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

Mengzhou Li1,*, Zheng Fang1,2,*, Wenxiang Cong1, Chuang Niu1, Weiren Wu1, Josef Uher3, James Bennett4, Jay T. Rubinstein5, and Ge Wang1,*

1 Department of Biomedical Engineering, Rensselaer Polytechnic, Troy, NY, 12180 USA
2 Department of Instrumental and Electrical Engineering, Xiamen University, Xiamen, 361102 China
3 Radalytics a.s., Prague, 17000 CZE
4 Hawkeye Spectral Imaging, Tucson, AZ, 85701 USA
5 Department of Otolaryngology-HNS, Virginia Merrill Bloedel Hearing Research Center, University of Washington, Seattle, WA, 98122 USA

Abstract—While micro-CT systems are instrumental in preclinical research, clinical micro-CT imaging has long been desired with cochlear implantation as a primary example. 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 technologies of micro-focus X-ray source, 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 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 potentials 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 arm, 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 invariable in detecting neoplasms with gadolinium contrast agent, but 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, dehiscent superior semicircular canal, congenital labyrinthine dysplasia, labyrinthine fistula are all disorders or therapies where temporal bone CT is either absolutely necessary or commonly desired adjacent to surgical management, for diagnosis or planning. In cochlear implantation, although psychophysical and physiological measures can be helpful, the ability to localize electrodes and depict their 3D anatomical environment in vivo is particularly instrumental in understanding variations in threshold, uncomfortable level, and channel interaction.

* Contributed equally.
* To whom correspondence should be addressed; E-mail: wangg6@rpi.edu.
duced 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 the phase contrast imaging applications. The key parameters were optimized in the range of 40 keV to 130 keV. This type 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 (CMCT) scanner to integrate a micro-focus tube, a PCD, interior tomography, and deep learning into a novel CMCT device that can be abutted to a clinical CT scanner or used separately. The proposed whole CMCT workflow starts with a 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 upgrade 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 50 µm with a high-power micro-focus source and a state-of-the-art PCD with fine detector elements for electron-noise-free spectral imaging. Furthermore, the robotic-arms adds a great mechanical flexibility compared to the C-arm gantry and enables a moving VOI mask [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]. Especially, [27] 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 dose reduction 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 consist 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.

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 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

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 the power consumption of 250 W, while the working radius of UR16e is 900 mm with the 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 the 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 the 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 the temporal bone imaging requirements [29]. The power supply can be from 100 to 240 VAC at 47 to 440 Hz.

**X-ray Source** The Hamamatsu L9181-02 tube is selected as the x-ray source for CMCT. It is an RS-232 controlled 40-130 kV x-ray tube with a focal spot size adjustable among 8, 20, and 40 µm. Its maximum tube power output is 39 W. It combines an X-ray tube high-voltage power supply and a controller in a compact package powered by a 24 VDC, 5A source. The anode target material is Tungsten. The tube window is made of Beryllium of 0.2 mm thickness. The effective cone beam angle is 45°. Its weight is 10.5 kg, manageable with the larger robot arm.

**Photon-counting Detector** The photon-counting detector ADVA-CAM WIDEPIX-5x5 of 1280 × 1280 pixels fits the need for human temporal bone micro-CT. Its continuous sensitive surface is supported by an array of 5 × 5 detector tiles. Each tile consists of one Timexip 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 Timexip technology also allows for the use of multiple energy thresholds. The intrinsic spatial resolution is defined by the detector pitch of 55 µm. That is, the imaging sensor covers a 7 cm × 7 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 µm, and the diameter of the volume of interest (VOI) is 3.5 cm, assuming a magnification factor 2. This should be sufficient to cover the human inner and middle ear. In addition to the Timexip 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.

**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 µm to 50 µm) 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. All these factors coupled together, on ballpark we should be able to maintain the current head CT dose for an interior micro-CT scan to achieve about 50 µm resolution.

**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'.

\[ r = \sqrt{\left(\frac{x}{M}\right)^2 + \left(\frac{d}{M'}\right)^2} \]  

In our initial CMCT system design, x is no more than 40 µm, and d is equal to 110 µm (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 50 µm.

**Geometric Alignment** In the CMCT process, global and local projection data are obtained in different scanning geometries. An optic 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 LLT2910-100 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 µm at a profiling frequency of 300 Hz. The compact size of 96 mm × 85 mm × 33 mm and low weight of 380 grams are ideal for static, dynamic and robotic applications. The wavelength of the semiconductor laser is switchable between 668 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 VDC 500 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.

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 \( P_{\text{background}} \) and the global sinogram \( P_{\text{global}} \), we can easily obtain the pure sinogram of the region of the interest (ROI) \( P_{\text{ROI}} \) as follows:

\[
P_{\text{ROI}} = P_{\text{global}} - P_{\text{background}}.
\]  

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

\[
\text{trunc}(P_{\text{ROI}}) = \text{trunc}(P_{\text{global}}) - \text{trunc}(P_{\text{background}}),
\]

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

\[
P_{\text{ROI}} = P_{\text{local}} - \text{trunc}(P_{\text{background}}).
\]

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 \( P_{\text{HR,local}} \), and an appropriate background estimation. Suppose that we have a prior LR CT scan of the object \( P_{\text{LR,global}} \), the LR background estimation \( P_{\text{background}} \) 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 with pathological potentials can be further 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 \( P_{\text{global}} \) with a scanning geometry \( G_{\text{global}} \);

2) Determine a VOI and then plan a scanning geometry \( G_{\text{local}} \) for a local micro-CT scan;

3) Scan the patient optically with a surface scanner to generate the surface model \( S_{\text{local}} \) for data/image registration between the global and local scans at the start of the local scan;

4) Perform the micro-CT scan \( P_{\text{local}} \) following \( G_{\text{local}} \) (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 \( P_{\text{local}} \) 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 \( G_{\text{global}} \) and \( G_{\text{local}} \) in reference to the facial surface model \( S_{\text{local}} \).
   a) Reconstruct the global volume \( V_{\text{global}} \) from \( P_{\text{global}} \);
   b) Render the surface model \( S_{\text{global}} \) from the global reconstruction \( V_{\text{global}} \);
   c) Register the two surface models to align the orientation and position of \( S_{\text{global}} \) with \( S_{\text{local}} \). The registration result is used to guide the micro-CT scan;
   d) From \( P_{\text{local}} \), directly reconstruct a volume of interest \( V_{\text{local}} \) 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 \( V_{\text{global}} \) and \( V_{\text{local}} \).

2) Compensation: Correct the attenuation offsets in \( P_{\text{local}} \) to form a pure sinogram of the VOI \( P_{\text{ROI}} \).
   a) Set the attenuation values inside VOI to zero in the aligned global reconstruction to form the background volume;
b) Digitally reproject the background volume \( V_{\text{align}} \) following the geometry \( G_{\text{local}} \) to form the LR background estimation \( P_{\text{align,LR}} \).

c) Interpolate \( P_{\text{background}} \) to the same resolution as the local HR projection of CMCT \( P_{\text{local}} \), and obtain \( P_{\text{HR}} \).

d) Truncate \( P_{\text{HR}} \) to the same size as \( P_{\text{local}} \), and obtain \( \text{trunc}(P_{\text{HR}}) \).

e) Correct \( P_{\text{local}} \) with the estimated attenuation background as \( P_{\text{local}} - \text{trunc}(P_{\text{background}}) \) to form the pure sinogram of the VOI \( P_{\text{VOI}} \).

3) Reconstruction: Reconstruct the VOI from \( P_{\text{VOI}} \) with geometry \( G_{\text{local}} \) 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 from the 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 size of 0.4981 mm. In our study, the phantom image was first interpolated to 10,240 × 10,240 pixels with a pixel size of 0.025 mm to generate the HR phantom. To be mentioned, 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 25 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 140kVp, 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 µm gap is not resolved while the 75 µm gap is well resolved, which indicates the resolution of the local scan protocol is between 50 µm and 75 µm. 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 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. 12. 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 compensated local reconstruction is within ±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.

The initial view angle mismatches were simulated in the range from −4.39° to 4.39° with an increment of 1.998°. The resultant reconstructions are in Fig. 13. 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 directional asymmetry of the background compensation. The maximum relative error is in the range [−10%, 5%] when the angular errors are within ±2.2°. Especially, the relative reconstruction errors are mostly contained within ±6% if the angular errors are within ±1.1°.

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. 14. 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 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 I, II and III 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 I and II. The attenuation deviation from GT in the reconstruction with aligned background compensation is very small, with MSE $0.454 \times 10^{-5}$ 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 1.992 mm after the bias correction, which demonstrate the robustness of our method. Similarly, in Table II the RMSRE remains below 2.0% for all angular errors within $\pm 4.39^\circ$.

The bias correction substantially improves the magnification-error-
Fig. 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.

Fig. 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 correspond to isocenter mismatches from 0 to 2.490 mm in a step of 0.498 mm.

Fig. 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 $\pm 2.2^\circ$ to $\pm 2.2^\circ$ with an increment of $1.1^\circ$.

Fig. 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 $5\%$. 
affected image reconstruction, as shown in Table III. 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 when the compensation is well aligned. The attenuation estimation from the global reconstruction cannot be affected by 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. DISCUSSIONS AND CONCLUSION

In this study, we have proposed a novel CMCT system which incorporates a micro-focus source, 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], the human facial surface measurement techniques become mature 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 a micro-focus has usually an insufficient flux to produce a decent 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 to find the optimal balance between the X-ray intensity and image resolution, a deep-learning based deblurring method can be used to undo the blurring induced by a finite focus size. Although the cone-beam projection with the finite focus spot is no longer a spatially invariant linear system, which is hard to handle with traditional deblurring method, advanced deep learning techniques have the capability of dealing with shift-variant system in 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 build 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. The feasibility of applying the model trained with synthetic data to the real scenarios has been demonstrated in our previous optical deblurring work [44].

B. Image Denoising

Deep denoising techniques can be used to reduce radiation dose and improve image quality. According to the level of supervision in the network training process, current deep learning methods can be grouped into three categories: 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 generates pairs of corresponding noisy-clean images. 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
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 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 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. We also believe advanced deep learning techniques can be leveraged for superior performance, as exemplified by the potential application envisions. Overall, our CMCT system holds a great promise for inner ear imaging as well as other clinical applications.

Acknowledgment

We would like to thank for the financial support from National Institutes of Health under award numbers R01CA237267 (NCI) and R01EB026646 (NIBIB), and from the State Scholarship Fund for the China Scholarship Council (CSC No.201906315030).

Appendix: 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. 12, 13 and 14.

TABLE I

| Position Error | aligned 0.498mm 0.996mm 1.498mm 1.992mm 2.490mm 2.980mm 3.487mm 3.985mm 4.483mm |
|----------------|-------------------------------------------------|
| SSIM           | 0.9998 0.9998 0.9996 0.9993 0.9989 0.9984 0.9977 0.9970 0.9964 0.9952 |
| PSNR           | 46.2150 44.9145 42.3074 39.7698 37.6195 35.8479 34.4174 33.2073 32.1431 31.2084 |
| MSER          | 0.0454 0.0613 0.1117 0.2004 0.3288 0.4944 0.6873 0.8981 1.1603 1.4389 |
| RMSRE         | 0.0107 0.0125 0.0217 0.0231 0.0298 0.0367 0.0434 0.0500 0.0566 0.0631 |
| SSIM (Cr.)     | 0.9998 0.9998 0.9998 0.9996 0.9993 0.9991 0.9986 0.9984 0.9980 0.9980 |
| PSNR (Cr.)     | 46.3106 46.2296 45.1858 43.6330 41.8485 40.1459 39.3240 37.5028 36.3361 35.4465 |
| MSER          | 0.0444 0.0451 0.0576 0.0823 0.1242 0.1838 0.2540 0.3363 0.4336 0.5423 |
| RMSRE (Cr.)   | 0.0164 0.0105 0.0120 0.0146 0.0182 0.0225 0.0266 0.0308 0.0350 0.0393 |

TABLE II

| Angle Error  | −4.30° −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 0.9998 0.9991 0.9989 0.9989 |
| PSNR         | 38.4776 40.0008 42.1055 44.8752 46.2150 44.3655 41.9105 39.9452 38.4152 |
| MSE×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 (Cr.)   | 0.9999 0.9999 0.9999 0.9998 0.9998 0.9998 0.9998 0.9998 0.9998 0.9998 |
| PSNR (Cr.)   | 41.8476 43.3071 44.8716 46.1801 46.3105 46.1733 45.0336 43.3808 41.7435 |
| MSE×10⁻⁴ (Cr.) | 0.1242 0.0887 0.0619 0.0458 0.0444 0.0459 0.0596 0.0873 0.1272 |
| RMSRE (Cr.)  | 0.0179 0.0105 0.0120 0.0146 0.0104 0.0108 0.0121 0.0144 0.0172 |

TABLE III

| Mag. Error | −20% −15% −10% −5% −5% aligned 5% 10% 15% 20% |
|-----------|---------------------------------|
| SSIM      | 0.9828 0.9907 0.9961 0.9991 0.9999 0.9999 0.9999 0.9993 0.9866 0.9785 |
| PSNR      | 26.3718 29.2189 33.1242 39.3940 49.0959 46.5190 31.3849 28.5055 26.5721 |
| MSE×10⁻⁴  | 4.3832 2.2750 0.9257 0.2185 0.0234 0.4236 1.3816 2.6812 4.1848 |
| RMSRE     | 0.0143 0.0751 0.0479 0.0232 0.0076 0.0323 0.0584 0.0813 0.1016 |
| SSIM (Cr.) | 0.9996 0.9996 0.9999 0.9999 0.9999 0.9999 0.9999 0.9999 0.9999 0.9999 |
| PSNR (Cr.) | 41.9195 44.8545 47.0394 48.2090 48.0992 46.9567 45.1736 43.6877 42.4808 |
| MSE×10⁻⁴ (Cr.) | 0.1222 0.0621 0.0376 0.0287 0.0294 0.0383 0.0577 0.0813 0.1074 |
| RMSRE (Cr.) | 0.0167 0.0122 0.0097 0.0084 0.0084 0.0095 0.0117 0.0138 0.0159 |

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. We also believe advanced deep learning techniques can be leveraged for superior performance, as exemplified by the potential application envisions. Overall, our CMCT system holds a great promise for inner ear imaging as well as other clinical applications.

Acknowledgment

We would like to thank for the financial support from National Institutes of Health under award numbers R01CA237267 (NCI) and R01EB026646 (NIBIB), and from the State Scholarship Fund for the China Scholarship Council (CSC No.201906315030).

Appendix: 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. 12, 13 and 14.
Fig. 12. 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 4.483 mm with an increment of 0.498 mm.

Fig. 13. ROI Reconstructions with misalignment of initial view angle in background compensation displayed in the same window (unit: $cm^{-1}$). The misalignment errors are from $-4.39^\circ$ to $4.39^\circ$ with an increment of $1.098^\circ$.

Fig. 14. 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 $5\%$. 
L. Chen, Y. Shen, C.-J. Lai, T. Han, Y. Zhong, S. Ge, X. Liu, T. Wang, D. Kolditz, T. Struffert, Y. Kyriakou, A. Bozzato, A. Dörfler, and G. Zan, D. J. Vine, R. I. Spink, W. Yun, Q. Wang, and G. Wang, “Design approach based on a truncated hilbert transform,” Optical Engineering, vol. 56, no. 4, pp. 043101, 2017.

S. A. Hendriksen, D. M. Pelt, and K. J. Batenburg, “Noise2inverse: Self-supervised deep convolutional denoising for tomography,” IEEE Transactions on Computational Imaging, 2020.

M. Feuer, 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,” Measurement science and technology, vol. 19, no. 9, p. 094001, 2008.

T. Bakirmam, M. U. Gunusay, 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,” Applied optics, vol. 56, no. 4, pp. 985–992, 2017.

N. M. Bell, T. N. Miller, P. J. Abbas, B. J. Gantz, S. W. Lee, and J. T. Rubinstein, “Design, analysis and simulation for optimization of a periodic microstructured array anode for hard x-ray ct system,” Physics in Medicine & Biology, vol. 54, no. 7, pp. 1601–1622, 2019.

M. Chen, Y. Xi, W. Cong, B. Liu, B. Wei, and G. Wang, “X-ray ct geometrical calibration via locally linear embedding,” Journal of X-ray Science and Technology, vol. 24, no. 2, pp. 241–256, 2016.

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,” Optical Engineering, vol. 57, no. 2, p. 021302, 2018.

M. J. Ackerman, “The visible human project,” Proceedings of the IEEE, vol. 91, no. 3, pp. 504–511, 1998.

O. M. Parkhi, A. Vedaldi, and A. Zisserman, “Deep face recognition,” 2015.

L. Ma, T. Xu, and J. Lin, “Validation of a three-dimensional facial scanning system based on structured light techniques,” Computer methods and programs in biomedicine, vol. 94, no. 3, pp. 290–298, 2009.

A. Braham, R. Ramam, S. Heist, C. Rulf, P. Kühlstedt, and G. Notni, “Fast 3d mir systems for facial measurement and lip-reading,” in Dimensional Optical Metrology and Inspection for Practical Applications VI, vol. 10220. International Society for Optics and Photonics, 2017, p. 102200P.

G. P. Khoyro, C. A. Beaumont, A. Borgna, N. Rodriguez-Flores, R. W. Breakey, W. Rodgers, F. Anguilla, N. O. Jeelani, S. Schievena, and D. J. Dunaway, “Comparison of three-dimensional scanner systems for craniomaxillofacial imaging,” Journal of Plastic, Reconstructive & Aesthetic Surgery, vol. 70, no. 4, pp. 441–449, 2017.

A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J.-C. Fillion-Robin, S. Pajot, C. Bauer, D. Jennings, F. Fennessy, M. Sonka et al., “3d slicer as an image computing platform for the quantitative imaging network,” Magnetic resonance imaging, vol. 30, no. 9, pp. 1323–1341, 2012.
[43] H. Wang, Y. Rivenson, Y. Jin, Z. Wei, R. Gao, H. G"unaydin, L. A. Bentolila, C. Kural, and A. Ozcan, “Deep learning enables cross-modality super-resolution in fluorescence microscopy,” Nature methods, vol. 16, no. 1, pp. 103–110, 2019.

[44] M. Li, H. Shan, S. Pryshchep, M. M. Lopez, and G. Wang, “Deep adversarial network for super stimulated emission depletion imaging,” Journal of Nanophotonics, vol. 14, no. 1, p. 016009, 2020.

[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 Transactions on Image Processing, vol. 26, no. 7, pp. 3142–3155, 2017.

[46] X. Mao, C. Shen, and Y.-B. Yang, “Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections,” in Advances in neural information processing systems, 2016, pp. 2802–2810.

[47] Y. Tai, J. Yang, X. Liu, and C. Xu, “Memnet: A persistent memory network for image restoration,” in Proceedings of the IEEE international conference on computer vision, 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 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 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,” arXiv preprint arXiv:1803.04189, 2018.

[50] D. Ulyanov, A. Vedaldi, and V. Lempitsky, “Deep image prior,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 9446–9454.

[51] A. Krull, T.-O. Buchholz, and F. Jug, “Noise2void-learning denoising from single noisy images,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 2129–2137.

[52] A. Krull, T. Vicar, and F. Jug, “Probabilistic noise2void: Unsupervised content-aware denoising,” arXiv preprint arXiv:1906.06631, 2019.

[53] S. Laine, T. Karras, J. Lehtinen, and T. Aila, “High-quality self-supervised deep image denoising,” in Advances in Neural Information Processing Systems, 2019, pp. 6970–6980.

[54] M. Li, D. S. Rundle, and G. Wang, “X-ray photon-counting data correction through deep learning,” arXiv preprint arXiv:2007.03119, 2020.