sensing”. In: *Signal processing* 86.3 (2006), pp. 549–571.

[110] Emil Y Sidky, Mark A Anastasio, and Xiaochuan Pan. “Image reconstruction exploiting object sparsity in boundary-enhanced X-ray phase-contrast tomography”. In: *Optics express* 18.10 (2010), pp. 10404–10422.

[111] Mohammad Toghi et al. “Projection Onto Convex Sets (POCS) Based Signal Reconstruction Framework with an associated cost function”. In: *arXiv preprint arXiv:1402.2088* (2014).

[112] Antonin Chambolle et al. “An introduction to total variation for image analysis”. In: *Theoretical foundations and numerical methods for sparse recovery* 9.263-340 (2010), p. 227.

[113] Tarmizi Adam, Mohd Fikree Hassan, and Raveendran Paramesran. “A Study on Staircase Artifacts in Total Variation Image Restoration”. In: *2021 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*. IEEE. 2021, pp. 83–88.

[114] Kristian Bredies, Karl Kunisch, and Thomas Pock. “Total generalized variation”. In: *SIAM Journal on Imaging Sciences* 3.3 (2010), pp. 492–526.

[115] Simone Parisotto et al. “Higher-order total directional variation: Imaging applications”. In: *SIAM Journal on Imaging Sciences* 13.4 (2020), pp. 2063–2104.

[116] Charles Fox. *An introduction to the calculus of variations*. Courier Corporation, 1987.

[117] Amir Beck and Marc Teboulle. “Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems”. In: *IEEE transactions on image processing* 18.11 (2009), pp. 2419–2434.

[118] Stephen Boyd et al. “Distributed optimization and statistical learning via the alternating direction method of multipliers”. In: *Foundations and Trends® in Machine learning* 3.1 (2011), pp. 1–122.

[119] Antonin Chambolle and Thomas Pock. “A first-order primal-dual algorithm for convex problems with applications to imaging”. In: *Journal of mathematical imaging and vision* 40.1 (2011), pp. 120–145.

[120] Ingrid Daubechies, Michel Defrise, and Christine De Mol. “An iterative thresholding algorithm for linear inverse problems with a sparsity constraint”. In: *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences* 57.11 (2004), pp. 1413–1457.

[121] Oludare Isaac Abiodun et al. “State-of-the-art in artificial neural network applications: A survey”. In: *Heliyon* 4.11 (2018), e00938.

[122] Grace W Lindsay. “Convolutional neural networks as a model of the visual system: Past, present, and future”. In: *Journal of cognitive neuroscience* 33.10 (2021), pp. 2017–2031.

[123] Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. “Spatial transformer networks”. In: *Advances in neural information processing systems* 28 (2015), pp. 2017–2025.

[124] Chuanqi Tan et al. “A survey on deep transfer learning”. In: *International conference on artificial neural networks*. Springer. 2018, pp. 270–279.

[125] Tero Karras et al. “Analyzing and improving the image quality of stylegan”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020, pp. 8110–8119.

[126] Jun-Yan Zhu et al. “Unpaired image-to-image translation using cycle-consistent adversarial networks”. In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2223–2232.

[127] Wim Van Aarle et al. “Fast and flexible X-ray tomography using the ASTRA toolbox”. In: *Optics express* 24.22 (2016), pp. 25129–25147.

[128] Wentao Zhu et al. “AnatomyNet: deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy”. In: *Medical physics* 46.2 (2019), pp. 576–589.

[129] Johannes Hofmanninger et al. “Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem”. In: *European Radiology Experimental* 4.1 (2020), pp. 1–13.

[130] Linxi Shi, N Robert Bennett, and Adam S Wang. “Characterization of x-ray focal spots using a rotating edge”. In: *Journal of Medical Imaging* 8.2 (2021), p. 023502.

Appendices

Appendix A

Multi-level Thresholding Algorithm for 2D Online Calibration

Let us consider the HLCC in the fan-beam case. As Fig. 2.3 shows, following the Beam Rebinning Strategy, we got $g_F(\beta, u) = g_P(\theta, l)$. Then projection data can be transferred as

$$\begin{aligned} J_k(\theta) &= \int l^k g_p(\theta, l) dl = \int l^k g_F(\beta, u) dl|_{l=\frac{Ru}{\sqrt{D^2+u^2}}, \beta=\theta-\arctan\frac{u}{D}}, \\ J_k(\theta) &= \int_{-\infty}^{\infty} g_F(\theta - \arctan\frac{u}{D}, u)(\frac{Ru}{\sqrt{D^2+u^2}})^k \frac{RD^2}{(D^2+u^2)^{3/2}} du. \end{aligned} \tag{A.1}$$

Considering object movement during measurement, i.e., $d(t) = (d_1(t), d_2(t))$ varies with time variable t, then distance variable l becomes

$$\bar{l} = \frac{Ru}{\sqrt{D^2+u^2}} - d_1(t)\cos\theta - d_2(t)\sin\theta. \tag{A.2}$$

For simplicity, assuming detector rotation follows constant speed, that is, incident angle $\lambda = \omega t$, then we got

$$\begin{aligned} t =& \frac{1}{\omega}(\theta + \frac{\pi}{2} - \arctan\frac{u}{D}), \\ J_k(\theta) = & \int_{-\infty}^{\infty} g_F(\theta - \arctan\frac{u}{D}, u)(\frac{Ru}{\sqrt{D^2+u^2}} - d_1(t)\cos\theta - d_2(t)\sin\theta)^k \\ & \times(\frac{RD^2}{(D^2+u^2)^{3/2}} + \frac{D(d_1'(t)\cos\theta + d_2'(t)\sin\theta)}{\omega(D^2+u^2)})du, \end{aligned} \tag{A.3}$$

where θ can be replaced by fanbeam view angle parameter β without loss of generality, and $\beta = \omega t - \frac{\pi}{2}$. Thus we got the HLCC in fanbeam case in regard to the parameter t as

$$\begin{aligned} J_k(\theta) = & \int_{-\infty}^{\infty} g_F(\omega t - \frac{\pi}{2} - \arctan\frac{u}{D}, u)(\frac{Ru}{\sqrt{D^2+u^2}} - d_1(\tilde{t})\sin(\omega t) + d_2(\tilde{t})\cos(\omega t))^k \\ & \times(\frac{RD^2}{(D^2+u^2)^{3/2}} + \frac{D(d_1'(\tilde{t})\sin(\omega t) - d_2'(\tilde{t})\cos(\omega t))}{\omega(D^2+u^2)})du|_{\tilde{t}=t-\frac{1}{\omega}\arctan\frac{u}{D}}. \end{aligned} \tag{A.4}$$

In short, its numerical format is close to

$$J_k(t) = \sum_{r=0}^{k} \frac{k!}{r!(k-r)!} m_{r,k-r}(-1)^{k-r} \cos^r(\omega t) \sin^{k-r}(\omega t). \tag{A.5}$$

Similar processing procedure can be found in Yu *et al.* [31].

Recalling that our aim is to estimate movement $d(t)$, we initialize $d(0) = 0$, and decompose $d(t)$ into N-order polynomial:

$$\begin{cases} d_1(t) = p_{1,0} + p_{1,1}t + ... + p_{1,N}t^N, & p_{1,0} = 0 \\ d_2(t) = p_{2,0} + p_{2,1}t + ... + p_{2,N}t^N, & p_{2,0} = 0 \end{cases}, \tag{A.6}$$

where $1, t, t^2, ...t^N$ is an example of N-order basis in Hilbert Space. Let $k = 0$, then HLCC becomes

$$C_0(t) - \sum_{n=0}^{N} C_{1,n}(t)p_{1,n} - \sum_{n=0}^{N} C_{2,n}(t)p_{2,n} = m_{0,0}, \tag{A.7}$$

where

$$\begin{aligned} C_0(t) &= \int_{-\infty}^{\infty} g_F(\omega t - \frac{\pi}{2} - \arctan\frac{u}{D}, u) \frac{RD^2}{(D^2+u^2)^{3/2}} du \\ C_1(t) &= -\sin(\omega t) \int_{-\infty}^{\infty} g_F(\omega t - \frac{\pi}{2} - \arctan\frac{u}{D}, u) \varphi_n^{'}(t - \frac{1}{\omega}\arctan\frac{u}{D}) \frac{D}{\omega(D^2+u^2)} du \\ C_2(t) &= \cos(\omega t) \int_{-\infty}^{\infty} g_F(\omega t - \frac{\pi}{2} - \arctan\frac{u}{D}, u) \varphi_n^{'}(t - \frac{1}{\omega}\arctan\frac{u}{D}) \frac{D}{\omega(D^2+u^2)} du, \end{aligned} \tag{A.8}$$

and

$$\begin{cases} d_1(0) = p_{1,0}\varphi_0(0) + p_{1,1}\varphi_1(0) + ... + p_{1,N}\varphi_N(0) = 0 \\ d_2(0) = p_{2,0}\varphi_0(0) + p_{2,1}\varphi_1(0) + ... + p_{2,N}\varphi_N(0) = 0 \end{cases}. \tag{A.9}$$

Therefore, Eq. (A.7) can be formulated in matrix format as

$$Q = MP, \tag{A.10}$$

where

$$\begin{aligned} M &= \begin{bmatrix} \varphi_0(0) & \cdots & \varphi_N(0) & 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & \varphi_0(0) & \cdots & \varphi_N(0) & 0 \\ C_{1,0}(t_0) & \cdots & C_{1,N}(t_0) & C_{2,0}(t_0) & \cdots & C_{2,N}(t_0) & 1 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ C_{1,0}(t_J) & \cdots & C_{1,N}(t_J) & C_{2,0}(t_J) & \cdots & C_{2,N}(t_J) & 1 \end{bmatrix}, \\ P &= \begin{bmatrix} p_{1,0} & \cdots & p_{1,N} & p_{2,0} & \cdots & p_{2,N} & m_{0,0} \end{bmatrix}^T, \\ Q &= \begin{bmatrix} 0 & 0 & c_0(t_0) & c_0(t_1) & \cdots & c_0(t_J) \end{bmatrix}^T. \end{aligned} \tag{A.11}$$

Normally, J is larger than $2N$, thus, we solve Eq. (A.10) by pseudoinversing

$$P = (M^T M)^{-1} M^T Q. \tag{A.12}$$

More complicated, if we consider the rotation effect during the measurement, where incident angle $\bar{\lambda} = \omega t - \phi(t)$ is a function of t, then analytical solution becomes unreachable. Hence, an iterative scheme with designed cost function is applied to find the solution,

$$\begin{aligned} E(k,t) &= v_k(t) - \sum_{r=0}^{k} \frac{k!}{r!(k-r)!}(-1)^{k-r} m_{r,k-r} \sin^r(\bar{\lambda(t)}) \cos^{k-r}(\bar{\lambda(t)}), \\ P &= \arg\min \sum_{k=0}^{K} (\omega(k) \int E^2(k,t)dt), \end{aligned} \tag{A.13}$$

where $\omega(k)$ is a weighting function equals one.

Appendix B

Structure Tensor based Two-step Online Method for System Geometry Calibration

In this material, we will present analytical solutions to some simplified cases. Fig. B.1 shows an object-biased cone beam geometry in CT scanning. In our experiment, detector plane $(u-v)$ and X-ray source (S) are fixed, object-rotated device is movable. Therefore, assuming X-ray source projected point S' as the origin and building upon a world coordinate system as Fig. B.1 shows. Given measured distance SDD and SOD, we have X-ray source coordinates $S(0,0,SDD)$ and object center coordinates $O(m,n,SDD-SOD)$, where m is bias in X-direction and n is bias in Y-direction, m, n are not equal 0. For object $(abcd)$ in cylinder shape, assume its length equals $2l$ and radius equals r.

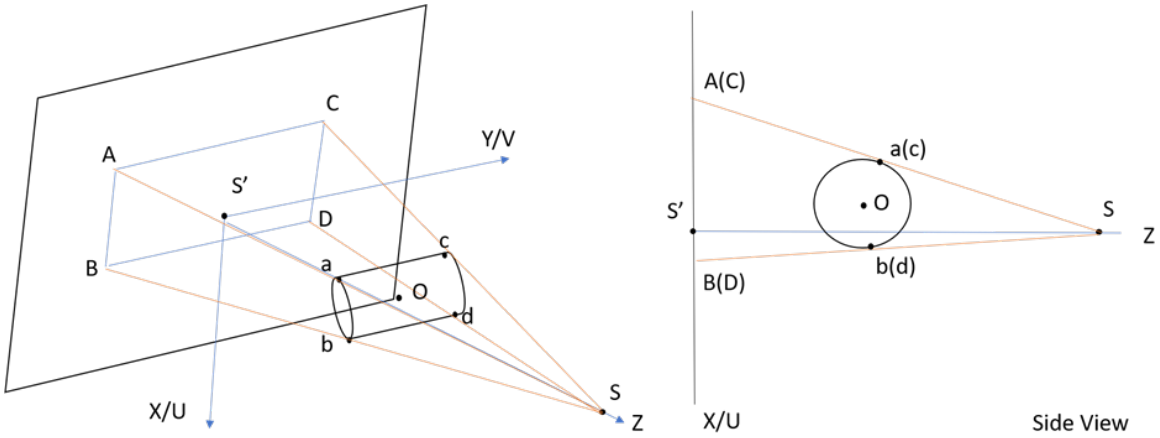


Figure B.1: Demonstration of scanning geometry

Let $z_1 = SDD - SOD$, $z_2 = SDD$, then in x-z plane, Fig. B.1 right part, object center O is (m, z_1) and four points on tangent boundaries, a, b, c, d, they have such relation:

$$\begin{cases} (x-m)^2 + (z-z_1)^2 = r^2 \\ \frac{x-m}{z-z_1}\frac{x}{z-z_2} = -1 \end{cases}. \tag{B.1}$$

Let $Z = z - z_1$, $c = z_2 - z_1$, we can infer the solution

$$\begin{aligned} z1 &= z_1 + \frac{r^2 c + mr\sqrt{m^2+c^2-r^2}}{m^2+c^2}, x1 = \frac{c(z1-z_1)+m^2-r^2}{m} \\ z2 &= z_1 + \frac{r^2 c - mr\sqrt{m^2+c^2-r^2}}{m^2+c^2}, x2 = \frac{c(z2-z_1)+m^2-r^2}{m}. \end{aligned} \tag{B.2}$$

where $\sqrt{m^2+c^2-r^2}$ is the length of the line segment $|Sa(c)| > 0$, thus $z1 > z2$. According to Fig. B.1, we have a, b, c, d coordinates as

$$a(x1, b-l, z1), b(x2, b-l, z2), c(x1, b+l, z1), d(x2, b+l, z2), \tag{B.3}$$

where b is center O coordinates in Y-direction. For $m > 0$ case,

$$a(x2, b-l, z2), b(x1, b-l, z1), c(x2, b+l, z2), d(x1, b+l, z1). \tag{B.4}$$

Assume in-plane rotation angle error, η, in x-y plane, and it rotates as Fig. B.2 shows, a clockwise rotation from dotted line to solid line. Then we have

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos\eta & \sin\eta \\ -\sin\eta & \cos\eta \end{bmatrix} \begin{bmatrix} x-m \\ y-n \end{bmatrix} + \begin{bmatrix} m \\ n \end{bmatrix}. \tag{B.5}$$

Thus points after rotation noted as a', b', c', d',

$$\begin{aligned} &a'((x1-m)\cos\eta - l\sin\eta + m, -(x1-m)\sin\eta - l\cos\eta + n, z1), \\ &b'((x2-m)\cos\eta - l\sin\eta + m, -(x2-m)\sin\eta - l\cos\eta + n, z2), \\ &c'((x1-m)\cos\eta + l\sin\eta + m, -(x1-m)\sin\eta + l\cos\eta + n, z1), \\ &d'((x2-m)\cos\eta + l\sin\eta + m, -(x2-m)\sin\eta + l\cos\eta + n, z2). \end{aligned} \tag{B.6}$$

and their corresponding projection onto detector plane, A, B, C and D as Fig. B.1 left shows, can be calculated through vector method

$$P_i = k_i\overrightarrow{Sp_i} + S, i = a', b', c', d', \tag{B.7}$$

where k_i is scaling factor and can be solved by

$$k_{a'} = k_{c'} = \frac{z_2}{z_2 - z1} = k_1, k_{b'} = k_{d'} = \frac{z_2}{z_2 - z2} = k_2, k_1 > k_2. \tag{B.8}$$

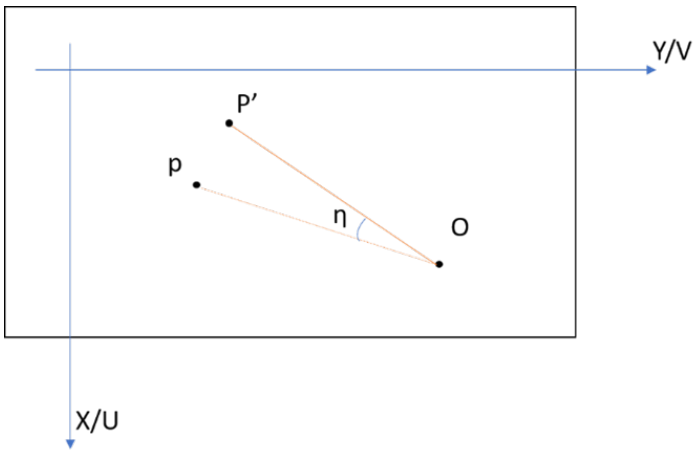


Figure B.2: Sketch of in-plain rotation

Then each projected point, A, B, C and D, its coordinate could be represented as

$$A(k_1 x_{a'}, k_1 y_{a'}, 0), B(k_2 x_{b'}, k_2 y_{b'}, 0), C(k_1 x_{c'}, k_1 y_{c'}, 0), D(k_2 x_{d'}, k_2 y_{d'}, 0). \tag{B.9}$$

We can observe that

$$AC = k_1 \times ac = k_1 \times (2l \sin\eta, 2l \cos\eta, 0), BD = k_2 \times bd = k_2 \times (2l \sin\eta, 2l \cos\eta, 0). \tag{B.10}$$

Thus, we can draw the conclusion that if central ray of X-ray source is perpendicular to the object and the detector plane, then in-plane rotation won't affect projection boundaries' parallel character. Therefore, if we can detect boundary information from the projection image, we can measure the in-plane rotation degree directly by

$$\eta = \arctan(\frac{x}{y}), \tag{B.11}$$

where x stands for X-axis component, y stands for Y-axis component.

Case 1: Only Tilt(θ) error, No in-plane rotation($\eta = 0$), no bias($m, n = 0$):
Tilt is a rotation in y-z plane around X-axis, as Fig. B.3 shows, denoted as θ. Due to the existence of the tilt, we can observe that projection image of a cylinder is no longer "rectangle" shape.

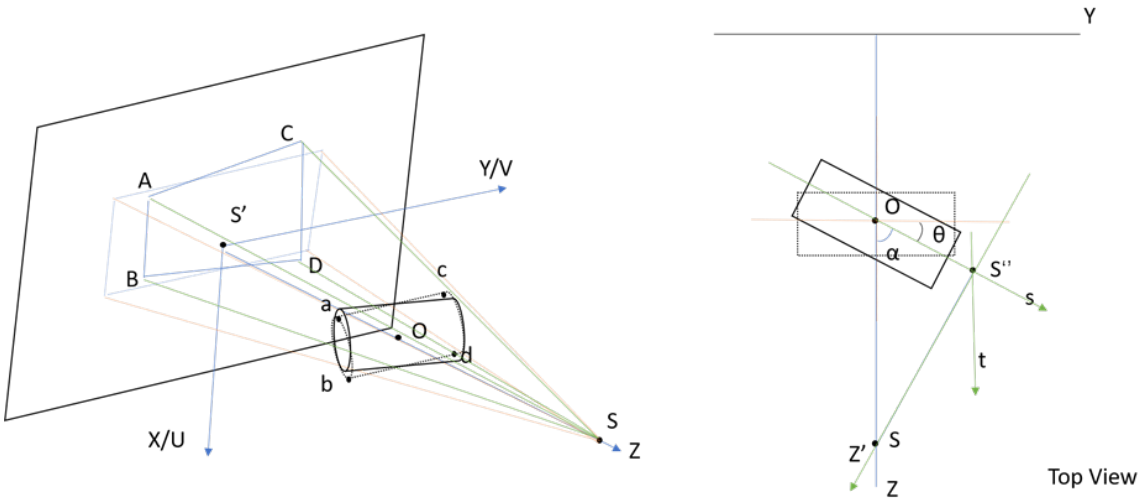


Figure B.3: Sketch of object cone beam scanning with Tilt rotation only

In order to determine the location of four points on the projection image boundaries, we build a local coordinate system as Fig. B.3 right shows, whose origin is the X-ray source point projected onto the Line Os, noted as S'' and the Length of OS'' is

$$|OS''| = |OS| \cos\alpha = |OS| \sin\theta, |OS| = SOD, \alpha = \frac{\pi}{2} - \theta. \tag{B.12}$$

Then as previous introduced, we can obtain the object points, a, b, c and d, coordinates as

$$\begin{aligned} &a(-t, -|OS|\sin\theta - L_a, z'), b(t, -|OS|\sin\theta - L_a, z'), \\ &c(-t, -|OS|\sin\theta + L_b, z'), d(t, -|OS|\sin\theta + L_b, z'), \end{aligned} \tag{B.13}$$

where (t, s) stands for local coordinate system variables, L_a and L_b are s-component distance away from the object origin S'', and $z' = r^2/|OS|\cos\theta$. Later we can show that boundary sloping rate is independent from the L_a and L_b elements, which meets the principle that any two given points can determine one line.

Next, transform local coordinates into world coordinate system according to

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} t \\ s \\ z' \end{bmatrix} + \begin{bmatrix} 0 \\ |OS|\sin\theta\cos\theta \\ |OS|\sin^2\theta + z_1 \end{bmatrix}. \tag{B.14}$$

We got a, b, c, d coordinates as

$$\begin{aligned} &a\left(-\sqrt{r^2 - \frac{r^4}{|OS|^2\cos^2\theta}}, -L_a\cos\theta - \frac{r^2}{|OS|\cos\theta}\sin\theta, -L_a\sin\theta + \frac{r^2}{|OS|} + z_1\right), \\ &b\left(\sqrt{r^2 - \frac{r^4}{|OS|^2\cos^2\theta}}, -L_a\cos\theta - \frac{r^2}{|OS|\cos\theta}\sin\theta, -L_a\sin\theta + \frac{r^2}{|OS|} + z_1\right), \\ &c\left(-\sqrt{r^2 - \frac{r^4}{|OS|^2\cos^2\theta}}, L_b\cos\theta - \frac{r^2}{|OS|\cos\theta}\sin\theta, L_b\sin\theta + \frac{r^2}{|OS|} + z_1\right), \\ &d\left(\sqrt{r^2 - \frac{r^4}{|OS|^2\cos^2\theta}}, L_b\cos\theta - \frac{r^2}{|OS|\cos\theta}\sin\theta, L_b\sin\theta + \frac{r^2}{|OS|} + z_1\right). \end{aligned} \tag{B.15}$$

Then following the Eq. (B.7), we got A,B,C,D coordinates

$$\begin{aligned} A &= \frac{z_2}{|OS| + L_a\sin\theta - \frac{r^2}{|OS|}}\left(-\sqrt{r^2 - \frac{r^4}{|OS|^2\cos^2\theta}}, -L_a\cos\theta - \frac{r^2}{|OS|\cos\theta}\sin\theta, 0\right), \\ B &= \frac{z_2}{|OS| + L_a\sin\theta - \frac{r^2}{|OS|}}\left(\sqrt{r^2 - \frac{r^4}{|OS|^2\cos^2\theta}}, -L_a\cos\theta - \frac{r^2}{|OS|\cos\theta}\sin\theta, 0\right), \\ C &= \frac{z_2}{|OS| - L_b\sin\theta - \frac{r^2}{|OS|}}\left(-\sqrt{r^2 - \frac{r^4}{|OS|^2\cos^2\theta}}, L_b\cos\theta - \frac{r^2}{|OS|\cos\theta}\sin\theta, 0\right), \\ D &= \frac{z_2}{|OS| - L_b\sin\theta - \frac{r^2}{|OS|}}\left(\sqrt{r^2 - \frac{r^4}{|OS|^2\cos^2\theta}}, L_b\cos\theta - \frac{r^2}{|OS|\cos\theta}\sin\theta, 0\right). \end{aligned} \tag{B.16}$$

Thus we can infer that

$$\overrightarrow{AC} = \left((k_a - k_c)\sqrt{r^2 - \frac{r^4}{|OS|^2\cos^2\theta}}, k_c\left(L_b\cos\theta - \frac{r^2\tan\theta}{|OS|}\right) + k_a\left(L_a\cos\theta + \frac{r^2\tan\theta}{|OS|}\right), 0\right)$$
$$\overrightarrow{BD} = \left((k_c - k_a)\sqrt{r^2 - \frac{r^4}{|OS|^2\cos^2\theta}}, k_c\left(L_b\cos\theta - \frac{r^2\tan\theta}{|OS|}\right) + k_a\left(L_a\cos\theta + \frac{r^2\tan\theta}{|OS|}\right), 0\right), \tag{B.17}$$

where $\overrightarrow{AC}$ and $\overrightarrow{BD}$ are detected boundaries on projection image. It is clear that two boundaries are symmetric about the Y-axis, which meets our expectation in no bias case. Thus we have slope

$$\begin{aligned} K = \frac{-\overrightarrow{AC}_x}{\overrightarrow{AC}_y} &= \frac{-(k_a - k_c)\sqrt{r^2 - \frac{r^4}{|OS|^2\cos^2\theta}}}{k_c\left(L_b\cos\theta - \frac{r^2\tan\theta}{|OS|}\right) + k_a\left(L_a\cos\theta + \frac{r^2\tan\theta}{|OS|}\right)} \\ &= \frac{(L_a + L_b)\sin\theta\sqrt{r^2 - \frac{r^4}{|OS|^2\cos^2\theta}}}{(L_a + L_b)\cos\theta(|OS| - \frac{r^2}{|OS|}) - \frac{r^2}{|OS|}\tan\theta(L_a + L_b)\sin\theta} \\ &= \frac{r\sin\theta\sqrt{1 - \frac{r^2}{|OS|^2\cos^2\theta}}}{|OS|\cos\theta - \frac{r^2}{|OS|}}. \end{aligned} \tag{B.18}$$

Let $c = \frac{r}{|OS|}$, we can simplify the K as

$$K = \frac{c\sin\theta}{\sqrt{\cos^2\theta - c^2}}, -\frac{\pi}{2} < \theta < \frac{\pi}{2}. \tag{B.19}$$

Hence, we can find the analytical solution to the tilt error when we figure out the detected boundary slope as follows

$$\theta = \pm\arccos\sqrt{\frac{c^2(1 + K^2)}{c^2 + K^2}}, \tag{B.20}$$

where θ symbol is the same as sloping rate K.

A phantom simulation was performed by MATLAB, in this case, we defined that $SDD = 2000$, $SOD = 1500 = |OS|$, Cylinder Radius $r = 150$, $2l = 400$, Tilt error $\theta = 30$. Phantom simulation is presented in Fig. B.4. According to Eq. (B.19), we can calculated theoretical slope $K = 0.0581$, while actual detected value based on edge detection method is $K = 0.0588$. Following Eq. (B.20), our estimated Tilt angle error is $\theta = 30.2868$, which is within error range.

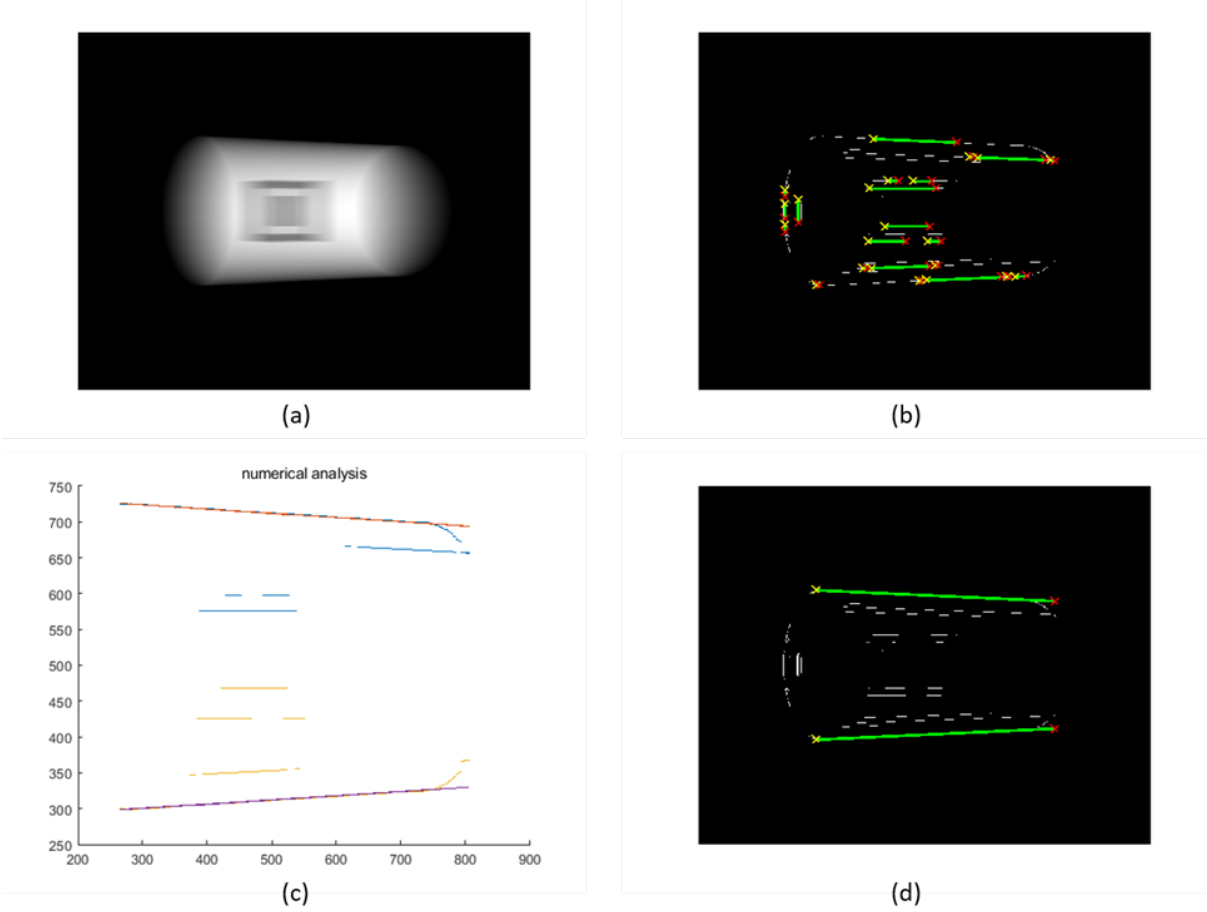


Figure B.4: Phantom simulation and edge detection (a) Phantom projection image after 30 degree tilt, up and bottom edges are no longer parallel to each other. (b) Edge detected after Canny edge detection and Hough Transform selection, certain unwanted edges are included in automatic selection. (c) A presentation of two detected boundaries, red lines, among Canny edges, which are estimated by RANSAC algorithm with adding extra boundary points into dataset pool for iterative selection. (d) Two detected boundaries presented in the binarized projection image, appearance shows fitting well.

Case 2: Tilt(θ) and slant(η) error exist, no bias($m, n = 0$):
In practice, besides the tilt describing the orientation error, slant is another parameter describing the misalignment of rotation axis orientation which refers to the inaccurate or improper placement of the add-on device, as Fig. B.6 shows. In projection domain, slant error, that rotates around the Z-axis, is similar to an in-plane rotation and noted as eta.

Refer Eq. (B.1), we have tangent point, a, b, c and d, coordinates as

$$a(-t, -L_a, z'), b(t, -L_a, z'), c(-t, L_b, z'), d(t, L_b, z'). \tag{B.21}$$

Then transform the local coordinates into world coordinates as

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \cos\eta & -\sin\eta & 0 \\ \sin\eta & \cos\eta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} t \\ s \\ z' \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ z_1 \end{bmatrix}. \tag{B.22}$$

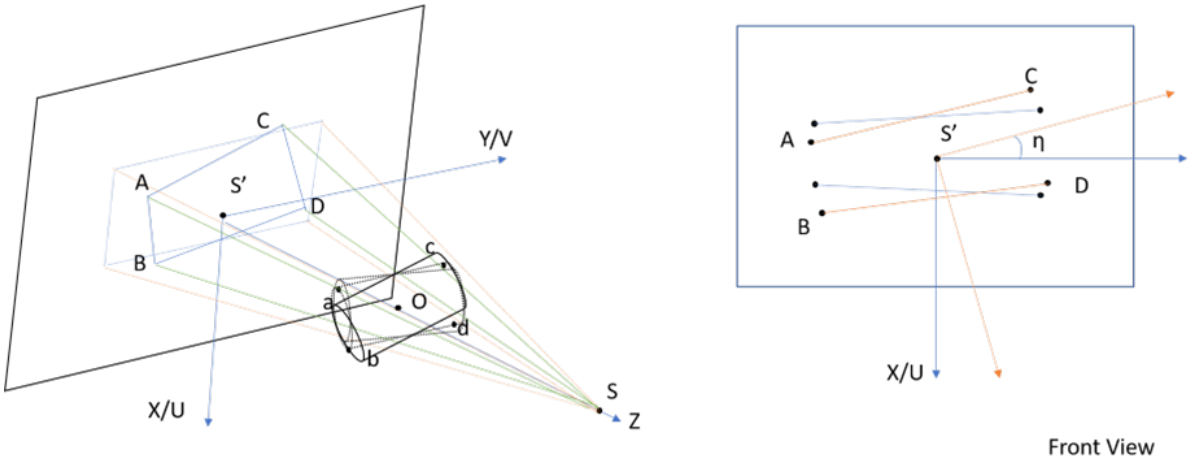


Figure B.5: Sketch of object cone beam scanning geometry with Tilt and In-plane Rotation

Note counterclockwise rotation around the axis equals the positive sign of the degree η and θ. Thus, a, b, c, d become

$$\begin{aligned}
&a(-t\cos\eta + \cos\theta\sin\eta L_a - \sin\eta\sin\theta z', -t\sin\eta - \cos\theta\cos\eta L_a + \sin\theta\cos\eta z', \\
&\quad \sin\theta L_a + z'\cos\theta + z_1), \\
&b(t\cos\eta + \cos\theta\sin\eta L_a - \sin\eta\sin\theta z', t\sin\eta - \cos\theta\cos\eta L_a + \sin\theta\cos\eta z', \\
&\quad \sin\theta L_a + z'\cos\theta + z_1), \\
&c(-t\cos\eta - \cos\theta\sin\eta L_b - \sin\eta\sin\theta z', -t\sin\eta + \cos\theta\cos\eta L_b + \sin\theta\cos\eta z', \\
&\quad -\sin\theta L_b + z'\cos\theta + z_1), \\
&d(t\cos\eta - \cos\theta\sin\eta L_b - \sin\eta\sin\theta z', t\sin\eta + \cos\theta\cos\eta L_b + \sin\theta\cos\eta z', \\
&\quad -\sin\theta L_b + z'\cos\theta + z_1).
\end{aligned} \tag{B.23}$$

Then following Eq. (B.17) and Eq. (B.18), let $c = |OS| = z_2 - z_1$, we have

$$\begin{aligned}
K_{AC} &= \frac{-\overrightarrow{AC}_x}{\overrightarrow{AC}_y} = \frac{-\sin\theta(t\cos\eta + \sin\eta\sin\theta z') + (c - z'\cos\theta)\cos\theta\sin\eta}{\sin\theta(t\sin\eta - \cos\eta\sin\theta z') + (c - z'\cos\theta)\cos\theta\cos\eta} \\
K_{BD} &= \frac{-\overrightarrow{BD}_x}{\overrightarrow{BD}_y} = \frac{\sin\theta(t\cos\eta - \sin\eta\sin\theta z') + (c - z'\cos\theta)\cos\theta\sin\eta}{-\sin\theta(t\sin\eta + \cos\eta\sin\theta z') + (c - z'\cos\theta)\cos\theta\cos\eta}.
\end{aligned} \tag{B.24}$$

For simplicity, let $m = \frac{r}{c}$, $X = \frac{m \sin\theta}{\sqrt{\cos^2\theta - m^2}}$, then

$$\begin{aligned} K_{AC} &= \frac{\sin\eta - X\cos\eta}{\cos\eta + X\sin\eta} \\ K_{BD} &= \frac{\sin\eta + X\cos\eta}{\cos\eta - X\sin\eta}. \end{aligned} \tag{B.25}$$

Note (θ, η) and $(-\theta, \eta + \pi)$ give rise same slopes given same K_AC and K_BD. Thus, in this research, degree η and θ are pre-assumed under constraint condition $-pi/2 < \theta, \eta < \pi/2$ based on the fact that their values are pretty small otherwise they will be recognized easily through human eyes. Then we have unique solution to these equations. For simplicity, denote $K_1 = K_{AC}$ and $K_2 = K_{BD}$, then

$$X = \frac{(K_1K_2 + 1) \pm \sqrt{k_1^2 + k_2^2 + k_1^2k_2^2 + 1}}{k_1 - k_2} = \frac{m\sin\theta}{\sqrt{\cos^2\theta - m^2}}. \tag{B.26}$$

Therefore, we got analytical solution

$$\begin{cases} \theta = \arccos\sqrt{\frac{m^2(X^2+1)}{m^2+X^2}}, & \eta = \arctan\sqrt{\frac{K_1+K_2}{2-(K_1-K_2)X}}, \quad K_1 - K_2 < 0 \\ \theta = -\arccos\sqrt{\frac{m^2(X^2+1)}{m^2+X^2}}, & \eta = \arctan\sqrt{\frac{K_1+K_2}{2-(K_1-K_2)X}}, \quad K_1 - K_2 > 0 \end{cases} \tag{B.27}$$

Similarly, a phantom simulation was taken by MATLAB, in this case, we defined that $SDD = 2000$, $SOD = 1500 = |OS|$, Cylinder Radius $r = 150$, $2l = 400$, Tilt error $\theta = -30$, Eta error $\eta = -30$. Phantom simulation is presented in Fig. B.6. According to Eq. (B.25), we can calculated theoretical slope $K_1 = -0.5024, K_2 = -0.6575$, while actual detected value based on edge detection method is $K_1 = -0.5052, K_2 = -0.6561$. Following Eq. (B.27), our estimated Tilt and Slant angle error are $\theta = -29.2984$,$\eta = -30.0359$, respectively.

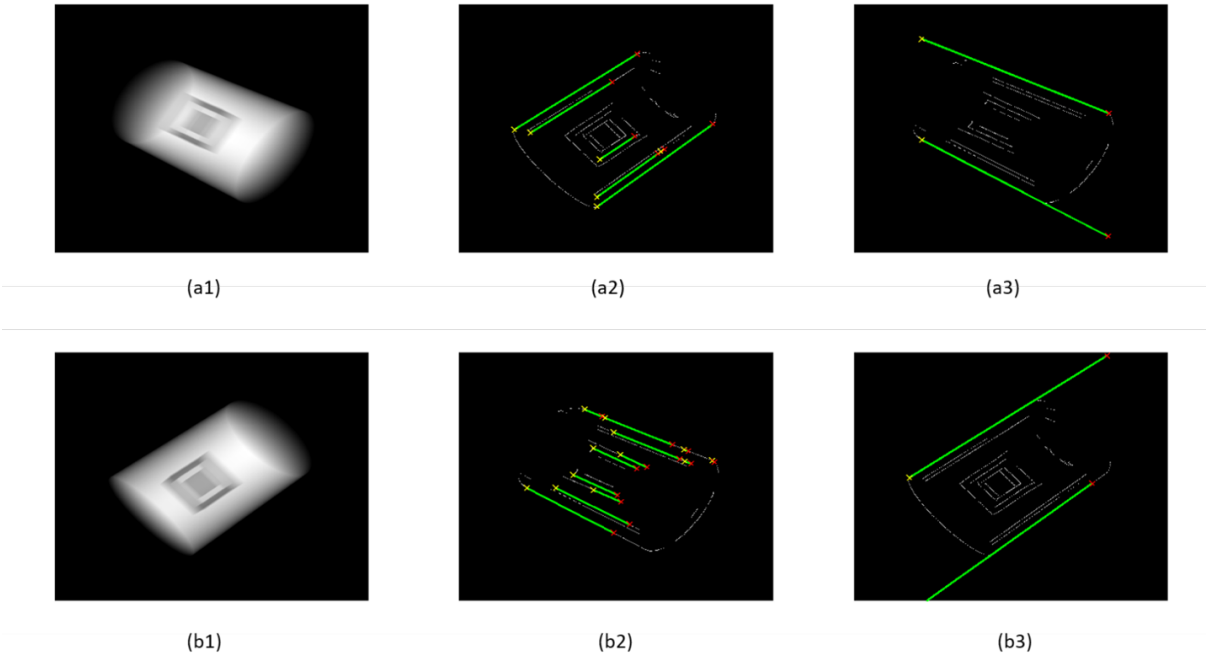


Figure B.6: Two examples of Tilt and Slant determination. (a) case, tilt = -30 degree, slant = -30 degree, (b) case, tilt = 20 degree, slant = 40 degree, estimated $K_1 = 0.7866$ and $K_2 = 0.9035$, estimated tilt = 18.7434, slant = 40.1420, detected $K_1 = 0.7792$ and $K_2 = 0.8974$, detected tilt = 19.0684, slant = 39.9151.