

Received November 16, 2020, accepted November 24, 2020, date of publication December 21, 2020, date of current version December 31, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3046187

Clinical Micro-CT Empowered by Interior Tomography, Robotic Scanning, and Deep Learning

MENGZHOU LI^{1,*}, (Graduate Student Member, IEEE), ZHENG FANG^{1,2,*}, WENXIANG CONG¹, CHUANG NIU¹, WEIWEN WU¹, JOSEF UHER³, JAMES BENNETT⁴, JAY T. RUBINSTEIN⁵, (Senior Member, IEEE), AND GE WANG¹, (Fellow, IEEE)

¹Department of Biomedical Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180, USA

²Department of Instrumental and Electrical Engineering, Xiamen University, Xiamen 361102, China

³Radalytica a.s., 17000 Prague, Czech Republic

⁴Hawkeye Spectral Imaging, Tucson, AZ 85701, USA

⁵Virginia Merrill Bloedel Hearing Research Center, Department of Otolaryngology-HNS, University of Washington, Seattle, WA 98195, USA

Corresponding author: Ge Wang (wangg6@rpi.edu)

*Mengzhou Li and Zheng Fang contributed equally to this work.

This work was supported in part by the National Institutes of Health [National Cancer Institute(NCI)] and [National Institute of Biomedical Imaging and Bioengineering (NIBIB)] under Award R01CA237267 and Award R01EB026646, and in part by the State Scholarship Fund from the China Scholarship Council (CSC) under Grant 201906315030.

ABSTRACT While micro-CT systems are instrumental in preclinical research, clinical micro-CT imaging has long been desired with cochlear implantation as a primary application. The structural details of the cochlear implant and the temporal bone require a significantly higher image resolution than that (about 0.2 mm) provided by current medical CT scanners. In this paper, we propose a clinical micro-CT (CMCT) system design integrating conventional spiral cone-beam CT, contemporary interior tomography, deep learning techniques, and the technologies of a micro-focus X-ray source, a photon-counting detector (PCD), and robotic arms for ultrahigh-resolution localized tomography of a freely-selected volume of interest (VOI) at a minimized radiation dose level. The whole system consists of a standard CT scanner for a clinical CT exam and VOI specification, and a robotic micro-CT scanner for a local scan of high spatial and spectral resolution at minimized radiation dose. The prior information from the global scan is also fully utilized for background compensation of the local scan data for accurate and stable VOI reconstruction. Our results and analysis show that the proposed hybrid reconstruction algorithm delivers accurate high-resolution local reconstruction, and is insensitive to the misalignment of the isocenter position, initial view angle and scale mismatch in the data/image registration. These findings demonstrate the feasibility of our system design. We envision that deep learning techniques can be leveraged for optimized imaging performance. With high-resolution imaging, high dose efficiency and low system cost synergistically, our proposed CMCT system has great promise in temporal bone imaging as well as various other clinical applications.

INDEX TERMS Clinical micro-CT, deep learning, high-resolution imaging, interior tomography, photon-counting detector, robotic arms, temporal bone imaging, X-ray computed tomography.

I. INTRODUCTION

In the clinical practice of otology and neurotology, medical imaging is critical for evaluation and treatment of many diseases [1]. Magnetic resonance imaging (MRI) is an effective imaging tool with excellent soft tissue contrast and invaluable in detecting neoplasms with gadolinium contrast agent, but

The associate editor coordinating the review of this manuscript and approving it for publication was Yi Zhang.

it is not good at evaluating bony structures [2]. Currently, temporal bone CT is the primary method of choice for otological imaging [3]. Chronic otitis media, otosclerosis, temporal bone fracture, congenital aural atresia, cochlear implantation, dehiscence superior semicircular canal, congenital labyrinthine dysplasia, labyrinthine fistula are all disorders or therapies where temporal bone CT is either absolutely necessary or a commonly desired adjunct to surgical management, for diagnosis or planning. In cochlear implantation, although

psychophysical and physiological measures are of primary importance, the ability to localize electrodes and depict their 3D anatomical environment in vivo is of great relevance to understanding variations in threshold, uncomfortable level, and channel interaction.

CT is widely used to image a variety of middle and inner ear pathologies but it is limited by suboptimal image resolution [4]. While most CT scanners are limited to, at best, a spatial resolution of about 0.3 mm, the latest Cannon CT scanner resolves details down to about 0.2 mm [5]. This level of image resolution is still insufficient. For example, it is common that a precise diagnosis is unavailable until direct assessment during otologic surgery allows visual inspection and palpation of the ossicular chain. Also, much-improved image resolution of pre- and post-operative inner ear imaging is required for detailed analysis of cochlear morphometry and its relationship to an implanted electrode array [6]. In these and many other research and clinical applications, there are critical and immediate needs for ultra-fine spatial resolution without compromising other image quality indices at a minimized radiation dose level. A breakthrough in CT image resolution may greatly benefit diagnosis and treatment in general, and otology and neurotology in particular, such as for rational design and implantation of prosthetic devices.

In contrast to the classic CT system, the robotic-arm-based X-ray imaging system allows a great flexibility of scanning. It supports different scanning trajectories optimized for diverse tasks focusing on various organs and locations. Siemens developed a robotic X-ray system named Multitom Rax, which can scan a patient under a natural weight bearing condition. It enables a variety of clinical examinations. Three-dimensional images acquired by Multitom Rax improve diagnostic and planning performance compared to what we obtained when a patient lies on a bed. Researchers with Johns Hopkins University developed a mathematical framework for the design of scanning trajectories optimal to a particular task with cone-beam CT [7]. FirstImaging also develops robotic image systems with excellent image quality.

The X-ray photon-counting detector (PCD) is an enabling technology on the horizon for high-resolution (HR) and low-noise imaging, which promises to add a spectral dimension to raw data and boost CT performance [8]. Different from the energy-integrating detector (EID), PCD works in a pulse-counting mode and directly converts individual X-ray photons into the corresponding charge signals which are then sorted into different energy bins based on the pulse heights. Thus, the intensity and wavelength information of incoming photons are simultaneously obtained. PCDs principally have no electronic noise and provide quite small effective pixel size; e.g., around 0.11 mm × 0.11 mm. In contrast, EIDs suffer from the dark current and readout noise with the element size of typically about 1 mm × 1 mm. In addition, EIDs put more weights on high-energy photons than on low-energy photons but high-energy photons usually get attenuated less than low-energy photons, leading to a reduced image contrast. On the other hand, with PCDs optimal weights can be

implemented on polychromatic photons for improved contrast and dose efficiency. More importantly, the energy discrimination ability of PCDs helps reduce beam hardening and metal artifacts, and enables K-edge imaging and material decomposition.

In parallel to the development of X-ray detectors, innovations of micro-focus X-ray tubes are also important. The NanoX tube is an example [9], and consists of electron emitting and receiving constructs. The receiving part is mainly an anode with a photoconductor. The emission part includes a backplate, a substrate, a cathode, a gate electrode, and an array of field emission electron sources. A microstructured array anode target (MAAT) X-ray source is another example, which was recently designed [10] to offer a significantly higher flux than an ordinary X-ray source in phase contrast imaging applications. The key parameters were optimized in the range of 40 keV to 130 keV. These types of technologies could be combined into a micro-focus X-ray tube for temporal bone CT imaging; for example, with a focal spot size of about 0.1 mm or less to match the PCD element size.

We envision a robotic-arm-based design of a clinical micro-CT scanner to integrate a micro-focus tube, a PCD, interior tomography, and deep learning into a novel device that can be attached to a clinical CT scanner to form a clinical micro-CT system (CMCT) or used separately. The proposed CMCT workflow starts with a conventional volumetric scan of a patient, and then proceeds to a subsequent image analysis session by an expert or a smart software analyzer. After that, the patient is smoothly translated into the robotic scanning space for an interior photon-counting micro-CT scan, in registration to the previously performed global scan. Based on interior tomography theory, the interior scan can target only a region of interest (ROI) with a small detector panel and a customized scanning trajectory so that the image quality can be optimized. To facilitate accurate registration between the global scan data/images and local counterparts, an optical 3D surface scanner is used at the start of the local scan and it continues to monitor the head movement of the patient [11] during the local scan for motion compensation. This design integrates all relevant cutting-edge hardware and software elements, and systematically upgrades our earliest CMCT design published in 2005 [12] as well as other follow-up designs for similar purposes [13]–[18].

Compared to the above-mentioned prior designs, our approach aims at a much higher resolution of 50 μm with a high-power micro-focus source and a state-of-the-art PCD with fine detector elements for electronic-noise-free spectral imaging. Furthermore, the robotic arms add a great mechanical flexibility compared to the C-arm gantry and enable a moving VOI mask scanning for higher dose efficiency [14].

As far as the image reconstruction is concerned, the current mainstream of interior tomography methods focuses on improving reconstruction quality from truncated projections with no or little extra information [19]–[24]. Nevertheless, in our setting a global scan is available, which is clinically natural and can be utilized for robust

reconstruction at least for the initial imaging task. Several methods that complete the HR projection data with a low-resolution global scan have been developed for truncation-artifacts-free reconstruction [12]–[15], [25], [26]. [27], in particular, is computationally-efficient without forward projection operations that directly fuses global and local reconstructions, but it is restricted to a certain type of scanning geometry and inappropriate for our application. Instead of completing the HR truncated projections to cover a global FOV, we propose to directly perform a background compensation on the HR local/interior scan with the global scan. The main benefit is that the subsequent interior reconstruction only involves a small portion of the sinogram, requiring much less memory space and computational time. This is critical since the projection resolution will be enhanced up to 10 folds (from 1 mm to 0.11 mm), hence the amount of interpolated 2D projection data will be increased 100 times compared to that of the global scan, and handling such big data for reconstruction can be very time-consuming or even prohibitive. Furthermore, for quality enhancement at reduced doses, deep learning techniques can be involved in multiple reconstruction stages; e.g., projection deblurring, image denoising and super-resolution, beam hardening correction and material decomposition.

II. SYSTEM DESIGN

A. SYSTEM DESCRIPTION

The proposed CMCT system consists of a standard medical CT scanner or a novel clinical CT system (such as a NanoX system) and a robotic micro-CT scanner, as shown in Fig. 1. The CT scanner performs a global scan. A standard CT algorithm can be used to reconstruct head slices, which show the inner ear region of interest. Then, the patient table transports the patient into the robotic micro-CT system.

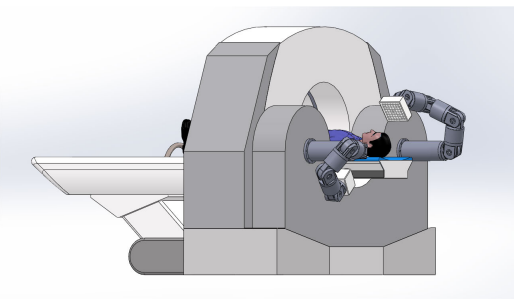


FIGURE 1. Schematic diagram of the proposed clinical micro-CT (CMCT) system.

This robotic micro-CT system uses two robotic arms holding the X-ray source and the X-ray detector respectively, as shown in Fig. 2. The micro-focus tube is a key component for high spatial resolution imaging. The flat-panel PCD is suitable for material decomposition and tissue characterization. The robotic arms can perform a scan along an arbitrary trajectory such as a circular or spiral trajectory. A high-performance computer is in overall control. It sends

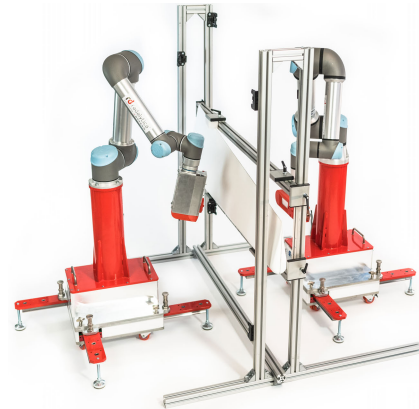


FIGURE 2. Robotic cone-beam micro-CT system.

instructions to the robotic arms, X-ray source and detector. It also acquires raw data from the detector. Each robotic arm needs one control box, which receives commands from the computer and drives the servo system.

B. DESIGN PRINCIPLES

1) COORDINATED ROBOTS

The lightweight, highly flexible, and collaborative UR5e and UR16e industrial robot arms with payloads of up to 5 kg and 16 kg, respectively, are suitable for our CMCT system [28]. The working radius of UR5e is 850 mm with power consumption of 250 W, while the working radius of UR16e is 900 mm with power consumption of 585 W. The communication between the control computer and the robotic arm control box is through an Ethernet cable using TCP/IP 1000BASE-T protocol. The 6-axis robotic arms have a high position repeatability of 30 μm . Some models can be used to achieve even higher accuracy. The existing Radalytica robotic imaging platform can be modified for CMCT, pushing the limits of robotic imaging considerably further. The robotic-arm cone-beam micro-CT sub-system can be made to achieve a spatial resolution of 50 μm or higher to meet temporal bone imaging requirements [29]. The power supply can be from 100 to 240 VAC at 47 to 440 Hz.

2) X-RAY SOURCE

The Hamamatsu L12161-07 tube is selected as the X-ray source for CMCT. It is an RS-232 controlled 40-150 kV X-ray source with a focal spot size adjustable among 7, 20, and 50 μm . Its maximum tube power output is 75 W. The anode target material is Tungsten. The tube window is made of Beryllium of 0.2 mm thickness. The effective cone beam angle is 43°. Its weight is 13.5 kg, manageable with the larger robot arm.

3) PHOTON-COUNTING DETECTOR

The photon-counting detector ADVACAM WIDEPIC-5 \times 5 of 1280 \times 1280 pixels fits the need for human temporal bone micro-CT. Its continuous sensitive surface is supported by an array of 5 \times 5 detector tiles. Each tile consists of a single

Timepix hybrid detector (256×256 pixels) with an edgeless CdTe sensor. Each pixel counts the number of X-ray photons, allowing a large dynamic range. The Timepix technology also allows for the use of multiple energy thresholds. The intrinsic spatial resolution is defined by the detector pitch of $55 \mu\text{m}$. That is, the imaging sensor covers a $7 \text{ cm} \times 7 \text{ cm}$ area. The detector weighs 3.3 kg and can be easily carried by the smaller robotic arm. If we use the detector in a 2×2 binning mode, image resolution is $50 \mu\text{m}$, and the diameter of the volume of interest (VOI) is 3.5 cm , assuming a magnification factor of 2. This should be sufficient to cover the human inner and middle ear. In addition to the Timepix detector, we can also customize the Medipix3 detector tiles, whose pixels have two integrated 12-bit digital counters and two energy discrimination thresholds. If we use the detector in a 2×2 binning mode, there are 8 spectral bins for data collection in a single scan.

4) RADIATION DOSE

Radiation dose is mainly determined by the tube voltage, current and exposure period. With the use of the PCD, there is no electronic noise when recording projection data but Poisson noise cannot be avoided. In the interior scanning mode, the X-ray source only radiates about 1/10 of the diameter of the field of view but improvement in image resolution by four times (roughly, from $200 \mu\text{m}$ to $50 \mu\text{m}$) would increase radiation dose significantly (two orders of magnitude) [12]. Thanks to the latest advancement in deep-learning-based low-dose CT imaging techniques [30], we can reduce radiation dose by an order of magnitude. With all of these factors coupled together, we should be able to maintain approximately the current head CT dose for an interior micro-CT scan to achieve about $50 \mu\text{m}$ resolution.

5) SPATIAL RESOLUTION

We define the center of the focal spot as S, the center of the detector as D, and the rotation center as O which is the center of a VOI. These three points should always be kept in a straight line. In Fig. 3, a is the source to VOI distance, b is the VOI to detector distance, and c is the source to detector distance. The geometric magnification factors M and M' are c/a and c/b for the focal spot and the detector aperture respectively [31]. The spatial resolution r of the imaging system can be approximated as a convolution of the detector

size d and the focal spot size x respectively scaled by M and M' :

$$r = \sqrt{\left(\frac{x}{M}\right)^2 + \left(\frac{d}{M'}\right)^2} \quad (1)$$

In our initial CMCT system design, x is no more than $50 \mu\text{m}$, and d is equal to $110 \mu\text{m}$ (due to 2×2 binning). The source to VOI distance can be from 150 to 250 mm , and the VOI to detector distance is in the same range. Consequently, the magnification factor can be adjusted from 1.60 to 2.67 with the imaging field of view from 26 to 44 mm in diameter. According to the above equations, the system spatial resolution can be made close to $50 \mu\text{m}$.

6) GEOMETRIC ALIGNMENT

In the CMCT process, global and local projection data are obtained in different scanning geometries. An optical 3D surface scanner is preferably used for image registration, as shown in Fig. 4. The clinical CT scan obtains a 3D image of a patient who may wear landmarks such as a firmly-attached helmet. A boundary detection algorithm extracts the facial/helmet surface and key points for mesh generation as a basis for image registration [32]. The fact is utilized that the inner ear and the face surface are in a rigid relation. As an example of 3D surface scanners, the laser scanner Micro-Epsilon LIT2910-100 is a high-quality profile sensor. The height range may be set from 125 to 390 mm , and the width range can be fixed to 143.5 mm , with spatial resolution of $12 \mu\text{m}$ at a profiling frequency of 300 Hz . The compact size of $96 \text{ mm} \times 85 \text{ mm} \times 33 \text{ mm}$ and low weight of 380 grams are ideal for static, dynamic and robotic applications. The wavelength of the semiconductor laser is switchable between 658 nm (red) and 405 nm (blue). The measurement data can be output via an Ethernet UDP, Modbus TCP, or serial communication protocol. The power supply is $24 \text{ VDC } 500 \text{ mA}$. The optical scanner and the two robotic arms are mounted on the same pedestal so that they share the same coordinate system. The optically scanned patient head surface needs to be registered with the clinical CT originated head surface. Then, the coordinate conversion is carried out to delineate a VOI for a robotic micro-CT scan. The temporal bone data and images by medical CT and micro-CT also need to be registered and fused to achieve the best imaging performance.

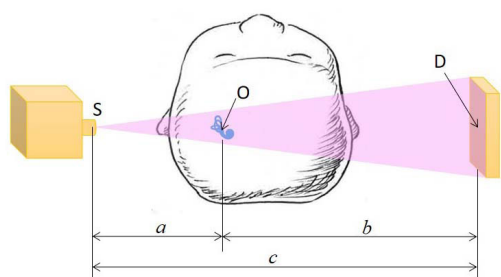


FIGURE 3. Geometry of a robotic imaging chain.

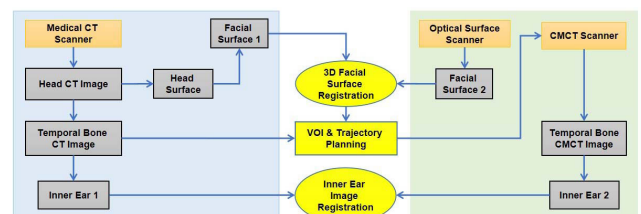


FIGURE 4. CMCT image registration integrating standard temporal bone CT and our robotic micro-CT imaging.

III. IMAGING ALGORITHM

It is well known that the CT interior problem is not uniquely solvable in an unconstrained space [33]. By introducing additional prior knowledge on the image to be reconstructed; e.g., an interior sub-region with known attenuation values or a piece-wise constant model of underlying images, the image reconstruction from local projection profiles that are truncated on both sides becomes uniquely solvable. However, those assumptions often do not exactly hold in practical cases, which potentially results in shifting and cupping in reconstructed attenuation values. To address this problem, we propose to use a low-resolution (LR) global CT scan to estimate the background attenuation in the sinogram of the HR local scan that involves the surrounding volume of the VOI, and obtain an accurate HR local reconstruction of the VOI.

The general idea of background compensation is illustrated in Fig. 5 in a 2D case without loss of generality. The transformation between the underlying/reconstructed image and its projection data is linear and invertible, and we can partition an underlying image into the two parts, the region of interest and the background. Given the sinogram of the background $P_{background}$ and the global sinogram P_{global} , we can easily obtain the pure sinogram of the region of the interest (ROI) P_{ROI} as follows:

$$P_{ROI} = P_{global} - P_{background}. \quad (2)$$

The above relationship becomes nontrivial with laterally truncated projection data. Let $trunc(\cdot)$ denote the truncation operation, and Eq. 2 becomes

$$trunc(P_{ROI}) = trunc(P_{global}) - trunc(P_{background}), \quad (3)$$

where $trunc(P_{global})$ stands for a local scan P_{local} . By intentionally letting the local scan cover the ROI, the truncated parts of P_{ROI} are all zeros, and we have

$$P_{ROI} = P_{local} - trunc(P_{background}). \quad (4)$$

This equation suggests that the ROI within a local scan can be accurately reconstructed from the laterally truncated scan after background subtraction. Clearly, this compensation will improve the stability of interior tomography.

For CMCT, a VOI can be accurately reconstructed at high resolution, given the HR local projection dataset P_{local}^{HR} and an appropriate background estimation. Suppose that we have a prior LR CT scan of the object P_{global}^{LR} , the LR background estimation $P_{background}^{LR}$ can be easily obtained as shown in Fig. 5(a), which can be used to approximate the HR background closely via interpolation. This is based on the assumption that residual high-frequency background estimation errors will mostly cancel out during the integration which makes the LR estimation sufficiently accurate for our purpose.

A. CMCT RECONSTRUCTION

In clinical applications, a standard (relatively low resolution) global CT scan is first performed. Those regions with possible pathology or of physiologic importance can be further

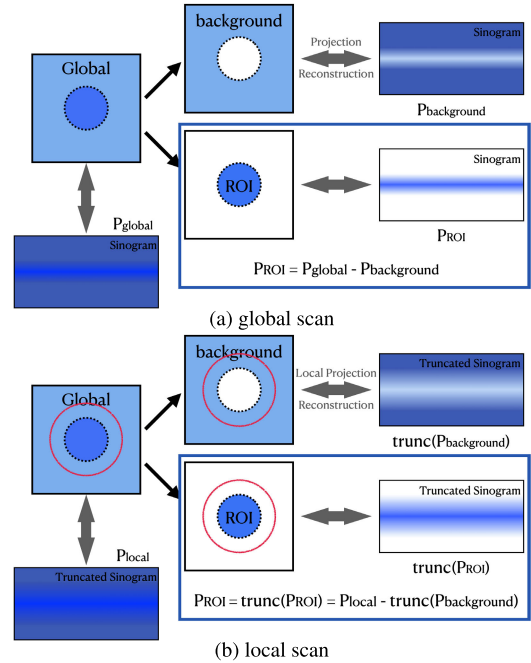


FIGURE 5. Illustration of the proposed ROI reconstruction through background compensation.

examined with a local micro-CT scan, which provides a HR local/interior reconstruction of a VOI. With interior tomography, a minimized additional dose will be involved. The prior information obtained through the global CT scan can be utilized to help the interior image reconstruction (at ultrahigh resolution).

The whole scanning procedure is summarized as follows:

- 1) Perform a global head CT scan P_{global} with a scanning geometry G_{global} ;
- 2) Determine a VOI and then plan a scanning geometry G_{local} for a local micro-CT scan;
- 3) Scan the patient optically with a surface scanner to generate the surface model S_{local} for data/image registration between the global and local scans at the start of the local scan;
- 4) Perform the micro-CT scan P_{local} following G_{local} (see the registration step below). During the local scan, the optical scanner continuously scans the patient head (preferably with optical markers) for tracing unconscious head movement.

Then, the VOI reconstruction mainly consists of the following three steps: data/image registration, background compensation, and image reconstruction. For brevity, in the following we assume that the head movement effects have already been compensated for in the local projections P_{local} with the feedback from the optical scanner using advanced correction techniques, like, locally linear embedding motion correction [34], [35].

- 1) **Registration:** Find the relative geometry between G_{global} and G_{local} in reference to the facial surface model S_{local} .

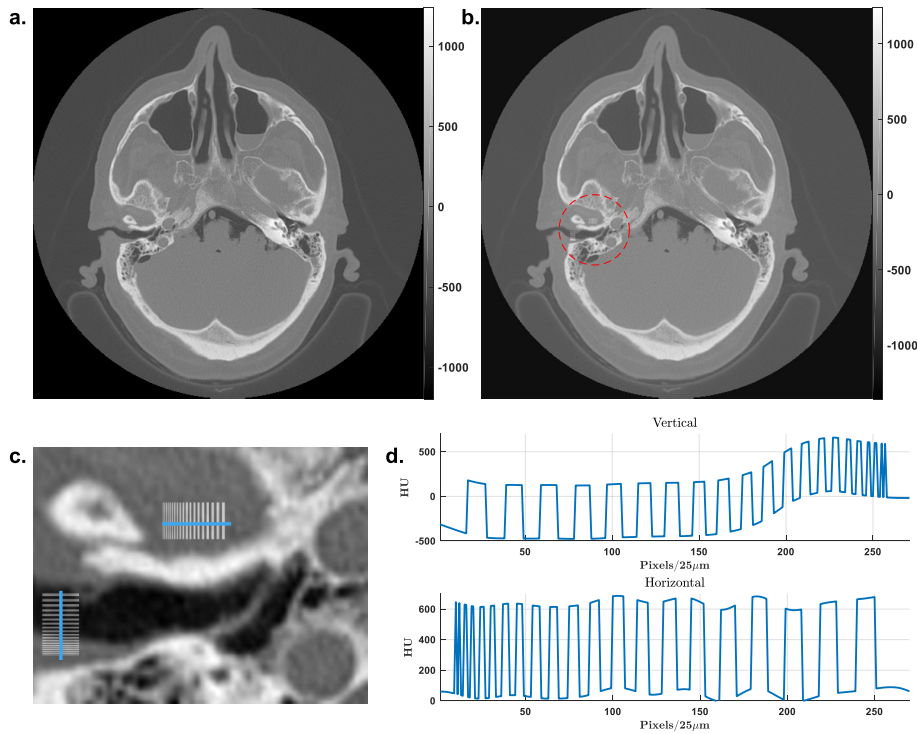


FIGURE 6. Head phantom with resolution patterns. (a) The original head image, (b) two resolution bar patterns placed along horizontal and vertical directions respectively with an amplitude of 600 HU in the inner ear region, (c) an enlarged region around the patterns, and (d) the profiles of the resolution patterns.

- a) Reconstruct the global volume V_{global} from P_{global} ;
 - b) Render the surface model S_{global} from the global reconstruction V_{global} ;
 - c) Register the two surface models to align the orientation and position of S_{global} with S_{local} . The registration result is used to guide the micro-CT scan;
 - d) From P_{local} , directly reconstruct a volume of interest V_{local} which contains fine structures but may be subject to distorted attenuation values;
 - e) Refine the registration parameters (obtained in Step 1c) in reference to the registration between V_{global} and V_{local} .
- 2) **Compensation:** Correct the attenuation offsets in P_{local} to form a pure sinogram of the VOI P_{VOI} .
- a) Set the attenuation values inside VOI to zero in the aligned global reconstruction to form the background volume $V_{background}^{align}$;
 - b) Digitally reproject the background volume $V_{background}^{align}$ following the geometry G_{local} to form the LR background estimation $P_{background}^{LR}$;
 - c) Interpolate $P_{background}^{LR}$ to the same resolution as the local HR projection P_{local} , and obtain $P_{background}^{HR}$;
 - d) Truncate $P_{background}^{HR}$ to the same size as P_{local} , and obtain $trunc(P_{background}^{HR})$;
 - e) Correct P_{local} with the estimated attenuation background as $P_{local} - trunc(P_{background}^{HR})$ to form the pure sinogram of the VOI P_{VOI} ;
- 3) **Reconstruction:** Reconstruct the VOI from P_{VOI} with geometry G_{local} using a cone-beam reconstruction algorithm, and preferably one developed in the deep learning framework.

IV. RESULTS

A. CMCT ACCURACY AND RESOLUTION

A simulation study was performed to demonstrate the feasibility of our proposed VOI reconstruction through background compensation. When the PCD is used, projections are collected in a number of energy bins. For inner ear imaging, we may initially focus on all the counts in a wide energy window to study the reconstruction performance; i.e., 40keV to 110keV (120 kVp source), for the following considerations: (1) to avoid the blurring from the X-ray fluorescence in the CdTe crystal; (2) to collect most photons that have penetrated through the head; and (3) to reduce the influence of pile-up effects. One head CT image containing the inner ear structures from the Visible Human Project [36] was used as a realistic image phantom, as shown in Fig. 6(a). The original phantom matrix is of 512×512 pixels with pixel

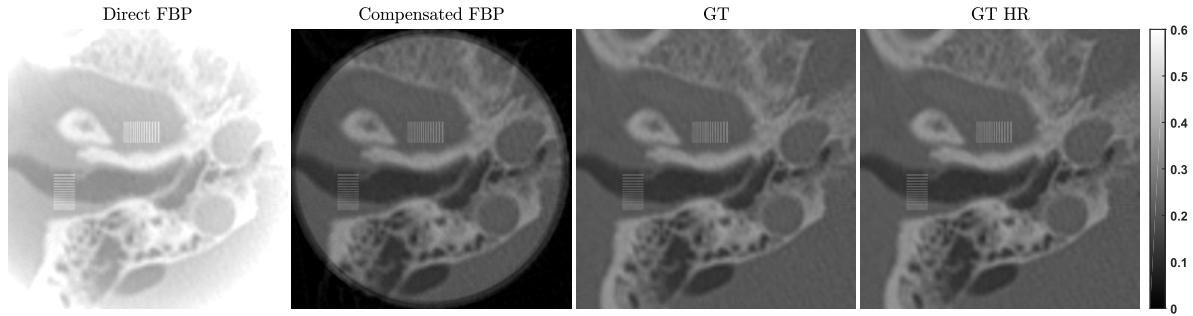


FIGURE 7. Reconstructed ROI in the same display window, including the direct reconstruction with FBP from local data, FBP reconstruction after background compensation, the ground truth reconstructed from the HR global scan with 0.11 mm pixel size, and the HR ground truth from the HR global scan with 0.04 mm pixel size. The attenuation value is in cm^{-1} .

size of 0.4981 mm . In our study, the phantom image was first interpolated to $10,240 \times 10,240$ pixels with a pixel size of 0.025 mm to generate the HR phantom. Note that the bicubic interpolation method was used to generate pixel values. Then, two resolution bar patterns in horizontal and vertical orientations were embedded in the inner ear region as shown in Fig. 6(b) to (d). The amplitude of the added patterns is 600 HU , and the radius of the ROI is 23 mm as marked with a red circle in Fig. 6(b). A global CT scan was performed with 1.024 mm detector pixel size, and the HR local CT scan covered a region of radius 35.05 mm with a detector pixel size 0.11 mm . Both scans had the same system magnification factor of 2 and tube voltage of 140 kVp , and were simulated with an industrial CT simulator CatSim [37]. Two additional HR global scans were also performed with fine detection pixels of 0.11 mm and 0.04 mm to produce the reference reconstructions as the ground truth (GT) of the attenuation value and image resolution, respectively. In Fig. 7, large deviations of attenuation values from the GT and strong cupping effects are observed for the direct local scan reconstruction with filtered back projection (FBP). In contrast, the reconstruction from the background compensated projections demonstrates accurate attenuation values. The negligible artifacts around the ROI boundary may come from minor mismatches between the background estimation. Interestingly, for the direct reconstruction, in spite of the attenuation value shifting, the fine details are still clearly discernible except for the distorted ROI boundary, which might be sufficient in those applications that only need structural features.

To evaluate the potential influence of the proposed method on the attenuation value and image resolution, the profiles along and around the vertical and horizontal midlines of the ROI and through the resolution bar patterns are shown in Fig. 8. The lines overlap well with the ground truth inside the ROI as demonstrated in Fig. 8(a) and 8(b), demonstrating the high fidelity of the reconstructed attenuation values. Similarly in Fig. 8(c) and (d). The overlapped profiles of the GT and FBP reconstruction show that the proposed local reconstruction method does not compromise image resolution as compared with the global HR scan. Particularly, the $50\text{ }\mu\text{m}$ gap is not resolved while the $75\text{ }\mu\text{m}$ gap is well resolved,

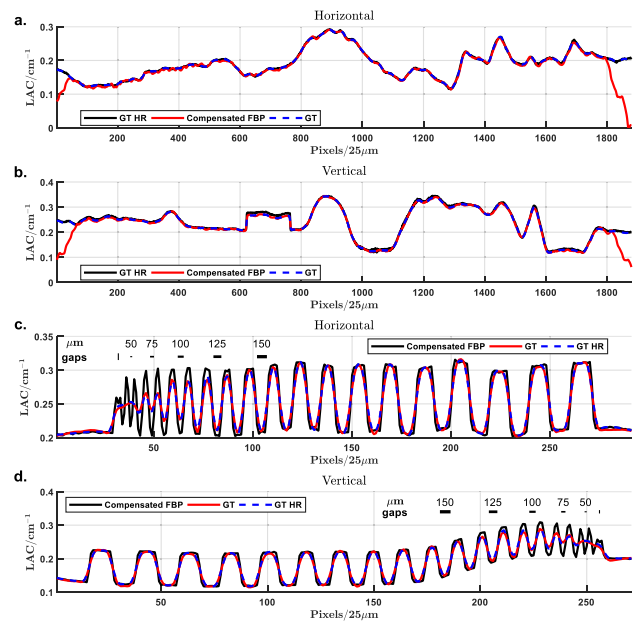


FIGURE 8. Attenuation profiles of the local reconstructions after background compensation relative to the ground truth. (a) and (b) The profiles along the horizontal and vertical midlines for attenuation value comparison; (c) and (d) the profiles through the horizontal and vertical resolution bar patterns for image resolution assessment.

which indicates the resolution of the local scan protocol is between $50\text{ }\mu\text{m}$ and $75\text{ }\mu\text{m}$. This agrees well with our analysis on the imaging parameters.

B. MIS-REGISTRATION EFFECTS

While the accuracy and resolution have been demonstrated above, the feasibility of the proposed reconstruction method will be illustrated in this subsection in terms of the robustness to the a potential mis-registration due to imperfect hardware components and their suboptimal coordination; i.e., with respect to mismatches in position, orientation and/or scale. The direct effects of these mismatches on the reconstruction process are an isocenter offset, a falsely tilted initial view angle, and an incorrect magnification factor messing up the interpolation between the local scan and the re-projection

through a globally reconstructed image volume for background estimation.

First, the mismatches in the isocenter position were set to the range from 0 to 4.483 mm with an increment of 0.498 mm along the horizontal direction. The corresponding results with misaligned background compensation are in Fig. 13. The absolute and relative error maps were calculated against the GT. Then, the profiles along the horizontal and vertical midlines of the error maps corresponding to 6 selected mismatches are in Fig. 9. It is observed that the maximum relative error of the well-aligned compensated local reconstruction is within $\pm 5\%$, which came mainly from the error in the background estimation, while the misalignment along the horizontal direction seems having little effect on the vertical profiles while there are significant drops off the horizontal profiles. Specifically, the maximum relative error remains around 6% when the misalignment is 0.498 mm, and it becomes close to 10% when the misalignment reaches 0.996 mm. The above observations suggest the robustness of the compensated local reconstruction with respect to the isocenter misalignment which remains quite accurate even with an up to 0.498 mm mismatch.

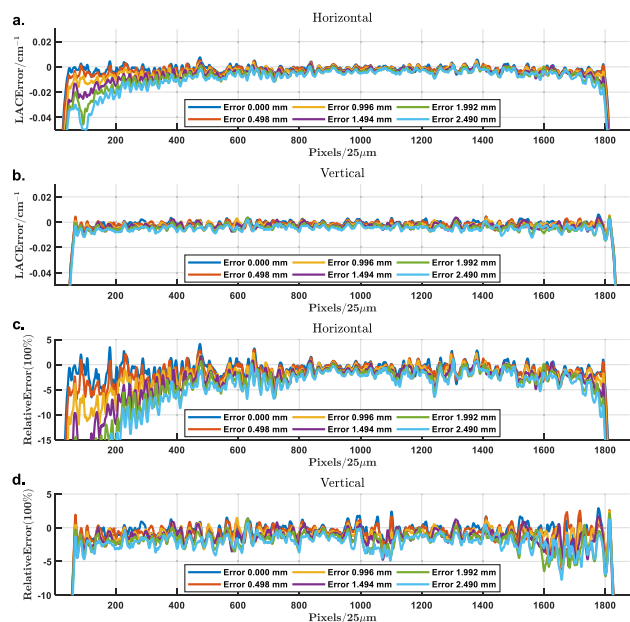


FIGURE 9. Profiles of the attenuation error in the reconstructions with isocenter misalignment after background compensation. (a) The horizontal and (b) vertical profiles of the attenuation error through the VOI center; (c) and (d) the corresponding relative error profiles. The profiles are associated with isocenter mismatches from 0 to 2.490 mm in a step of 0.498 mm.

The initial view angle mismatches were simulated in the range from -4.39° to 4.39° with an increment of 1.098° . The resultant reconstructions are in Fig. 14. The error maps were calculated in the same way as for the positional mismatches, and the through-center profiles of the error maps are in Fig. 10. Similarly, the mismatches have stronger influences on the horizontal profiles than on the vertical ones, by the

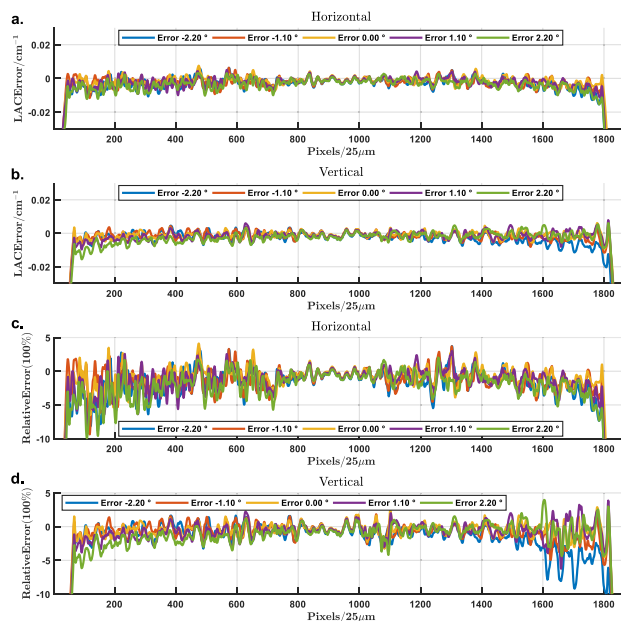


FIGURE 10. Profiles of the attenuation error in the reconstructions with misaligned initial view angles for background compensation. (a) and (b) The horizontal and vertical profiles through the center of the attenuation error map, (c) and (d) the corresponding relative reconstruction error profiles. The profiles correspond to angular mismatches from -2.2° to 2.2° with an increment of 1.1° .

directional asymmetry of the background compensation. The maximum relative error is in the range $[-10\%, 5\%]$ when the angular errors are within $\pm 2.2^\circ$. Especially, the relative reconstruction errors are mostly contained within $\pm 6\%$ if the angular errors are within $\pm 1.1^\circ$.

The magnification errors were also simulated in the range from -20% to 20% in a step of 5% , and the corresponding reconstructions are in Fig. 15. Interestingly, the “effective” ROI of reconstruction is determined by the applied magnification factor, with the regions outside the effective ROI having opposite attenuation shifting as compared to that inside the effective ROI. The relative reconstruction error profiles are in Fig. 11 with the corresponding magnification mismatches from -10% to 10% . Different from the distortions observed in the cases of positional and angular misalignments, which are mainly concentrated around a peripheral region, the magnification error causes a global attenuation shift inside the effective ROI. As shown in Fig. 11(a) and (b), the attenuation shift is proportional to the magnification mismatch and seems much more sensitive than other types of misalignments. Fortunately, this global shift can be effectively addressed by a bias correction method introduced below.

C. BIAS CORRECTION

The global shift, especially observed in the case with magnification errors, can be addressed with bias correction to make the attenuation value of a known region agree with the target value; i.e., to make the air value close to zero. Other known values can be used for the same purpose, such as that

TABLE 1. Quantitative metrics of the reconstructions with isocenter misalignment for background compensation ($r=21mm$). Bold indicates the reference.

Position Error	aligned	0.498mm	0.996mm	1.494mm	1.992mm	2.490mm	2.989mm	3.487mm	3.985mm	4.483mm
SSIM	0.9998	0.9998	0.9996	0.9993	0.9989	0.9984	0.9977	0.9970	0.9961	0.9952
PSNR	46.2150	44.9155	42.3074	39.7698	37.6195	35.8479	34.4174	33.2073	32.1431	31.2084
MSE $\times 10^{-4}$	0.0454	0.0613	0.1117	0.2004	0.3288	0.4944	0.6873	0.9081	1.1603	1.4389
RMSRE	0.0107	0.0125	0.0171	0.0231	0.0298	0.0367	0.0434	0.0500	0.0566	0.0631
SSIM (Crt.)	0.9998	0.9998	0.9998	0.9997	0.9996	0.9993	0.9991	0.9988	0.9984	0.9980
PSNR (Crt.)	46.3106	46.2296	45.1858	43.6330	41.8483	40.1459	38.7410	37.5219	36.4183	35.4465
MSE $\times 10^{-4}$ (Crt.)	0.0444	0.0453	0.0576	0.0823	0.1242	0.1838	0.2540	0.3363	0.4336	0.5423
RMSRE (Crt.)	0.0104	0.0105	0.0120	0.0146	0.0182	0.0225	0.0266	0.0308	0.0350	0.0393

TABLE 2. Quantitative metrics of the reconstructions with initial angular misalignment for background compensation ($r=21mm$). Bold indicates the reference.

Angle Error	-4.39°	-3.29°	-2.20°	-1.10°	aligned	1.10°	2.20°	3.29°	4.39°
SSIM	0.9991	0.9994	0.9996	0.9998	0.9998	0.9998	0.9996	0.9994	0.9991
PSNR	38.4776	40.0098	42.1055	44.8752	46.2150	44.3655	41.9105	39.9452	38.4152
MSE $\times 10^{-4}$	0.2699	0.1896	0.1170	0.0619	0.0454	0.0696	0.1224	0.1925	0.2738
RMSRE	0.0273	0.0226	0.0176	0.0127	0.0107	0.0128	0.0166	0.0206	0.0244
SSIM (Crt.)	0.9995	0.9997	0.9998	0.9998	0.9998	0.9998	0.9998	0.9997	0.9995
PSNR (Crt.)	41.8476	43.3071	44.8716	46.1801	46.3105	46.1733	45.0336	43.3808	41.7435
MSE $\times 10^{-4}$ (Crt.)	0.1242	0.0887	0.0619	0.0458	0.0444	0.0459	0.0596	0.0873	0.1272
RMSRE (Crt.)	0.0179	0.0149	0.0123	0.0105	0.0104	0.0108	0.0121	0.0144	0.0172

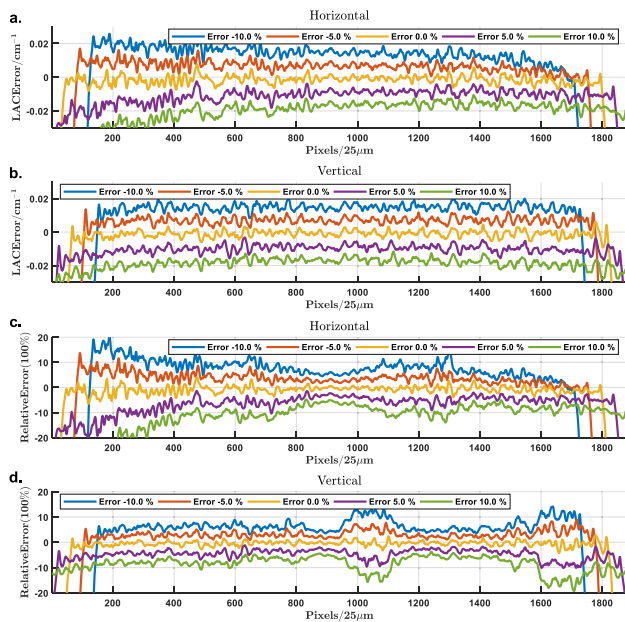


FIGURE 11. Profiles of the attenuation error in the reconstructions with incorrect magnification factors for background compensation. (a) and (b) The horizontal and vertical profiles through the reconstruction error maps; (c) and (d) the corresponding relative reconstruction error profiles. The profiles correspond to magnification mismatches from -10% to 10% with an increment of 5%.

obtained from the global reconstruction; for example, we can select a relatively flat region and calculate its mean value as a benchmark.

Quantitative metrics, including SSIM, Peak Signal-to-Noise Ratio (PSNR), MSE and Root Mean-Squared Relative Error (RMSRE), were used to quantify the reconstructions with these misalignments in reference to the ROI in the GT,

with and without additional bias correction. The actual radius of the ROI was set to 21 mm. The evaluation results with respect to isocenter position, initial angle, and magnification errors are summarized in Tables 1, 2 and 3 respectively. The results after bias correction are denoted with the abbreviation ‘Crt’ in the Tables.

The reconstructions are quite robust with respect to the isocenter positional and initial angular errors, and the bias correction method can further improve the accuracy, as shown in Tables 1 and 2. The attenuation deviation from GT in the reconstruction with aligned background compensation is very small, with MSE 0.454×10^{-5} and RMSRE only 1.07%. By increasing the position error, the SSIM and PSNR metrics decrease while MSE and RMSRE increase, as expected. The RMSRE value is still below 2.0% when the position error reaches 0.996 mm, and the tolerance is extended to up to 1.992mm after the bias correction, which demonstrate the robustness of our method. Similarly, in Table 2 the RMSRE remains below 2.0% for all angular errors within $\pm 4.39^\circ$.

The bias correction substantially improves the magnification-error-affected image reconstruction, as shown in Table 3. To be noted, since the effective ROI is scaled with the magnification factor, the intersection of these effective ROIs and the original ROI (radius 21 mm) was used for evaluation, with the radius being set to 16.8 mm. The metrics on the reconstruction before the bias correction dramatically change as the magnification error varies, demonstrating a relative high sensitivity. Fortunately, with the bias correction method RMSRE in the case of -20% magnification error is significantly reduced to 1.67% from 10.43% within the effective ROI, demonstrating a decent robustness.

An interesting phenomenon is that if we perform the bias correction on the aligned reconstruction, the metric scores drop slightly in the inner region within the radius of 16.8 mm

TABLE 3. Quantitative metrics of the reconstructions with magnification mismatch for background compensation ($r=16.8$ mm). Bold indicates the reference.

Mag. Error	-20%	-15%	-10%	-5%	aligned	5%	10%	15%	20%
SSIM	0.9828	0.9907	0.9961	0.9991	0.9999	0.9980	0.9933	0.9866	0.9785
PSNR	26.3718	29.2189	33.1242	39.3940	49.0959	36.5190	31.3849	28.5055	26.5721
MSE $\times 10^{-4}$	4.3823	2.2750	0.9257	0.2185	0.0234	0.4236	1.3816	2.6812	4.1848
RMSRE	0.1043	0.0751	0.0479	0.0232	0.0076	0.0323	0.0584	0.0813	0.1016
SSIM (Crt.)	0.9996	0.9998	0.9999	0.9999	0.9999	0.9999	0.9998	0.9997	0.9996
PSNR (Crt.)	41.9195	44.8545	47.0394	48.2090	48.0992	46.9567	45.1736	43.6877	42.4808
MSE $\times 10^{-4}$ (Crt.)	0.1222	0.0621	0.0376	0.0287	0.0294	0.0383	0.0577	0.0813	0.1074
RMSRE (Crt.)	0.0167	0.0122	0.0097	0.0084	0.0084	0.0095	0.0117	0.0138	0.0159

in contrast to a tiny boost when evaluated on the whole ROI region (radius 21 mm). This suggests that the reconstruction must already be very accurate in the inner region when the compensation is well aligned. The attenuation estimation from the global reconstruction cannot be perfect due to differences in resolution and existence of artifacts, and the linearity of the model to support our compensation method is just an approximation to the polychromatic X-ray imaging process, which may bring residual errors into the peripheral region of the ROI. Thus, there is no need to perform bias correction if the system is well calibrated and we are sure about the accuracy of the registration for background compensation. Otherwise, bias correction can be used for better performance.

V. DISCUSSION

In this study, we have proposed a novel CMCT system which incorporates a micro-focus source, a PCD, robotic arms and advanced imaging algorithms into a synergistic companion of a conventional CT scanner. The designed HR local scan protocol not only improves the dose efficiency but also reduces the area of detectors. The cost of PCDs currently remains high due to the complex manufacturing techniques, and the situation is likely to continue in the near future. Hence, this hybrid system for interior tomography can reduce the system cost and radiation dose without compromising the required performance. In the local/interior scan, the advanced robotic arms allow the free selection of a VOI, which is preferable compared to the traditional rotating gantry. In addition, the mobility of the robotic system enables surgeons to take projections from any view angle without moving the patient, which can be extremely helpful in many applications, such as high-quality evaluation in emergencies and real-time feedback in surgeries. It is worth mentioning that besides the exemplary application in inner ear imaging, the system can also work for other clinical imaging tasks that demand high resolution in a VOI/ROI, such as tumor examination in breast, nodule characterization in lung, bone quality analysis, and plaque imaging in the heart and the neck.

Besides the hybrid design as one imaging system shown in Fig 1, the robotic micro-CT scanner can also be separately used. The reconstruction results from a traditional CT scanner can be used as the prior knowledge following the same procedure described in Section III. However, this may impose extra

work in registration due to different positions of the patient in the local scan and an earlier global scan. Luckily, as analyzed in Section IV-B, the proposed reconstruction method has a good tolerance to geometric misalignment. Although the reconstruction is relatively sensitive to the magnification mismatch, the resultant attenuation shifting can be addressed with the proposed bias correction method. In addition, with the rapid development of face recognition technology [38], human facial surface measurement techniques have matured with high accuracy in real-time [32], [39]–[41]. The registration between surfaces have been studied for many years. High-quality toolboxes, such as 3D slicer [42], can be directly used or adapted for our purpose.

A. PROJECTION DEBLURRING

Another challenge for CMCT comes from the X-ray source. The intensity of an X-ray source with micro-focus has usually an insufficient flux to produce an appropriate signal to noise ratio through a human head during a reasonably short time. To obtain an appropriate contrast, an X-ray source with a slightly larger focal spot may be used to provide enough power. The increased focal spot could generate shadows in the projections and blur structural details. In addition, the balance between the X-ray intensity and image resolution can be optimized using a deep learning deblurring method. Although the cone-beam projection with the finite focus spot is no longer a spatially invariant linear system, which is a challenge with traditional deblurring methods, advanced deep learning techniques have the capability to perform shift-variant deblurring tasks [43]. The big data of paired blurred-original projections may be difficult to obtain for training a deblurring network. Alternatively, a forward projection model can be easily built to realistically synthesize the paired data. Then, the network trained with simulated data is fine-tuned with a small amount of paired real projection data. Finally, the trained network can be applied on blurred projections for inference, similar to what we did in an optical deblurring study [44].

B. IMAGE DENOISING

Deep denoising techniques can be used to reduce radiation dose and improve image quality. According to the level of supervision during training, three types of deep denoising methods have been developed, i.e., supervised learning, weakly-supervised learning, and unsupervised

learning. Supervised learning methods were designed for image denoising and achieved the best performance, such as deep CNNs with residual learning [45], [46] or with recurrent persistent memory units [47]. Weakly-supervised learning methods relax the requirement of paired noisy-clean data to unpaired noisy-clean data [48] or paired noise-to-noise data [49]. Using the unpaired noisy-clean data, [48] proposed GAN-based learning to create pairs of corresponding noisy-clean images as the training data. Recently, [49] demonstrated that paired noise-to-noise images are equivalent to the paired noisy-clean images in training a model, achieving a denoising performance competitive with supervised learning methods. For the applications where even the unpaired noisy-clean or paired noise-noise images are unavailable, unsupervised learning methods were proposed using only single noisy images for training. Deep image prior [50] is a generation process that maps the random noise to a single noisy image, and when they terminate the training process at the right moment the network produces a denoised image. Most recently, Noise2Void [51] and its variants [52], [53] achieved promising results only using individual noisy images in training a network. Basically, a Noise2Void network estimates a blind-spot in an image so that the network learns to map the surrounding pixels to the blind-spot, achieving excellent denoising results.

In our inner ear imaging application, paired noise-clean images can be synthesized via Monte-Carlo simulation, and single real noisy images can be acquired with the proposed CMCT system. The former data type can support supervised training although the noise may not perfectly match the real counterpart, while the latter type of data contains realistic noise and texture. Combining these two types of datasets, we can design a semi-supervised learning method to learn from the data with and without ground truth labels simultaneously. For example, the model can be trained in the Noise2Void mode first and then fine tuned with the paired noisy-clean data or vice versa. An alternative is to train the model in the Noise2Void and supervised modes simultaneously.

C. MATERIAL DECOMPOSITION AND BEAM HARDENING CORRECTION

In addition to the high resolution advantage, if the local/interior scan is performed in the multi-channel photon-counting mode, the energy information can be used for spectral analysis; e.g., K-edge imaging, material decomposition, beam hardening correction, and metal artifact reduction. Compared with traditional dual-energy CT, the PCD provides more energy channels and is more informative while the potential spectral distortion issues at high imaging speed can be overcome with a deep learning based correction method [54]. In principle, the direct spectral measurement with the PCD allows better spectral separation than dual-source, fast kVp-switching, and dual-layer detector techniques. Most relevant to this inner ear imaging application is

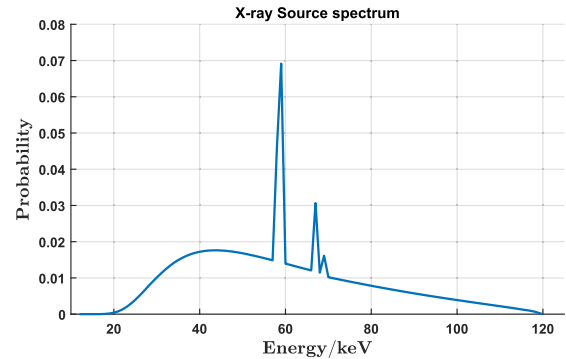


FIGURE 12. The simulated spectrum $p(E)$ of our tungsten X-ray source operated at 120 kV.

to utilize the X-ray energy dependent attenuation information for beam hardening correction, metal artifact reduction and material decomposition [8] so that the effects of the implanted electrodes and micro-environments can be optimally modeled.

VI. CONCLUSION

In conclusion, we have proposed a clinical micro-CT (CMCT) system empowered with a number of cutting-edge technologies for regional ultrahigh resolution imaging of a VOI, which is particularly suitable for human inner ear/temporal bone imaging. The background compensation technique has been proposed for fast and accurate local reconstruction with high resolution at a minimized X-ray dose, taking full advantage of prior information from a conventional medical CT scan. The HR feature, attenuation fidelity and robustness to geometric misalignment in the registration between the global and local scans have been demonstrated, establishing the feasibility of CMCT. Overall, our CMCT system is promising for inner ear imaging and other clinical applications.

APPENDIX I. THE X-RAY TUBE POWER JUSTIFICATION

To justify the selection of our X-ray source, we measured the number of photons emission per solid angle per second from a similar micro-focus X-ray source (Source-Ray, SB-120-350, a spot size of $s_0 = 75 \mu\text{m}$) with a PCD (CZT-Medipix3RX, $110 \mu\text{m}$ pitch size d_0 , with anti-charge sharing ASICs) as a reference. The source to detector distance c_0 was 272 mm . The tube was operated at 120 kV and $I_0 = 19 \mu\text{A}$ with a 1.96 mm aluminum filter. The measured average counts per second is 1.03×10^4 per pixel, corresponding to $N_0 = 6.30 \times 10^{10}$ counts per solid angle per second. The distribution of photons $p(E)$ over the energy range was simulated with SpekCalc [55], as shown in Fig. 12.

According to one study in the United States [56], the average human head circumference is 56.9 cm in males and 53.4 cm in females, corresponding to diameters of $\sim 18.2 \text{ cm}$ and $\sim 17.0 \text{ cm}$ respectively. Without loss of generality, the value 20 cm is a reasonable estimation of the maximum length of the X-ray beam passing through the head during

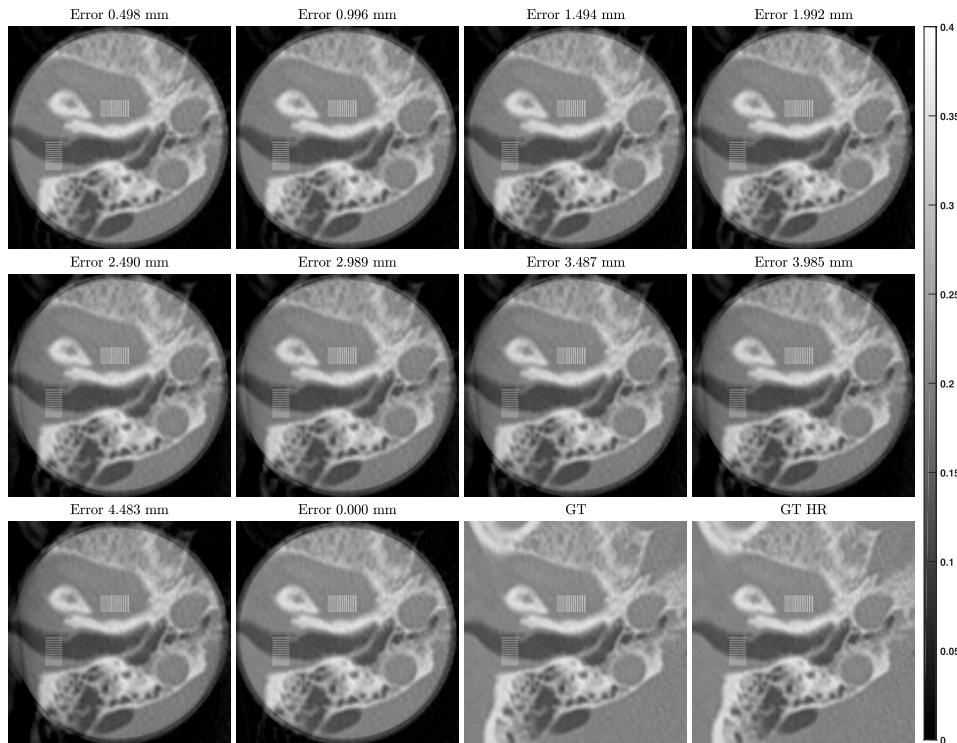


FIGURE 13. ROI Reconstructions with misalignment of isocenter position in background compensation, together with the ground truths (GT and GT HR), are displayed in the same window (unit of cm^{-1}). The misalignment errors are from 0 to 4.483mm with an increment of 0.498mm.

a scan. Considering the mean thickness of skull of 6.32 mm [57], we further assume that the path consists of $l_b = 1.2\text{ cm}$ bones and $l_t = 18.8\text{ cm}$ brain tissues.

Based on the CMCT design (the source to detector distance of $c_m = 400\text{mm}$, and the detector element size of $d_m = 0.11\text{mm}$), when the selected source is operated at 120 kV and $I_m = 500\mu A$ with a focus spot of $s_m = 50\mu m$ and assumed with a similar electron-to-photon conversion efficiency as that of our experiment source, the minimum counts (with the maximum length of path through head) received per second by one pixel on the detector can be estimated after the corrections of the tube current and focus size as follows:

$$N_m = N_0 \frac{d_m^2 I_m s_m^2}{c_m^2 I_0 s_0^2} \int p(E) e^{-l_b \mu_b(E) - l_t \mu_t(E)} dE = 474.3 \quad (5)$$

Note that this is the counts received by the darkest pixel in the projections which still provides a decent signal. In our VOI-centered scan, the projections of the other views could be much brighter than this worst case; e.g., if we reduce the l_t by half, the counts per second reaches $N_m = 3185.0$. Not to mention that the selected source can be operated at 150 kV which offers a much higher penetration ability. Hence, by taking 720 projections with 1-second exposure, we can finish the scan within 12 minutes and obtain a good-quality reconstruction.

In addition, innovative technologies like NanoX tube [9] and MAAT tube [10] can significantly raise the brightness of the micro-focus X-ray source by several folds, which will

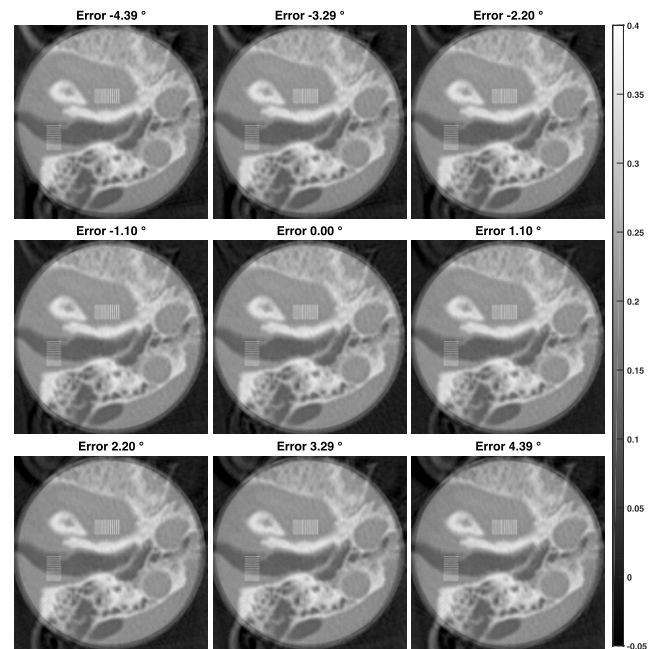


FIGURE 14. ROI Reconstructions with misalignment of initial view angle in background compensation displayed in the same window (unit: cm^{-1}). The misalignment errors are from -4.39° to 4.39° with an increment of 1.098° .

definitely facilitate our undertaking and shorten the scan time. On the other hand, deep learning techniques can be employed for further scan acceleration and dose reduction. For example,

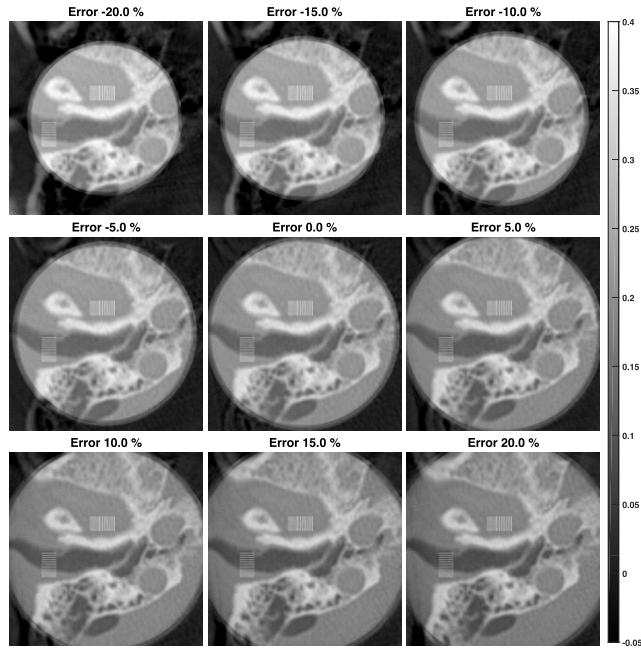


FIGURE 15. ROI Reconstructions with misalignment of magnification factor in background compensation displayed in the same window (unit: cm^{-1}). The misalignment errors are from -20% to 20% with a step of 5% .

deep denoising and few-view reconstruction can be utilized to reduce exposure time and the number of views, respectively, while maintaining the reconstruction quality. Thus, it is feasible to reduce the scanning time to 3 minutes or less and still provide a good-quality reconstruction.

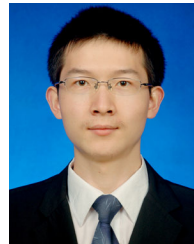
APPENDIX II. ROI RECONSTRUCTIONS WITH MIS-ALIGNMENTS

ROI reconstructions through background compensation with various mis-alignments in terms of isocenter position, initial view angle, and magnification factor, as shown in Fig. 13, 14 and 15.

REFERENCES

- [1] R. Saadi, J. Liaw, A. Baker, and H. Sildak, "Improving microscopic imaging in otology and neurotology," *Otol. Neurotol.*, vol. 41, no. 4, pp. e464–e467, 2020.
- [2] C. Rasch, I. Barillot, P. Remeijer, A. Touw, M. van Herk, and J. V. Lebesque, "Definition of the prostate in CT and MRI: A multi-observer study," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 43, no. 1, pp. 57–66, 1999.
- [3] R. Balachandran, D. Schurzig, J. M. Fitzpatrick, and R. F. Labadie, "Evaluation of portable CT scanners for otologic image-guided surgery," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 7, no. 2, pp. 315–321, 2012.
- [4] G. S. Phillips, S. E. LoGerfo, M. L. Richardson, and Y. Anzai, "Interactive Web-based learning module on CT of the temporal bone: Anatomy and pathology," *Radiographics*, vol. 32, no. 3, pp. E85–E105, 2012.
- [5] A. Urikura, T. Hara, T. Yoshida, E. Nishimaru, T. Hoshino, Y. Nakaya, and M. Endo, "Overranging and overbeaming measurement in area detector computed tomography: A method for simultaneous measurement in volume helical acquisition," *J. Appl. Clin. Med. Phys.*, vol. 20, no. 7, pp. 160–165, 2019.
- [6] H. M. Kjer, J. Fagertun, W. Wimmer, N. Gerber, S. Vera, L. Barazzetti, N. Mangado, M. Ceresa, G. Piella, T. Stark, M. Stauber, M. Reyes, S. Weber, M. Caversaccio, M. Á. G. Ballester, and R. R. Paulsen, "Patient-specific estimation of detailed cochlear shape from clinical ct images," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 13, no. 3, pp. 389–396, 2018.
- [7] S. Capostagno, J. W. Stayman, M. W. Jacobson, T. Ehtiati, C. R. Weiss, and J. H. Siewerdsen, "Task-driven source-detector trajectories in cone-beam computed tomography: II. Application to neuroradiology," *J. Med. Imag.*, vol. 6, no. 2, 2019, Art. no. 025004.
- [8] M. J. Willeminck, M. Persson, A. Pourmorteza, N. J. Pelc, and D. Fleischmann, "Photon-counting CT: Technical principles and clinical prospects," *Radiology*, vol. 289, no. 2, pp. 293–312, 2018.
- [9] T. Hori, H. Masuya, and H. Kenmotsu, "Image capture device," U.S. Patent 9922 793, Mar. 20, 2018.
- [10] G. Zan, D. J. Vine, R. I. Spink, W. Yun, Q. Wang, and G. Wang, "Design optimization of a periodic microstructured array anode for hard X-ray grating interferometry," *Phys. Med. Biol.*, vol. 64, no. 14, 2019, Art. no. 145011.
- [11] L. Li, Z. Chen, X. Jin, H. Yu, and G. Wang, "Experimental measurement of human head motion for high-resolution computed tomography system design," *Opt. Eng.*, vol. 49, no. 6, 2010, Art. no. 063201.
- [12] G. Wang, S. Zhao, H. Yu, C. A. Miller, P. J. Abbas, B. J. Gantz, S. W. Lee, and J. T. Rubinstein, "Design, analysis and simulation for development of the first clinical micro-CT scanner," *Acad. Radiol.*, vol. 12, no. 4, pp. 511–525, 2005.
- [13] V. Patel, K. Hoffmann, C. Ionita, C. Keleshis, D. Bednarek, and S. Rudin, "Rotational micro-CT using a clinical C-arm angiography gantry," *Med. Phys.*, vol. 35, no. 10, pp. 4757–4764, 2008.
- [14] L. Chen, Y. Shen, C.-J. Lai, T. Han, Y. Zhong, S. Ge, X. Liu, T. Wang, W. T. Yang, G. J. Whitman, and C. C. Shaw, "Dual resolution cone beam breast CT: A feasibility study," *Med. Phys.*, vol. 36, no. 9, pp. 4007–4014, 2009.
- [15] D. Kolditz, Y. Kyriakou, and W. A. Kalender, "Volume-of-interest (VOI) imaging in C-arm flat-detector CT for high image quality at reduced dose," *Med. Phys.*, vol. 37, no. 6, pp. 2719–2730, 2010.
- [16] D. Kolditz, T. Struffert, Y. Kyriakou, A. Bozzato, A. Dörfler, and W. Kalender, "Volume-of-interest imaging of the inner ear in a human temporal bone specimen using a robot-driven C-arm flat panel detector CT system," *Amer. J. Neuroradiol.*, vol. 33, no. 10, pp. E124–E128, 2012.
- [17] A. Jain, H. Takemoto, M. Silver, S. Nagesh, C. Ionita, D. Bednarek, and S. Rudin, "Region-of-interest cone beam computed tomography (ROI CBCT) with a high resolution CMOS detector," *Proc. SPIE*, vol. 9412, Mar. 2015, Art. no. 94124L.
- [18] A. Reshef, C. Riddell, Y. Troussel, S. Ladjal, and I. Bloch, "Dual-rotation C-arm cone-beam computed tomography to increase low-contrast detection," *Med. Phys.*, vol. 44, no. 9, pp. e164–e173, 2017.
- [19] Y. Ye, H. Yu, Y. Wei, and G. Wang, "A general local reconstruction approach based on a truncated Hilbert transform," *Int. J. Biomed. Imag.*, vol. 2007, Jun. 2007, Art. no. 063634.
- [20] H. Yu and G. Wang, "Compressed sensing based interior tomography," *Phys. Med. Biol.*, vol. 54, no. 9, p. 2791, 2009.
- [21] K. S. Sharma, C. Holzner, D. M. Vasilescu, X. Jin, S. Narayanan, M. Agah, E. A. Hoffman, H. Yu, and G. Wang, "Scout-view assisted interior micro-CT," *Phys. Med. Biol.*, vol. 58, no. 12, p. 4297, 2013.
- [22] S. Jin, K. Shin, S. Yoo, J. Kim, K. Kim, Y. Huh, S. Lee, and O.-K. Kwon, "Dual-resolution image reconstruction for region-of-interest CT scan," *J. Instrum.*, vol. 9, no. 7, 2014, Art. no. C07008.
- [23] H. Kudo, "Advanced compressed sensing image reconstruction for interior tomography," *Proc. SPIE*, vol. 11113, Sep. 2019, Art. no. 111130U.
- [24] Z. Zhu, A. Katsevich, and S. Pang, "Interior X-ray diffraction tomography with low-resolution exterior information," *Phys. Med. Biol.*, vol. 64, no. 2, 2019, Art. no. 025009.
- [25] S. Azevedo, P. Rizo, and P. Grangeat, "Region-of-interest cone-beam computed tomography," Lawrence Livermore Nat. Lab., Livermore, CA, USA, Tech. Rep. UCRL-ID-121668 ON: DE96002174; TRN: 95:008155, 1995. [Online]. Available: <https://www.osti.gov/biblio/125412>, doi: 10.2172/125412.
- [26] I. K. Chun, M. H. Cho, S. C. Lee, M. H. Cho, and S. Y. Lee, "X-ray micro-tomography system for small-animal imaging with zoom-in imaging capability," *Phys. Med. Biol.*, vol. 49, no. 17, p. 3889, 2004.
- [27] C. Maaß, M. Knaup, and M. Kachelrieß, "New approaches to region of interest computed tomography," *Med. Phys.*, vol. 38, no. 6, pp. 2868–2878, 2011.
- [28] V. Subramaniam, S. Jain, J. Agarwal, and P. Y. V. Alvarado, "Design and characterization of a hybrid soft gripper with active palm pose control," *Int. J. Robot. Res.*, vol. 39, no. 14, pp. 1668–1685, 2020, doi: 10.1177/0278364920918918.
- [29] G. Li, S. Luo, C. You, M. Getzin, L. Zheng, G. Wang, and N. Gu, "A novel calibration method incorporating nonlinear optimization and ball-bearing markers for cone-beam CT with a parameterized trajectory," *Med. Phys.*, vol. 46, no. 1, pp. 152–164, 2019.

- [30] A. A. Hendriksen, D. M. Pelt, and K. J. Batenburg, "Noise2inverse: Self-supervised deep convolutional denoising for tomography," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 1320–1335, 2020.
- [31] M. Feser, J. Gelb, H. Chang, H. Cui, F. Duewer, S. Lau, A. Tkachuk, and W. Yun, "Sub-micron resolution CT for failure analysis and process development," *Meas. Sci. Technol.*, vol. 19, no. 9, 2008, Art. no. 094001.
- [32] T. Bakirman, M. U. Gumusay, H. C. Reis, M. O. Selbesoglu, S. Yosmaoglu, M. C. Yaras, D. Z. Seker, and B. Bayram, "Comparison of low cost 3D structured light scanners for face modeling," *Appl. Opt.*, vol. 56, no. 4, pp. 985–992, 2017.
- [33] F. Natterer, *The Mathematics of Computerized Tomography*. Philadelphia, PA, USA: SIAM, 2001.
- [34] M. Chen, Y. Xi, W. Cong, B. Liu, B. Wei, and G. Wang, "X-ray CT geometrical calibration via locally linear embedding," *J. X-Ray Sci. Technol.*, vol. 24, no. 2, pp. 241–256, Mar. 2016.
- [35] M. Chen, P. He, P. Feng, B. Liu, Q. Yang, B. Wei, and G. Wang, "General rigid motion correction for computed tomography imaging based on locally linear embedding," *Opt. Eng.*, vol. 57, no. 2, 2018, Art. no. 023102.
- [36] M. J. Ackerman, "The visible human project," *Proc. IEEE*, vol. 86, no. 3, pp. 504–511, Mar. 1998.
- [37] B. De Man, S. Basu, N. Chandra, B. Dunham, P. Edic, M. Iatrou, S. McOlash, P. Sainath, C. Shaughnessy, B. Tower, E. Williams, "CatSim: A new computer assisted tomography simulation environment," *Proc. SPIE*, vol. 6510, Mar. 2007, Art. no. 65102G.
- [38] O. M. Parkhi, A. Vedaldi, and A. Zisserman, *Deep Face Recognition*. British Machine Vision Association, 2015, pp. 1–12. [Online]. Available: <https://ora.ox.ac.uk/objects/uuid:a5f2e93f-2768-45bb-8508-74747f85cad1>
- [39] L. Ma, T. Xu, and J. Lin, "Validation of a three-dimensional facial scanning system based on structured light techniques," *Comput. Methods Programs Biomed.*, vol. 94, no. 3, pp. 290–298, 2009.
- [40] A. Brahm, R. Ramm, S. Heist, C. Rulff, P. Kühmstedt, and G. Notni, "Fast 3D NIR systems for facial measurement and lip-reading," *Proc. SPIE*, vol. 10220, May 2017, Art. no. 102200P.
- [41] P. G. M. Knoops, C. A. A. Beaumont, A. Borghi, N. Rodriguez-Florez, R. W. F. Breakey, W. Rodgers, F. Angullia, N. U. O. Jeelani, S. Schievano, and D. J. Dunaway, "Comparison of three-dimensional scanner systems for craniomaxillofacial imaging," *J. Plastic, Reconstructive Aesthetic Surg.*, vol. 70, no. 4, pp. 441–449, 2017.
- [42] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J.-C. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. Fennessy, M. Sonka, J. Buatti, S. Aylward, J. V. Miller, S. Pieper, and R. Kikinis, "3D slicer as an image computing platform for the quantitative imaging network," *Magn. Reson. Imag.*, vol. 30, no. 9, pp. 1323–1341, 2012.
- [43] H. Wang, Y. Rivenson, Y. Jin, Z. Wei, R. Gao, H. Günaydin, L. A. Bentolila, C. Kural, and A. Ozcan, "Deep learning enables cross-modality super-resolution in fluorescence microscopy," *Nature Methods*, vol. 16, pp. 103–110, Jan. 2019.
- [44] M. Li, H. Shan, S. Pryschchep, M. M. Lopez, and G. Wang, "Deep adversarial network for super stimulated emission depletion imaging," *J. Nanophoton.*, vol. 14, no. 1, 2020, Art. no. 016009.
- [45] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian Denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.
- [46] X. Mao, C. Shen, and Y.-B. Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 2802–2810.
- [47] Y. Tai, J. Yang, X. Liu, and C. Xu, "MemNet: A persistent memory network for image restoration," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 4539–4547.
- [48] J. Chen, J. Chen, H. Chao, and M. Yang, "Image blind denoising with generative adversarial network based noise modeling," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3155–3164.
- [49] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2Noise: Learning image restoration without clean data," 2018, *arXiv:1803.04189*. [Online]. Available: <http://arxiv.org/abs/1803.04189>
- [50] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9446–9454.
- [51] A. Krull, T.-O. Buchholz, and F. Jug, "Noise2void-learning denoising from single noisy images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 2129–2137.
- [52] A. Krull, T. Vicar, and F. Jug, "Probabilistic Noise2Void: Unsupervised content-aware denoising," 2019, *arXiv:1906.00651*. [Online]. Available: <http://arxiv.org/abs/1906.00651>
- [53] S. Laine, T. Karras, J. Lehtinen, and T. Aila, "High-quality self-supervised deep image denoising," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 6970–6980.
- [54] M. Li, D. S. Rundle, and G. Wang, "X-ray photon-counting data correction through deep learning," 2020, *arXiv:2007.03119*. [Online]. Available: <http://arxiv.org/abs/2007.03119>
- [55] G. Poludniowski, G. Landry, F. Deblois, P. Evans, and F. Verhaegen, "SpekCalc: A program to calculate photon spectra from tungsten anode X-ray tubes," *Phys. Med. Biol.*, vol. 54, no. 19, p. N433, 2009.
- [56] R. P. Ching, "Relationship between head mass and circumference in human adults," Univ. Washington Tech. Brief, Seattle, WA, USA, Tech. Rep., 2007. Accessed: Dec. 12, 2020. [Online]. Available: <http://www.smf.org/docs/articles/pdf/chingtechbrief.pdf>
- [57] A. Moreira-Gonzalez, F. E. Papay, and J. E. Zins, "Calvarial thickness and its relation to cranial bone harvest," *Plastic Reconstructive Surg.*, vol. 117, no. 6, pp. 1964–1971, 2006.



MENGZHOU LI (Graduate Student Member, IEEE) received the B.S. degree in optoelectronic information engineering and the M.S. degree in instrument science and technology from the Harbin Institute of Technology, in 2014 and 2016, respectively. He is currently pursuing the Ph.D. degree with the Department of Biomedical Engineering, Rensselaer Polytechnic Institute, Troy, NY, USA.

His research interests include X-ray tomographic imaging, super-resolution light microscopy, and machine learning. He received awards and honors include the National Scholarship, Ministry of Education of China, the China Instrument and Control Society Scholarship, in 2014, and the First-Class Science and Technology Progress Award (Chinese Society for Measurement), in 2018.



ZHENG FANG was born in Dushan, Guizhou, China, in 1976. He received the B.S. degree in automation from the Huazhong University of Science and Technology, Wuhan, Hubei, China, in 1998, the M.S. degree in automation from the Naval University of Engineering, Wuhan, in 2001, and the Ph.D. degree in biomedical engineering from the Huazhong University of Science and Technology, in 2008.

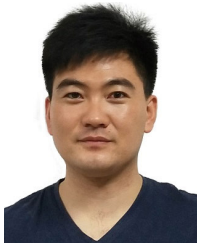
From 2001 to 2003, he was an Engineer with ZTE Corporation, Shenzhen, China. From 2008 to 2011, he held a postdoctoral position in automation with the Huazhong University of Science and Technology. From 2011 to 2016, he was an Associate Professor with the Department of Mechanical and Electrical Engineering, Xiamen University, Xiamen, China. Since 2016, he has been a Professor with the Department of Instrumental and Electrical Engineering, Xiamen University. He has published 60 articles and holds ten patents. His research interests include X-ray detection, biomedical imaging and non-destructive testing.

Dr. Fang was awarded a scholarship under the State Scholarship Fund from China Scholarship Council, in 2019. He was a Communication Reviewer of the National Natural Science Foundation of China.



WENXIANG CONG received the B.S. degree in mathematics from Heilongjiang University, Harbin, China, in 1982, the M.S. degree in applied mathematics from the Harbin Institute of Technology, Harbin, in 1988, and the Ph.D. degree in optical imaging from the Beijing University of Science and Technology, Beijing, China, in 1998. He is currently a Research Scientist with the Department of Biomedical Engineering, Rensselaer Polytechnic Institute, Troy, NY, USA. His research interests

include optical molecular tomographic imaging, X-ray computed tomographic imaging, and machine learning.



CHUANG NIU received the B.S. and Ph.D. degrees from Xidian University, Xi'an, China, in 2015 and 2020, respectively. He is currently a Postdoctoral Research Associate with Rensselaer Polytechnic Institute, Troy, NY, USA. His research interests include machine learning, computer vision, and deep imaging.



WEI WEN WU was born in Loudi, Hunan, China, in 1991. He received the Ph.D. degree in optoelectronic engineering from Chongqing University, Chongqing, in 2019. From 2017 to 2018, he was a Visiting Student with the University of Massachusetts Lowell under the supervision of Dr. H. Yu. From 2019 to 2020, he was a Postdoctoral Fellow with The University of Hong Kong. He has published over 35 academic journal articles in peer-reviewed journal publications. His research interests include deep learning, image reconstruction, compressed sensing, X-ray CT, MRI, and spectral CT. His awards and honors include the Chinese Scholarship Council, and the Non-Destructive Testing Scholarship, in 2019. He is a Guest Editor of the *EURASIP Journal on Advances in Signal Processing*.



JOSEF UHER received the Ph.D. degree in physics on research of sensors for neutron detection that were based on the Medipix technology. Medipix/Timepix chips are imaging devices designed in an international collaboration lead by CERN, Geneva, Switzerland. He is currently a Physicist with focus on radiation imaging, and CTO of Radalytica a.s. and ADVACAM s.r.o., Czech Republic. He worked in various research positions where most of them were focused on radiation detection and imaging. He was also part of research team developing DIRC detector at Stanford Linear Accelerator Center (SLAC, Menlo Park, CA). He participated in the chip evaluations and development of applications. Starting from his postdoctoral position in CSIRO, Australia. His research interests include X-ray imaging, spectral X-ray imaging, and computed tomography. He have been leading several Research and Development projects on application of these novel X-ray imaging detectors in variety of fields ranging from mining industry, going through art inspection to biology and preclinical research.



JAMES BENNETT received the B.S.E. degree in biomedical engineering from The University of Iowa, Iowa City, IA, USA, in 2005, and the Ph.D. degree in biomedical engineering from the School of Biomedical Engineering and Sciences, Virginia Tech, Blacksburg, VA, USA, and Wake Forest University, Winston-Salem, NC, USA, in 2014. From 2005 to 2010, he was a Consultant with Accenture, Chicago, IL, USA. From 2014 to 2015, he trained as a Postdoctoral Research Fellow in internal medicine-nuclear cardiology with the School of Medicine, Yale University, New Haven, CT, USA. Since 2015, he has been working as a Consultant and founded several companies. He is currently the Owner and CEO of Hawkeye Spectral Imaging, LLC. His research interests include spectral X-ray CT, computed tomography, biomedical imaging, multi-modality fusion, nuclear imaging, and non-destructive testing.

Dr. Bennett was a recipient of a NIH T32 Ruth L. Kirschstein National Research Service Award, in 2014.



JAY T. RUBINSTEIN (Senior Member, IEEE) received the Sc.B. and Sc.M. degrees in engineering from Brown University, in 1981 and 1983, respectively, and the M.D. and Ph.D. degrees in bioengineering from the University of Washington, in 1987 and 1988, respectively. He completed postdoctoral research training at MIT and residency in Otolaryngology with the Massachusetts Eye and Ear Infirmary, in 1994. He completed a Neurotology Fellowship with The University of Iowa, in 1995, staying as Assistant then Associate Professor of otolaryngology and bioengineering. He was the Boerhaave Professor with Leiden University, The Netherlands, in 2003 and 2004. He is currently a Virginia Merrill Bloedel Professor of otolaryngology and bioengineering and the Director of the Bloedel Hearing Research Center, University of Washington. He has been President of the American Auditory Society and the Association for Research in Otolaryngology. He is president-elect of the Politzer Society, an international Otologic Research Organization. He has published over 130 peer-reviewed articles in both clinical and basic science journals and has mentored 18 predoctoral and postdoctoral trainees in basic and translational research, as well as providing clinical training to a large number of otolaryngology residents and fellows. He maintains a busy surgical practice at both UW Medical Center and Seattle Childrens. His NIH funded laboratory studies models of, signal processing in and perception with cochlear implants, and is collaborating in the development of a vestibular implant. He is a member of the Collegium Otorhinolaryngologicum and the American Institute of Medical and Biological Engineering.



GE WANG (Fellow, IEEE) is currently a Clark & Crossan Endowed Chair Professor and the Director of the Biomedical Imaging Center, Rensselaer Polytechnic Institute, Troy, NY, USA. He published the first spiral cone-beam CT paper in 1990s and since then many papers systematically contributed to theory, algorithms, artifact reduction, and biomedical applications in this important area. There are 200 million medical CT scans yearly with a majority in the spiral/helical cone-beam/multi-slice mode. His group and collaborators developed interior tomography to solve the long-standing interior problem for high-fidelity local reconstruction, and proposed omni-tomography (all-in-one and all-at-once) with CT-MRI as an example. He initiated bioluminescence tomography as a new molecular imaging modality. He wrote the first perspective on deep-learning-based tomographic imaging in 2016 as the new frontier of machine learning, two papers on deep imaging in Nature Machine Intelligence, and a textbook on this topic with his coauthors published by IOP, in 2019. He wrote over 500 journal articles and 100 issued and pending patents, receiving a high number of citations and academic recognitions/awards. His results were featured in Nature, Science, PNAS, and news media. His team has been continuously and strongly funded by federal agencies and companies. His main research interests include medical imaging and artificial intelligence. He is an Outstanding Associate Editor of the IEEE TRANSACTIONS ON MEDICAL IMAGING and a Lead Guest Editor of six special issues for this flagship journal. He is a Fellow of SPIE, OSA, AIMBE, AAPM, AAAS, and National Academy of Inventors.

...