Accepted for publication in *Medical Physics*, 2016

An iterative projection-based motion estimation and compensation scheme for head X-ray CT

Tao Sun* and Johan Nuyts

KU Leuven-University of Leuven, Department of Imaging and Pathology, Nuclear Medicine and Molecular Imaging, Medical Imaging Research Center (MIRC), B-3000 Leuven, Belgium

Jung-Ha Kim

Discipline of Medical Radiation Sciences, University of Sydney, NSW 2006, Australia

Roger Fulton

Discipline of Medical Radiation Sciences, University of Sydney, NSW 2006, Australia and Department of Medical Physics, Westmead Hospital, Westmead, NSW 2145, Australia

(Dated: September 12, 2016)

Abstract

Purpose: Although current CT systems can scan the head in a very short time, patient motion sometimes still induces artifacts. If motion occurs, one has to repeat the scan; to avoid motion, sedation or anesthesia is sometimes applied.

Methods: We propose a method to iteratively estimate and compensate this motion during the reconstruction. In every iteration, the rigid motion was estimated view-by-view and then used to update the system matrix. A multi-resolution scheme was used to speed up the convergence of this joint estimation of the image and the motion of the subject. A final iterative reconstruction was performed with the last motion estimate.

Results: The method was evaluated on simulations, patient scans and a phantom study. The quality of the reconstructed images was improved substantially after the compensation. In simulation and phantom studies, root-mean-square error (RMSE) was reduced and mean structural similarity (MSSIM) was increased. In the patient studies, most of motion blurring in the reconstructed images disappeared after the compensation.

Conclusions: The proposed method effectively eliminated motion-induced artifacts in head CT scans. Since only measured raw data is needed for the motion estimation and compensation, the proposed method can be applied retrospectively to clinical helical CT scans affected by motion.

* Email: tao.sun@uzleuven.be

I. INTRODUCTION

A slight movement of the patient can lead to a reduction of spatial resolution in Computed Tomography (CT), in severe cases resulting in corrupted images unsuitable for diagnosis or further processing. To reduce the likelihood of motion artifacts, CT manufacturers have made scans faster by increasing the number of detector rows and the rate of rotation of the x-ray source and detector. Other ways to reduce the patient motion include general anesthesia, sedation [1] and the use of restraining devices for head and neck imaging [2].

In practice it is difficult to completely eliminate motion, and compensating motion artifacts is of considerable general interest in tomography. A variety of methods for assessing motion in CT exist, including directly estimating motion using a camera system with visual markers [3–6] or without markers [7]. Artificial or anatomical landmarks can be also tracked in the image or projection domains [8, 9]. Indirect estimation methods have been proposed where motion is estimated through the minimization of errors in consistency conditions [10–13], or iteratively updating the motion together with the reconstruction process [14–17]. Another approach has used similarity measures to quantify changes between successive projections to measure subject motion [18]. Once motion parameters have been estimated, a compensation for the motion can be applied, either to the measured raw data or during the reconstruction process.

Among these methods, some addressed the problem in 2D parallel-beam or fan-beam geometries [10–13]. Other retrospective motion estimation and compensation methods addressed the problem for 3D circular cone beam CT (CBCT) [8, 9, 14, 15]. The methods in [14, 15] minimize an image-based cost function (which essentially detects motion artifacts) to estimate the motion. Motion estimation and compensation is arguably simpler in CBCT since the entire object is normally in the field of view at all times. In contrast, in helical CT, the object is always truncated in the axial direction, limiting the amount of information that can be used to verify consistency of the projections. Relatively few studies have been done for clinical helical CT, and some of those require additional measurement to acquire the motion [3–6].

In this study, we propose an approach to reduce or eliminate motion artifacts in helical CT reconstruction. The proposed motion estimation and compensation method only needs the measured raw data. The method assumes that for each view, the pose of the measured object may be different. Consequently, for every view, a rigid transformation representing the object pose is estimated. An initial compensation for changes in pose (motion) during projection acquisition is applied during reconstruction by incorporating the motion estimates into the system matrix [5]. Then the motion and the reconstructed image can be updated alternately in an iterative scheme until an optimal motion estimate is found. The proposed approach has been validated on simulations and a phantom study by comparing reconstructed images with and without motion compensation. Results on patient scans are also presented.

II. MATERIALS AND METHODS

A. Coordinate system

A clinical helical CT system usually has a cylindrical detector surface, with a radius equal to the detector source distance. We define the world coordinate system $c = (x, y, z) \in \Re^3$ in Fig.1. It is fixed with respect to the scanner, and its z-axis coincides with the rotation axis

of the scanner. The detector coordinate system $c' = (u, v, z) \in \Re^3$ is fixed with respect to the rotating source-detector system: its origin moves along z-axis while the system moves, u is tangent and v is orthogonal to the detector. For one projection view, we define the rigid motion transform in the coordinate system c:

$$S_{world} = (\varphi_x, \varphi_y, \varphi_z, t_x, t_y, t_z)^{\mathrm{T}} \tag{1}$$

where φ_x, φ_y, φ_z are 3 rotations, t_x, t_y, t_z are 3 translations. The motion can be mapped in a detector coordinate system c':

$$S_{detector} = (\varphi_u, \varphi_v, \varphi_z, t_u, t_v, t_z{}')^{\mathrm{T}} \tag{2}$$

where φ_u, φ_v, φ_z are rotations, t_u, t_v, t'_z are translations. A small motion in the direction perpendicular to the detector t_v, results in a very small magnification of the projection, which is assumed negligible [19]. In every projection view, then, we set t_v to zero and only 5 parameters need to be estimated in our scheme in the detector coordinate system c':

$$S_{detector} = (\varphi_u, \varphi_v, \varphi_z, t_u, t_z{}')^{\mathrm{T}} \tag{3}$$

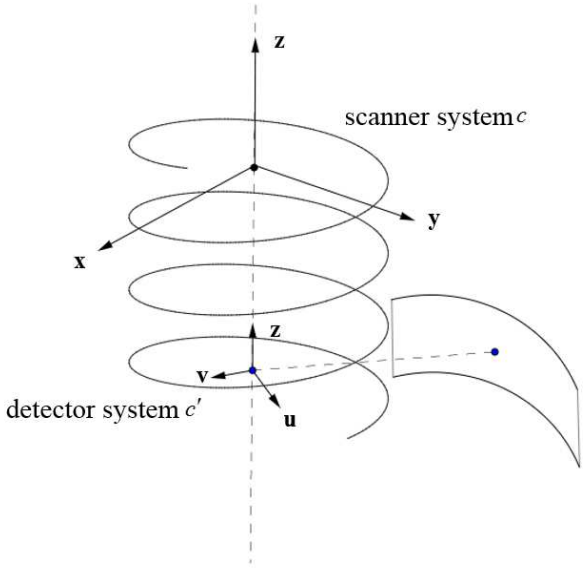

FIG. 1. The scanner and detector system on which motion estimation and compensation is based. The offset along the rotation axis between origins of the two systems is t_{offset}.

B. OSEM reconstruction

In the presence of object motion, the helical CT-orbit is distorted into an effective orbit with arbitrary shape [5]. Because this is problematic for analytical reconstruction, an iterative reconstruction algorithm is needed. We used Ordered Subset Expectation Maximization (OSEM) as the reconstruction algorithm [20]:

$$\mu_j^{n+1} = \frac{\mu_j^n}{\sum\limits_{i \in S_b} a_{ij}} \sum_{i \in S_b} a_{ij} \frac{f_i}{\sum\limits_k a_{ik} \mu_k^n} \tag{4}$$

where f is the log converted sinogram, S_b is one subset (consisting of b views). We used the OSEM algorithm for convenience, but if the use of a better noise model would be required, it can be replaced with a dedicated iterative algorithm for transmission tomography [21].

C. General motion estimation and compensation scheme

The aim is to estimate the pose of the object for each of the acquired CT views. This is achieved by a 3D registration of the object to each of the 2D views independently. The first estimate of the 3D object is obtained with an initial reconstruction without motion compensation. As a result, that first image suffers from motion artifacts, which will adversely affect the accuracy of pose estimates associated with each view. Nevertheless, we find that the 2D-3D registration process described below still captures part of the true motion, such that reconstruction with motion compensation based on these (poor) motion estimates improves the reconstruction. Reiterating the process with this improved reconstruction in turn produces more accurate motion estimates. This leads to an iterative algorithm which alternately estimates the motion for each view and the motion compensated image. This algorithm is explained in more detail in the following paragraphs.

Although the initial reconstructed image is motion-contaminated, it can be used to generate a first rough motion estimate. This motion is taken into account in a reconstruction process to generate a motion-corrected image at the first iteration. Then the motion-corrected image and the motion estimate are alternately updated to increase the likelihood, the iterations are stopped when the updated motion seems to have converged (Fig. 2). The algorithm consists of two parts: (1) the joint image and motion estimation (JIM) and (2) the final reconstruction (motion compensation). Each JIM-iteration consists of 2 steps: a motion update and an image update. The image update is done by applying multiple iterations of the OSEM algorithm.

The implementation involves four steps: (1) a motion update — a 2D-3D image registration to update the pose estimate for each view in the current JIM-iteration; (2) an image update, computed with an iterative reconstruction algorithm incorporating the updated motion estimate in its system matrix at the current JIM-iteration; (3) alternate updates of both image and motion within a multi-resolution scheme; (4) final reconstruction with a system matrix based on the last motion estimate. Details on each part of the framework are described below.

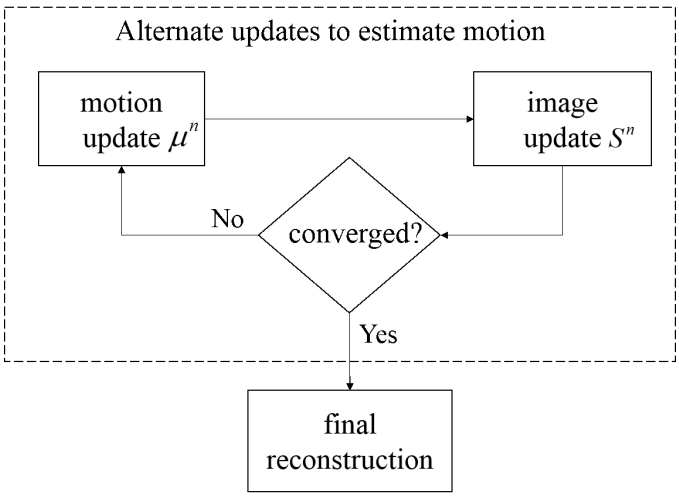

FIG. 2. General motion estimation and compensation scheme. μ is the update of the attenuation image, S is the update of the rigid motion.

1. Motion update

For one projection line i, we integrate along the projection line to define the forward projection of the estimated image μ at current JIM-iteration:

$$f_i = \sum_j a_{ij}\mu_j \tag{5}$$

where i is the projection line index, j is the voxel index, a_{ij} is the effective intersection length of the line i with voxel j. In helical CT, the line integrals are organized in views, where view θ contains all line integrals associated with a single source position:

$$f_\theta = \{f_i\} \tag{6}$$

Suppose the general motion estimation and compensation scheme (Fig. 2) is at the JIM-iteration n, hence the current motion estimate is s^n. For view θ, the current pose estimate is s_θ^n and the 5 motion parameters in Eq. (3) are estimated one after the other. Let r be one of these parameters (a rotation or translation) to be estimated. Assuming that the change in the pose parameter represented by $\hat{r}$ is small, the derivative of projection f with respect to r can be approximated as a finite difference of the intensities:

$$\frac{\partial f_\theta}{\partial r} \approx \frac{f_{\theta,m} - f_\theta(s_\theta^n)}{\hat{r}} \tag{7}$$

where $f_\theta(s_\theta^n)$ is the calculated re-projection (using the current estimates of the image and motion), $f_{\theta,m}$ is the measured projection for view θ. For view θ, $\hat{r}$ minimizes the difference between $f_\theta(s_\theta^n + \hat{r})$ and $f_{\theta,m}$. To estimate $\hat{r}$ in Eq. (7), we need to know the derivative on the left hand side. Therefore, we introduce another equation which is very similar to Eq. (7):

$$\frac{\partial f_\theta}{\partial r} \approx \frac{f_\theta(s_\theta^n + \Delta r) - f_\theta(s_\theta^n)}{\Delta r} \tag{8}$$

where Δr is a known small increment of the parameter to be estimated. When Δr represents a translation, $f_\theta(s_\theta^n + \Delta r)$ can be approximated as a simple translation of current re-projection $f_\theta(s_\theta^n)$; for in-plane rotation, again $f_\theta(s_\theta^n + \Delta r)$ can be approximated as a simple rotation of the re-projection $f_\theta(s_\theta^n)$, as shown in Fig. 3. For the two out-of-plane rotations, we calculated $f_\theta(s_\theta^n + \Delta r)$ with a forward projection using a system matrix adjusted with Δr.

Eq. (7) and Eq. (8) assume that a small increment of one degree-of-freedom rigid motion only results in a linear change of the intensities in the projection. All the above lead to a least squares minimization problem for view θ at the current JIM-iteration n:

$$\hat{r} = \arg\min_r \left\| \Delta r \left[f_{\theta,m} - f_\theta(s_\theta^n) \right] - r \left[f_\theta(s_\theta^n + \Delta r) - f_\theta(s_\theta^n) \right] \right\|^2 \tag{9}$$

To find $\hat{r}$, Eq. (9) was solved analytically. Defining

$$\begin{aligned} P_\theta &= f_{\theta,m} - f_\theta(s_\theta^n) \\ Q_\theta &= f_\theta(s_\theta^n + \Delta r) - f_\theta(s_\theta^n) \end{aligned} \tag{10}$$

and setting the derivative of the right hand side in Eq. (9) with respect to r to zero, one obtains:

$$\hat{r} = \frac{\sum\limits_N P_\theta \cdot Q_\theta}{\left\| \sum\limits_N {Q_\theta}^2 \right\|} \Delta r \tag{11}$$

where N is total number of voxels in projection view θ.

The above procedure showed how to estimate one parameter in one projection view. For view θ, this procedure was applied to estimate all five parameters in Eq. (3). The sequence of the estimation was translation first, then rotation. The newly estimated parameter values were used immediately when estimating the value of next parameter. This sequential estimation of five motion parameters for all projection views completes the update of the rigid motion at the current JIM-iteration (Fig. 2).

Now the estimated motion is obtained in the detector coordinate system c' (Fig.1). It is transformed to the motion in the world coordinate system c, as motion-corrected reconstruction requires the motion in the world coordinate system:

$$\left\{ S_\theta^{n,inc} \mapsto T_\theta^{n,inc} \right\}_{\theta=0,\ldots,M} \tag{12}$$

where T is the 4×4 homogeneous matrix representation of the estimated motion in the world coordinate system. More details about Eq. (12) are given in the Appendix.

The transformation matrix obtained in the nth JIM-iteration was then used to update the previous motion estimate for every view, which was used in the next JIM-iteration $(n+1)$:

$$\left\{T_\theta^{n+1} = T_\theta^n T_\theta^{n,inc}\right\}_{\theta=0,\ldots,M} \tag{13}$$

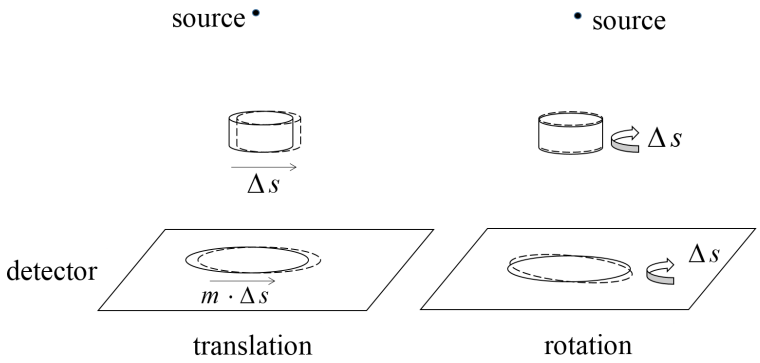

FIG. 3. In the detector coordinate system, the effect of object translation or rotation parallel to the detector can be well approximated as translation and rotation of the projection. For simplicity, the curvature of the detector is ignored. In the left half figure, m is the magnification factor from the object to detector.

2. Image update

After obtaining the motion, the image representing the attenuation coefficients can be updated with iterative reconstruction. We used OSEM as the reconstruction algorithm.

Instead of moving the reconstructed image in every view, rigid motion compensation is done by considering a coordinate system fixed to the object and incorporating the motion (now associated to the source-detector pair) into the system matrix. This corresponds to an arbitrary 3D motion of a virtual gantry around the object being scanned, created by the superposition of the inverse of the object motion on the helical trajectory [5]. Motion compensation is enabled by introducing a modified version of standard OSEM:

$$\begin{aligned} \hat{T}_i^{n+1} &= invert(T_i^{n+1}) \\ \mu_j^{n+1} &= \frac{\mu_j^n}{\sum\limits_{i \in S_b} \hat{T}_i^{n+1}(a_{ij})} \sum_{i \in S_b} \hat{T}_i^{n+1}(a_{ij}) \frac{f_i}{\sum\limits_k \hat{T}_i^{n+1}(a_{ik}) \mu_k^n} \end{aligned} \tag{14}$$

where $\hat{T}_i$ is a 4×4 transformation matrix applied to the projection line i. If T_i is the identity matrix for all projection lines, then Eq. (14) is the same as standard OSEM (Eq. (4)). In helical CT, T_i is constant for all projection lines in one projection view, hence the inversion

is done for every single view. Because of the high rotation speed and the large number of views, the motion within a single view is negligible.

Distance-driven projection is used for interpolation during the (back) projection [22]. The new estimate of the attenuation image is then used for the next motion update (step 1).

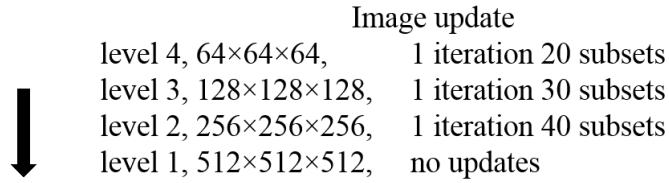

FIG. 4. The number of OSEM-iterations and subsets applied for the image update at each resolution level. Note that we stop the estimation at the second last level, hence no image and motion updates were computed at level 1.

3. Multi-resolution alternate updates

By repeating steps 1 and 2 we can update the motion (Eq. (13)) and reconstruction (Eq. (14)) alternately. Because the image and the motion parameters are jointly estimated from the measured data, the problem of error propagation is minimized. An approach to reduce computation time is to apply a multi-resolution technique. We utilized this by running the algorithm from a coarse to fine representation of the image. For example, the starting image resolution level is $8 \times 8 \times 8$, i.e. a down-sampling factor of eight was applied in all directions. There is a resampling with a factor of two between adjacent levels. Image updates were reconstructed at coarse resolution at early JIM-iterations, while the resolution increased as the iteration numbers increased. The number of OSEM-iterations applied for the image update was the same within one resolution level. These numbers were optimized based on simulations which had a similar configuration as the patient study (Fig. 4). A possible additional advantage of the multi-resolution technique is that it may help avoiding convergence to an undesired local maximum.

It was not obvious how to define good stopping criteria when estimating the motion at each level, especially considering that the ground truth image was missing for clinical studies. In our implementation, the summation of projection errors between the re-projected and measured data over all views was computed, and at each resolution level, the iterations were stopped when the relative change of this error measure did not exceed 0.2 %. In our experiments, we observed that the motion estimate hardly changed during the computations at the finest resolution. Since these computations (if included) are the most expensive ones in the multi-resolution scheme, we stopped the scheme at the second finest resolution.

The motion updates were smoothed (by filtering each degree-of-freedom independently along the projection views) to remove outliers. We chose the Savitzky-Golay filter [23] to do the smoothing. The optimal size of the smoothing kernel depends on both the view sampling rate of the measured data and the axial detector extent. The clinical data usually have a high view sampling rate, while we used a lower rate in simulations. The axial extent varies with different scan configuration (slice collimation). The optimal smoothing kernels are determined by simulations with several common collimation configurations, as shown

in Table I. If the number of views per rotation is increased, the number of kernel points is increased accordingly.

TABLE I. The motion smoothing kernel width for common slice collimation

	Case 1	Case 2	Case 3	Case 4
Angles used during motion estimation	150	150	150	150
Collimation (mm)	96×0.6	64×0.6 (32×1.2)	32×0.6	16×0.6
Smoothing kernel (points)	17	23	75	105

4. Final reconstruction

When the motion estimate has converged, a final reconstructed image with diagnostic quality must be produced for a clinical scan (Fig. 2). One way to speed up the final reconstruction is choosing an initial image which is close to the maximum likelihood solution. Faster convergence is achieved if the iterative reconstruction could be started from a sharper image. In simulations we started the final reconstruction with the last image update from the alternate updates. Since the alternate multi-resolution scheme was terminated at a coarser grid, the initial image must be created by interpolating to the finer grid, and as a result, the initial image is relatively smooth.

For that purpose, we implemented an approximate helical Feldkamp-Davis-Kress (FDK) algorithm. We utilized all the data in each projection view. A first order motion compensation is obtained by taking the motion for each view into account in the backprojection step. This approximation creates artifacts, but these have typically low spatial frequencies. Since low frequencies tend to converge relatively fast in iterative algorithms such as OSEM, only few updates are needed to eliminate them. For the clinical studies, this FDK-initialisation was used since it was found to be more efficient to reach convergence.

To further accelerate the final reconstruction, Nesterov's momentum approach [24] was applied (using all previous iterates to compute the momentum) in the final reconstruction. All forward and backward projection operations were implemented in OpenCL and run on a GPU (NIVIDIA Tesla C2075).

D. Design of the experiments

1. Simulations

In simulations, measured motion segments from volunteers were applied to a phantom to generate simulated CT scans subject to patient motion. Details about measuring these motions are given in [6]. The phantom was a 3D voxelized phantom from the Visible Human Project [25]. The image intensities were converted from Hounsfield (HU) to attenuation coefficients (cm^{-1}) at an effective energy of 70 keV. Image size was $256 \times 256 \times 240$; pixel size was $1 \times 1 \times 1$ mm^3.

All helical scans were simulated as being scanned with a Siemens Definition AS CT scanner (Siemens Medical Solutions USA, Inc., Malvern, PA), with reduced angular sampling

to reduce computation times. The scan parameters were: angles per rotation 150, pitch 1.0, collimation 32×1.2 mm. Six measured motion segments (referred as studies 1 to 6) were applied to the phantom for the simulated helical scans. Examples of volunteer motion segments are shown in Fig. 5. To avoid artifacts, all simulated helical scans covered a bit more than the entire object. Reconstructed images from these scans were analyzed quantitatively to assess the performance of the proposed method.

Alternate updates of both image and motion were performed within the multi-resolution scheme to obtain the final motion estimate. OSEM was used for all reconstructions, with motion compensation enabled (Eq. (14)). During the JIM, the attenuation image was updated using the OSEM iteration schemes shown in Fig. 4. For the final reconstruction, the pixel size was $1 \times 1 \times 1$ mm^3, and 4 iterations with 60 subsets were applied.

Poisson noise was added to the raw simulated data before the reconstruction (assuming 1000 photons were detected on each detector element in the blank scan). Again OSEM was used for this reconstruction, but as mentioned in II. B, it can be replaced by a dedicated iterative algorithm for transmission tomography.

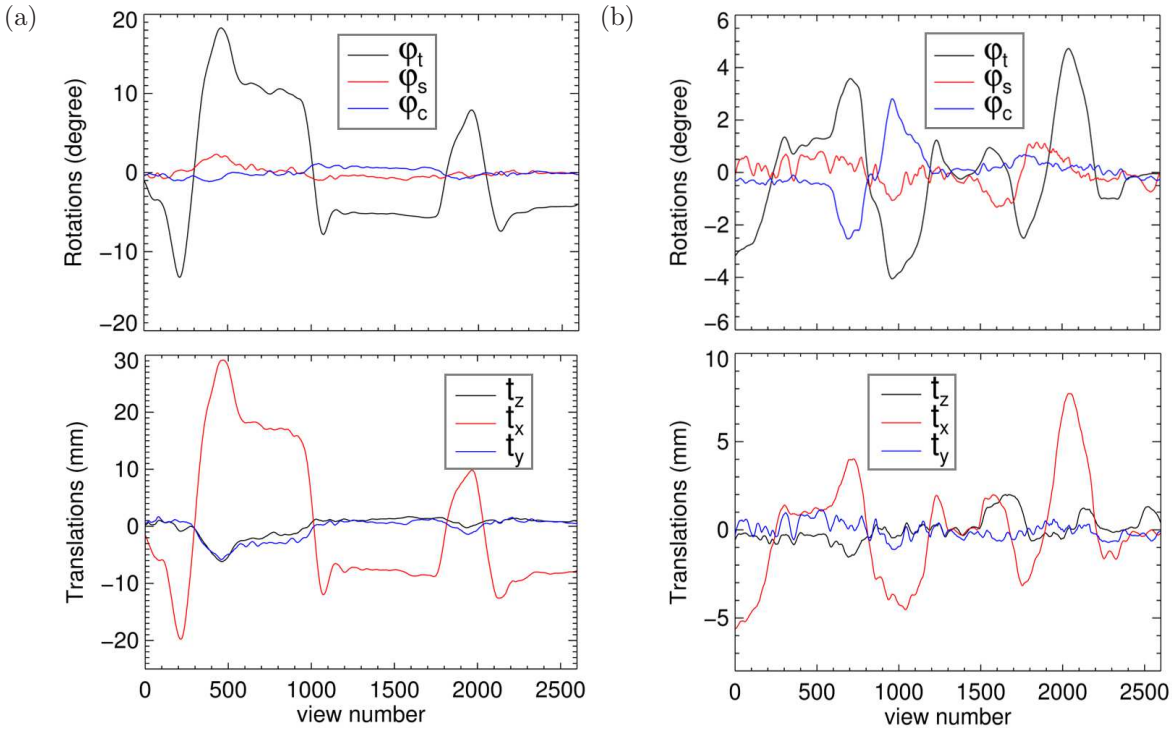

FIG. 5. Examples of the simulated volunteer head motion. (a) Moderate motion from study 1. (b) Slight motion from study 6. Details of motion tracking are given in [6].

2. Patient scans

The method has been applied to clinical studies in which motion artifacts had been observed. The anonymized raw data of four patients who had previously undergone head CT scans in the Department of Radiology at Westmead Hospital, Sydney, Australia, were collected with the approval of the Human Research Ethics Committee of the Western Sydney

Local Health District. The scans were performed on a Siemens Force scanner (Siemens Medical Solutions USA, Inc., Malvern, PA). The scan parameters are listed in Table II.

Because the head support and the bed do not move with the patient during the scan and would compromise the estimation of the patient motion, they were removed from the raw data prior to further processing. The head support and (or) bed were firstly segmented from an initial low-resolution reconstructed image. The segmented portion of image was then forward-projected to generate a new set of projections, which were subtracted from the measured projections before the motion estimation and compensation scheme was executed.

Because of the large size of the raw data, motion was estimated at every 8^{th} view to accelerate both motion and image updates. This resulted in approximately 500 views per rotation. With a rotation time of 1 s, this yields a temporal sampling of 500 Hz, which was considered sufficient for motion estimation. The multi-resolution JIM scheme was applied as in II. C. 3. For all patient studies, the motion smoothing kernel sizes were selected based on Table I.

For the final OSEM reconstruction, the starting image was computed with helical FDK reconstruction with motion compensation enabled. Six iterations with 30 subsets were applied in combination with Nesterov's acceleration. Other reconstruction parameters are listed in Table II. Also, the entropy of the reconstructed images was computed for reconstructions with and without motion compensation.

TABLE II. Scan and reconstruction parameters

	Patient 1	Patient 2	Patient 3	Patient 4
Tube voltage (kVp)	120	120	120	120
Tube current (mA)	120	154	150	150
Rotation time (s)	1.0	1.0	1.0	1.0
Pitch	0.55	0.55	0.55	0.55
Angles per rotation	4200	4200	4200	4200
Collimation (mm)	64×0.6	96×0.6	32×0.6	96×0.6
Flying focus	phi, z	phi, z	phi, z	phi, z
Pixel size (mm$\times$ mm$\times$ mm)	$0.4 \times 0.4 \times 0.75$	$0.45 \times 0.45 \times 0.5$	$0.455 \times 0.455 \times 0.5$	$0.451 \times 0.451 \times 0.5$
Dimension	$512 \times 512 \times 219$	$512 \times 512 \times 550$	$512 \times 512 \times 376$	$512 \times 512 \times 404$

3. Phantom scan

In our previous work [5], we developed and optimized a rigid motion compensation technique for helical CT brain scanning, in which the motion information was obtained using an optical motion tracking system. We observed that after motion compensation based on the tracking data, some residual artifacts were still present. We attributed these artifacts to the finite accuracy of the motion tracker. In this experiment, we verified whether the proposed method can be used to refine the motion estimate from the optical system.

The scan was performed in the Department of Nuclear Medicine and Ultrasound at Westmead Hospital, on a Siemens Sensation 16 scanner (Siemens Medical Solutions USA, Inc., Malvern, PA). The scan parameters were: pitch 1.0, tube voltage 120kVp, tube current 280 mA and tube collimation 16×0.6 mm. Flying focus was turned off.

An optical motion tracking system (Polaris Spectra, Northern Digital Inc., Waterloo, Canada) was placed at the rear of the scanner. A 3D Hoffman brain phantom, which contained air inside was used in this experiment. The phantom was placed off-center on the curved bed and held in place with a wedge (Fig. 6a). During the scan, the wedge was removed by pulling a string from outside the room. The phantom then started rolling left and right on the bed to finally come to rest at a stable position at the center of the bed. This motion was too severe to be compensated only by the proposed method, since the initial reconstruction was corrupted severely. The tracked rigid motion is shown in Fig. 6b, relative to its pose at the start of the scan.

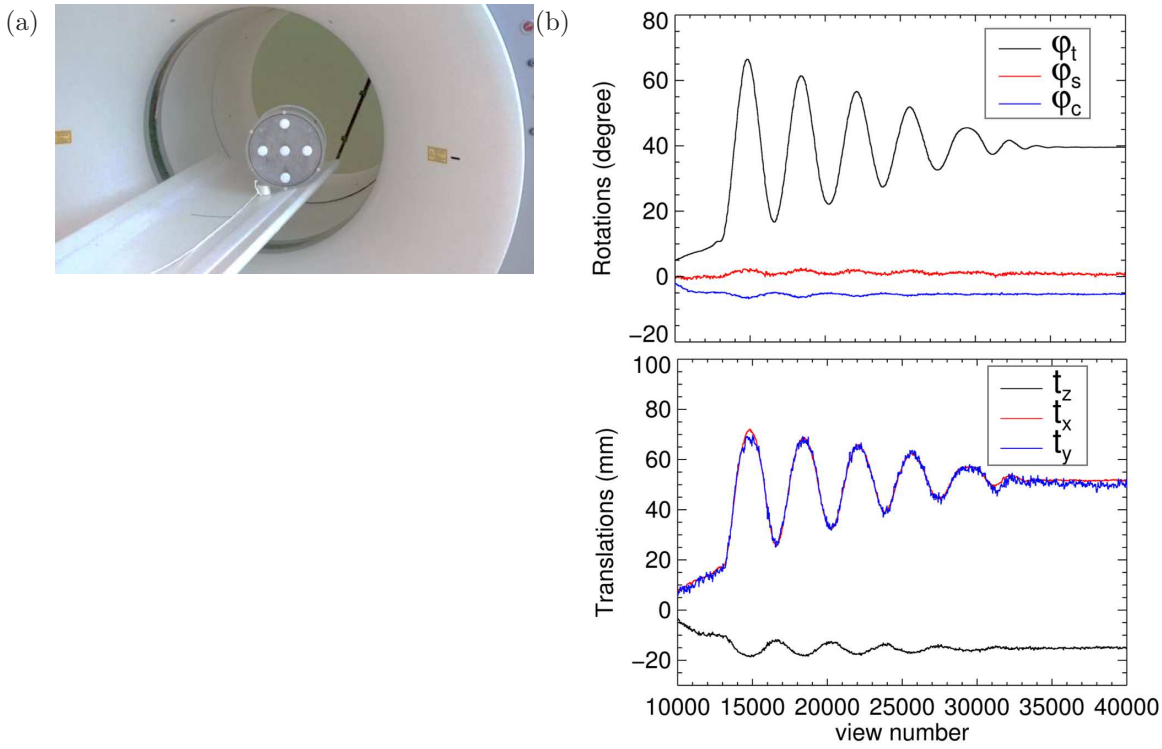

FIG. 6. (a) Setup of the Hoffman phantom used in the experiment, and (b) measured motion.

E. Evaluation of the results

In the simulation studies, the effects of motion compensation were evaluated by visual assessment and with quantitative analysis. The reconstructed images and the ground truth images were compared in all planes with similarity metrics. We chose root-mean-square error (RMSE) and mean structural similarity (MSSIM) [26] as the metrics. Note that before the calculation of these metrics, both motion-corrected and uncorrected images were registered to the true image plane by plane (Fig. 7). This was done because motion and motion-compensation may introduce positional differences which are irrelevant for image quality and therefore should not affect the evaluation of similarity.

In the patient studies, image entropy, with and without motion compensation applied, was also computed (plane-by-plane), based on the assumption that motion-induced artifacts would tend to increase the image entropy.

In the phantom study, the quality of the reconstructed images was assessed for reconstructions with and without motion compensation by comparison to reconstructions from motion-free CT-scans.

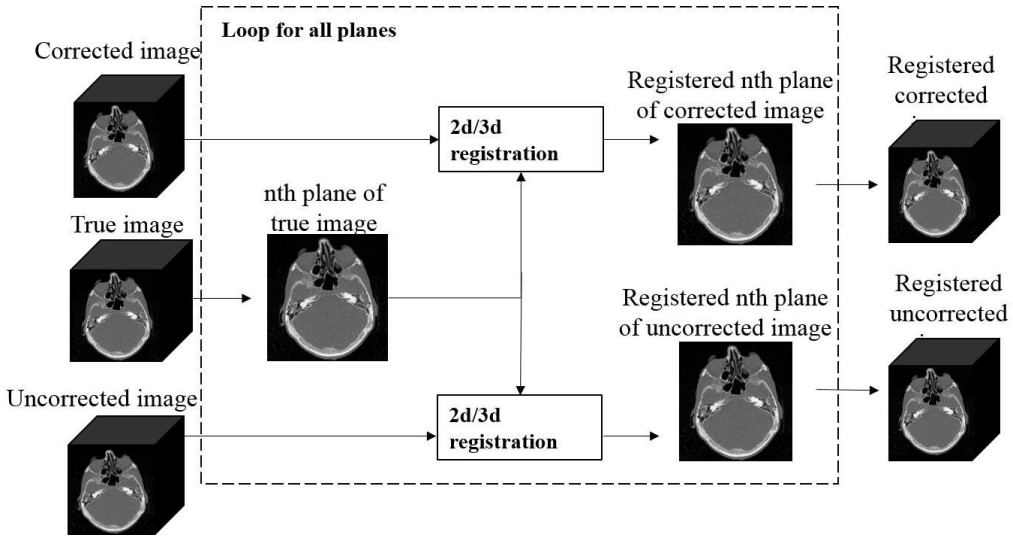

FIG. 7. Pre-processing of the motion-corrected and uncorrected images before quantitative analysis in simulation studies.

III. RESULTS

A. Simulations

Fig. 8 shows the corrected image of a selected simulation using a moderate motion segment. Most of the distortions are eliminated. Fig. 9 shows the quantitative analysis of a selected simulation using a slight motion segment. The true image is much more similar to the corrected image than the uncorrected one. Fig. 10 shows the overall improvement across all image planes in all 6 studies with different motions. Fig. 11 shows the result of the simulation study 1 with relatively high noise. The motion estimation and compensation is still effective on data with high noise level.

B. Patient scans

Fig. 12 and Fig. 13 show the non-motion corrected image (reconstructed with the scanner system software) and motion-corrected reconstructed images from patient 1 and patient 2, respectively. Fig. 14 compares the image of a repeat scan (which was done because of the observed motion in the first scan) with the reconstructions, with and without motion compensation for patient 3. Fig. 15 shows the change of the total image entropy, as an indicator of artifact reduction.

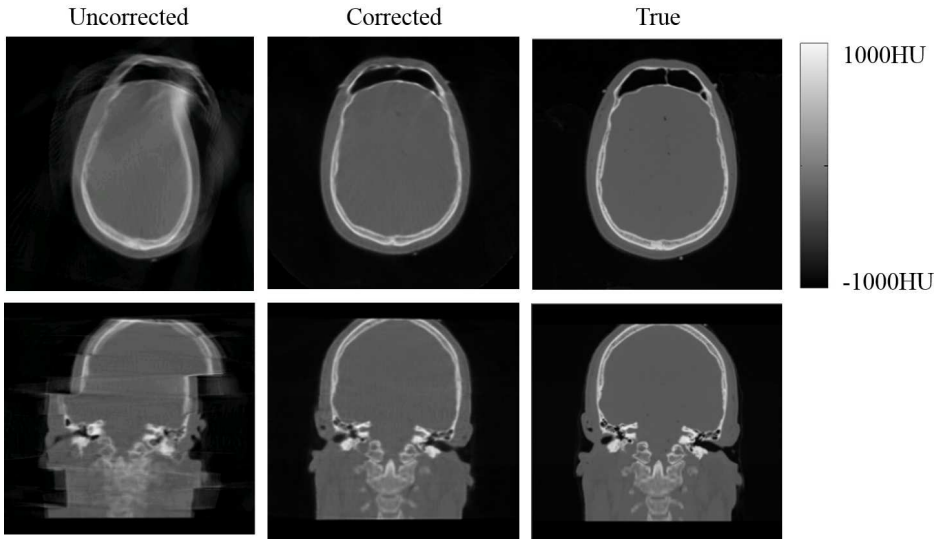

FIG. 8. Results from the simulation study 1 using moderate motion (Fig. 5a). Selected transaxial (top) and coronal (bottom) slices from reconstructions without and with motion compensation, and also from the true image.

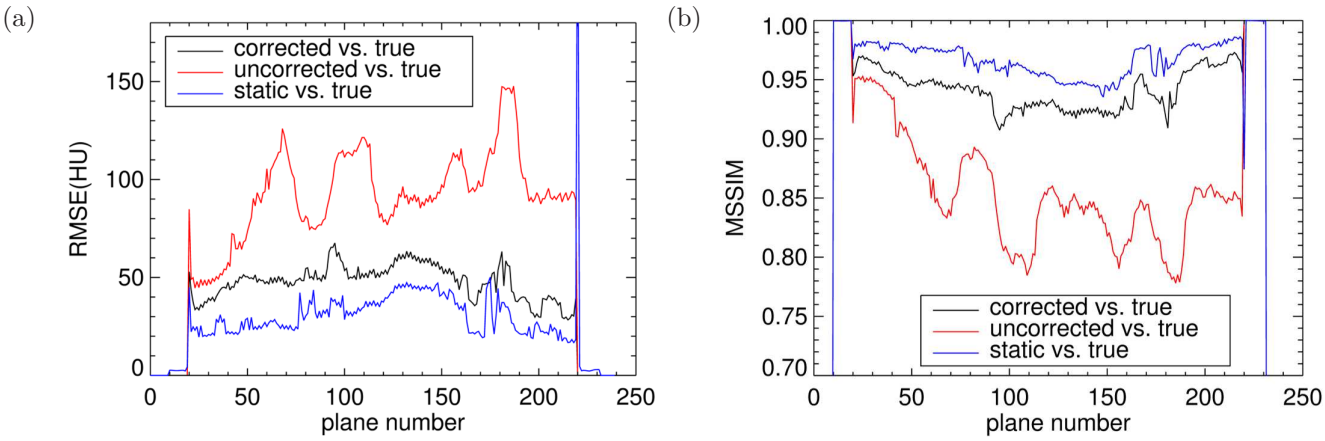

FIG. 9. Results from the simulation study 6 using slight motion (Fig. 5b). Similarity metrics with and without motion compensation over reconstructed planes. Top: RMSE. Bottom: MSSIM. The difference between the static reconstruction and true image was displayed as the reference.

C. Phantom scan

As shown in the top row of Fig. 16, some small irregularities were visible at the edges of the phantom in the reconstructed image after a first compensation using the tracked motion. A possible reason is the finite accuracy of the pose measurements. For this scan, we applied the proposed method to compensate these residual "jagged" artifacts due to the imperfect motion recording. The motion estimation process was identical to the one applied in the patient studies, except that the measured motion was used as the initial motion estimate. The proposed method removed the artefacts (Fig. 16, middle). Quantitative analysis was done similarly to what has been done in simulation studies. The true image was obtained from a static scan of the same phantom. Fig. 10 (box 7) shows the overall improvement of

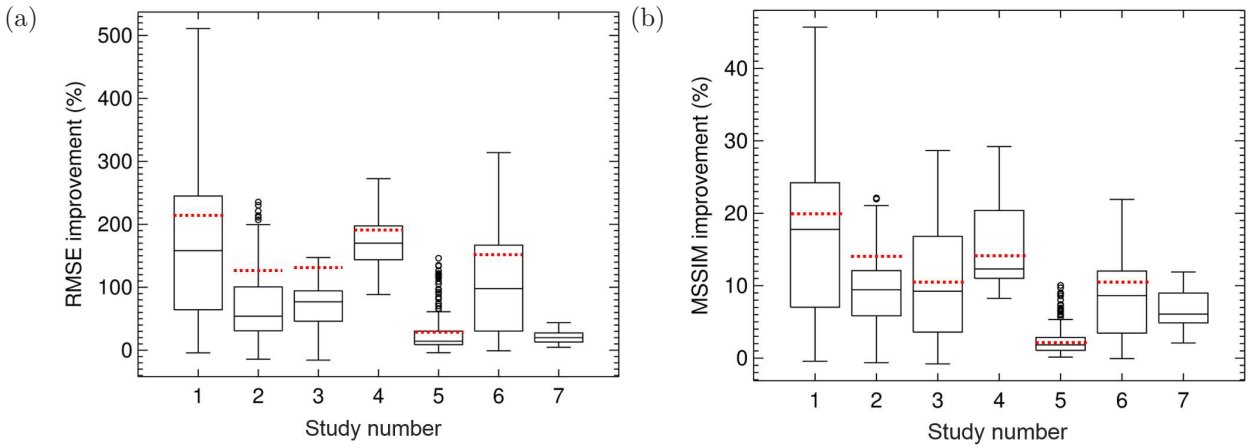

FIG. 10. Box plots of improvement over reconstructed planes of each individual study — simulation studies (1-6), phantom scan (7). Left: RMSE improvement. Right: MSSIM improvement. The upper and the lower limit of the bar are the maximum and minimum. The upper and the lower limit of the box are the first and third quartiles. The central line is the median. The circles are the outliers. As a reference for each study, the red dashed lines represent the medians of the improvement of the static reconstructed image over the uncorrected image.

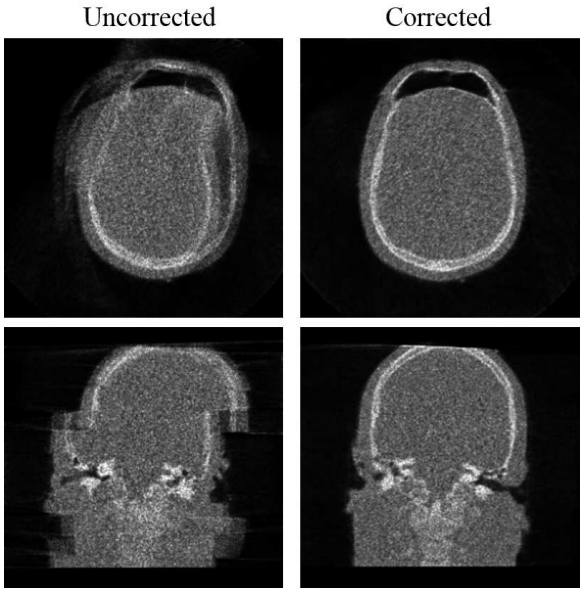

FIG. 11. Noisy simulation of study 1. Selected transaxial (top) and coronal (bottom) slices show the improvement on image quality.

the RMSE and MSSIM across image planes.

IV. DISCUSSION

In this paper, the simulation studies were performed with a (simulated) 64-row CT scanner, while the patient studies were performed with a 96-row CT, and the phantom study was performed with a 16-row CT. In other tests (not shown here), the method performed

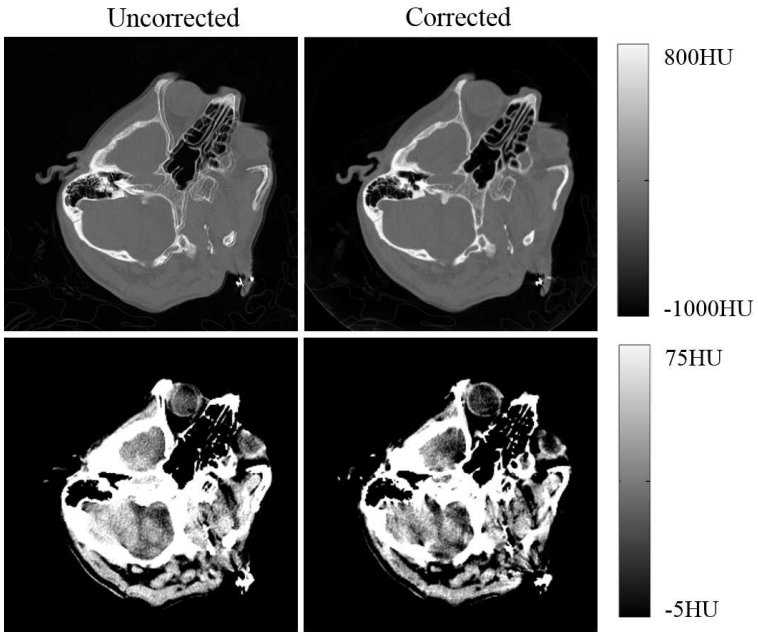

FIG. 12. Top: Selected transaxial plane, without (left) and with (right) compensation for motion artifacts in a scan from patient 1. The uncorrected image was from a standard vendor reconstruction. Bottom: the same plane that is shown in a narrow window (Gaussian smoothed with Full width at half maximum (FWHM) = 2 mm).

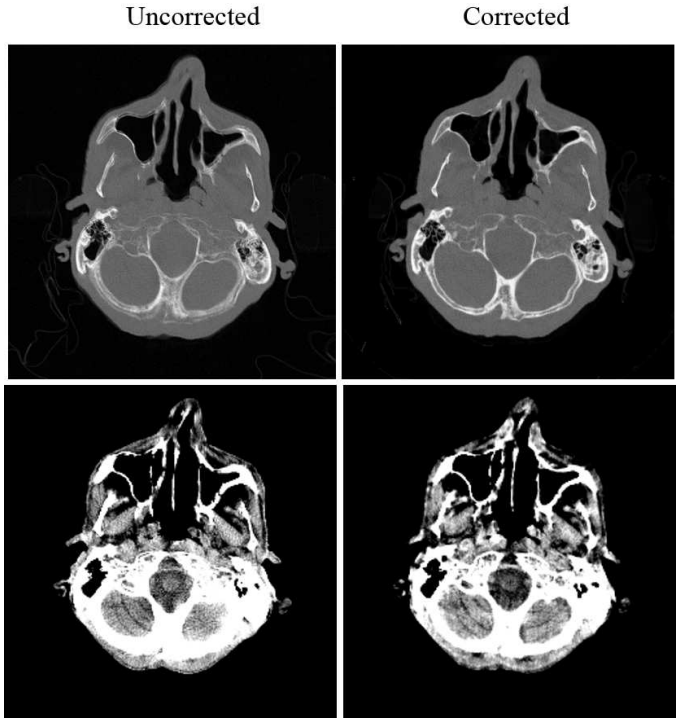

FIG. 13. Top: Selected transaxial plane, without (left) and with (right) compensation for motion artifacts in a scan from patient 2. Bottom: the same plane that is shown in the same narrow window used in Fig. 12 (Gaussian smoothed with FWHM = 2 mm).

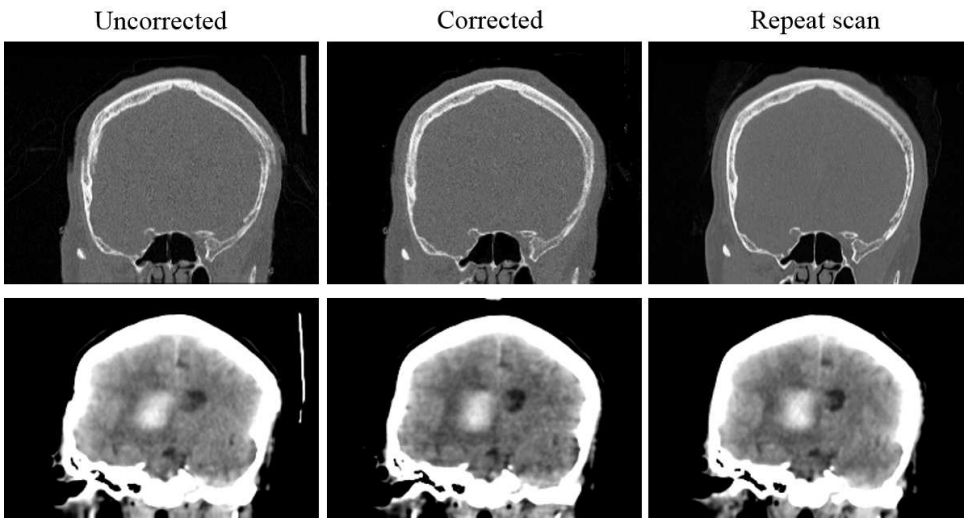

FIG. 14. Top: Selected coronal plane without (left) and with (middle) compensation in a scan (effective mAs 272) from patient 3. As a reference, the repeated scan (effective mAs 327, registered to the first scan) is displayed on the right. Bottom: another plane contains a lesion that is shown in the same narrow window used in Fig. 12 (Gaussian smoothed with FWHM = 2.5 mm).

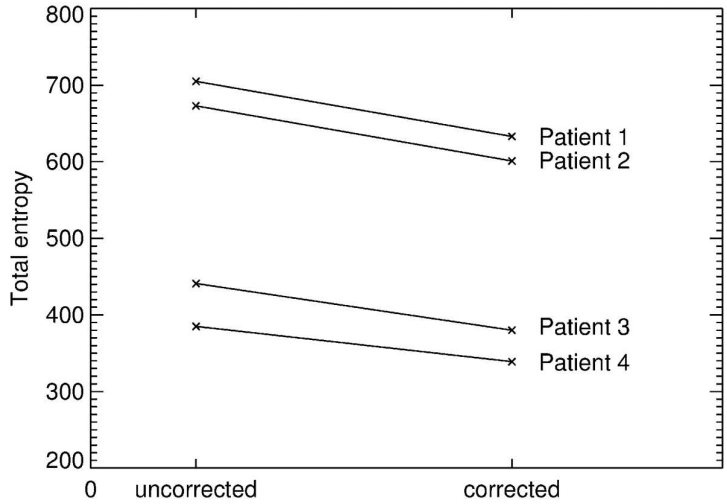

FIG. 15. The total entropy change for 4 patient scans. The entropy in each individual plane (not shown here) decreased for all the studies.

well for data from scanners with different row numbers as well, but the performance of the proposed method was found to be better in the case of a higher number of detector rows, as expected since a wider detector provides more information in a single projection view. For the data from a scanner with a narrower detector, a stronger smoothing was needed to suppress the noise on the estimated motions (Table I).

The large number of CT views in clinical scans creates an impressive computational challenge. Currently, the time for estimating the motion is about 2 hours for a single patient scan (from Siemens Force scanner), while the time for the final iterative reconstruction is about 12 hours (a high number of updates is required to obtain a quality comparable to the

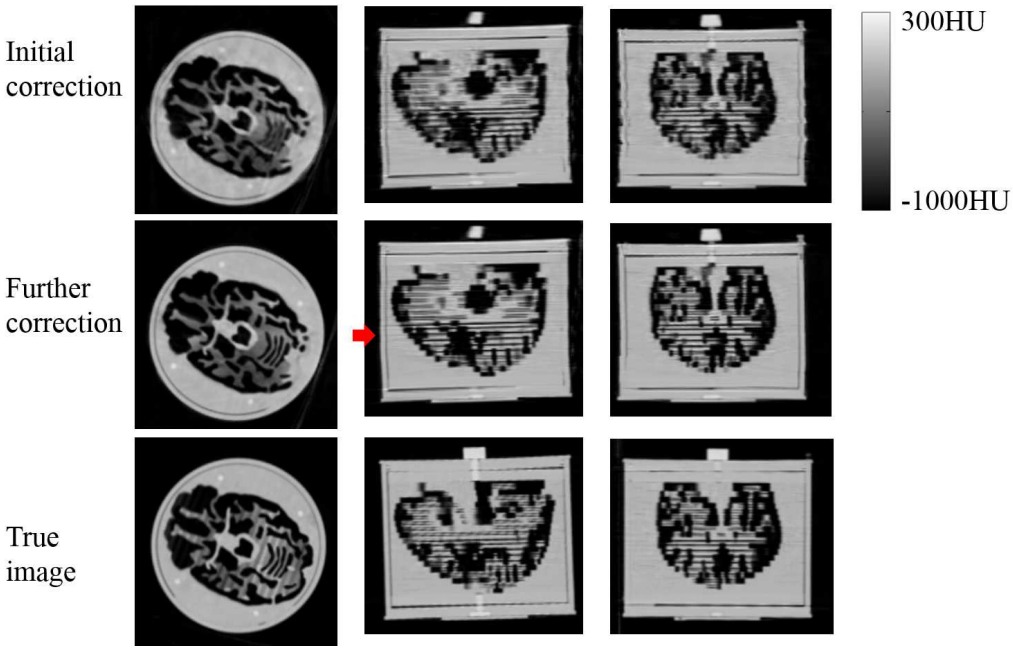

FIG. 16. Selected transaxial, sagittal and coronal planes, with and without compensation for residual motion. Top: reconstructed image with motion compensation based on optical tracking data; Middle: reconstructed image with further compensation by the proposed approach; Bottom: reference image reconstructed from a static scan.

vendor images). A possible way to accelerate the final reconstruction is to replace the iterative reconstruction algorithm with a dedicated motion-sensitive analytical reconstruction algorithm. This is under development by the authors.

As illustrated in Fig 8, the proposed method performed well even for relatively large patient motion. In our simulations, the proposed method usually worked well when the amplitude of the rotations was less than 10° and amplitude of the translations was less than 20 mm, which in our opinion are unlikely to be exceeded in most clinical scans. Nevertheless, we observed that it did not perform well in cases of severe motion such as that of Fig. 6b. We have shown previously that optical motion tracking methods are effective, even for very severe motion [5, 6]. As shown in III. C, the proposed method can refine the tracker based motion estimation. Consequently, combining both methods would relax the specifications for the tracking device significantly and at the same time provide accurate motion compensation even in the presence of very severe motion.

The multi-resolution technique accelerated the motion estimation algorithm. Using too coarse a starting resolution, however, should be avoided because the excessive blurring may suppress important high-frequency features. We observed that this can lead to an overestimation of the motion.

We ignored estimating the translation perpendicular to the detector in every projection view. We repeated the simulation studies in III. A, where that translation was estimated and compensated too. The results (not shown) indicated that compensating for that translation had no or a negligible effect on the quality of the corrected image. Considering the additional time to estimate that particular motion, we did not take it into account in our studies.

The estimated motion is not always identical to the true motion. First, the pose of the reconstructed object is arbitrary, and probably roughly corresponds to the average pose

during the scan. Moreover, slow components of the motions may not be fully estimated, but instead partly be incorporated as a gradual and almost rigid distortion of the image along the z-axis. Such a small distortion can be observed in the phantom image of Fig. 16 (middle). These small distortions are not expected to have an adverse effect on the diagnostic value of the image.

The proposed method relies on 2D-3D image registration, and is therefore expected to be less effective when the contrast in the object is low. In CT brain imaging, the high contrast between the skull and soft tissue was found to provide adequate information for estimating the motion. For other possible applications where the contrast would be lower, pre-processing to enhance the contrast of the raw projection data might be necessary.

We only considered the application on diagnostic CT where the noise in the raw data is typically very low. We performed a noisy simulation, in which the proposed method worked fine even with relatively high Poisson noise. We think one reason is that most of the artifacts were already eliminated during the early resolution levels, which had smooth image updates as the pixel size was large. Still, it would be interesting to investigate how the proposed method works on the data from low dose CT scans.

The motion was estimated using an analytical expression based on a linearization. Instead, a more accurate non-linear least squares algorithm could be used. However, that approach would have a much larger computational cost. The linearization approximation becomes better closer to convergence, and the experimental results indicate that even for large motions it is good enough to improve the estimate in every JIM-iteration.

In our current approach, an independent rigid motion was estimated for every view. The method could be further improved by modelling the motion as a parameterized function of time (or view number) [13, 16]. This would reduce the number of unknowns and impose a physically meaningful smoothness to the estimated motion.

V. CONCLUSION

In this paper, we proposed a motion estimation and compensation approach for helical X-ray CT of the head, for which the only required input is the measured raw data. Since no additional measurements are needed, it can be applied retrospectively to standard helical CT data. We believe that, when sufficiently accelerated, it can become a valuable clinical tool, since it would reduce the need for anesthesia or sedation in children and other patients who are likely to move, and decrease the number of repeat scans. Further testing of the method with more clinical data is ongoing.

ACKNOWLEDGEMENT

This work was supported by the IWT MIRIAD SBO project, the KU Leuven IMIR PF-project and National Research Council Australia Project Grant 632677. The authors want to thank Krystal Moore from the Department of Radiology at Westmead Hospital for collecting the datasets. We also want to thank Karl Stierstorfer, Siemens Healthcare, Forcheim, Germany, for help with reading the raw CT data.

DISCLOSURE OF CONFLICTS OF INTEREST

The authors have no relevant conflicts of interest.

APPENDIX A

This section explains how the rigid motion parameters in the detector coordinate system are transformed to a homogenous matrix in the world coordinate system. From Eq. (3), we have 5 degrees-of-freedom for each projection view θ:

$$S_{detector} = (\varphi_u, \varphi_v, \varphi_z, t_u, t_z)^{\mathrm{T}}$$

Transform $S_{detector}$ into homogeneous matrix:

$$T_{detector} = \begin{bmatrix} \cos\varphi_z & -\sin\varphi_z & 0 & 0 \\ \sin\varphi_z & \cos\varphi_z & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\varphi_u & -\sin\varphi_u & 0 \\ 0 & \sin\varphi_u & \cos\varphi_u & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$
$$\begin{bmatrix} \cos\varphi_v & 0 & \sin\varphi_v & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\varphi_v & 0 & \cos\varphi_v & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & t_u \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Now the motion is in detector coordinate system, we still need to map $T_{detector}$ into world coordinate system:

$$T_{world} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 & 0 \\ \sin\theta & \cos\theta & 0 & 0 \\ 0 & 0 & 1 & t_{offset} \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot T_{detector}$$

where t_{offset} is the offset between the world and detector system in the direction of bed movement.

REFERENCES

[1] R E Wachtel, F Dexter and A J Dow, "Growth rates in pediatric diagnostic imaging and sedation," Anesth. Analg. **108**, 161621 (2009).

[2] J F Barrett and N Keat, "Artifacts in CT: recognition and avoidance," Radiographics **24**, 167991 (2004).
[3] U K Bhowmik and R R Adhami, "A head motion measurement system suitable for 3D cone-beam tomography using markers," IEEE Eng. Med. Biol. Soc. Conf. Proc. 59758 (2012).
[4] J-H Kim, J Nuyts, Z Kuncic and R Fulton, "The feasibility of head motion tracking in helical CT: a step toward motion correction," Med. Phys. **40**, 41903 (2013).
[5] J-H Kim, J Nuyts, A Kyme, Z Kuncic and R Fulton, "A rigid motion correction method for helical computed tomography (CT)," Phys. Med. Biol. **60**, 204773 (2015).
[6] J-H Kim, T Sun, A R Alcheikh, Z Kuncic, J Nuyts and R Fulton, "Correction for human head motion in helical x-ray CT," Phys. Med. Biol. **61**, 141638 (2016).
[7] P Noonan and J Howard, "Realtime markerless rigid body head motion tracking using the Microsoft Kinect," IEEE Nucl. Sci. Symp. Med. Imaging Conf. Rec. 22416 (2012).
[8] C J Ritchie, C R Crawford, J D Godwin, K F King and Y Kim, "Correction of Computed Tomography Motion Artifacts Using Pixel-Specific Back-Projection," IEEE Trans. Med. Imaging **15**, 33342 (1996).
[9] W Lu and T R Mackie, "Tomographic motion detection and correction directly in sinogram space," Phys. Med. Biol. **47**, 126784 (2002).
[10] S Leng, B Nett, M Speidel and G-H Chen, "Motion artifact reduction in fan-beam and cone-beam computed tomography via the fan-beam data consistency condition (FDCC)," Proc. SPIE 6510, Medical Imaging 2007: Physics of Medical Imaging **6510**, 13 (2007).
[11] H Yu, Y Wei, J Hsieh and G Wang, "Data Consistency Based Translational Motion Artifact Reduction in Fan-Beam CT," IEEE Trans. Med. Imaging **25**, 792803 (2006).
[12] H Yu and G Wang, "Motion Artifact Reduction in Fan-Beam CT," IEEE Trans. Med. Imaging **26**, 24960 (2007).
[13] R Clackdoyle and L Desbat, "Data consistency conditions for truncated fanbeam and parallel projections," Med. Phys. **42**, 83145 (2015).
[14] Y Kyriakou, R M Lapp, L Hillebrand, D Ertel and W Kalender, "Simultaneous misalignment correction for approximate circular cone-beam computed tomography," Phys. Med. Biol. **53**, 626789 (2008).
[15] A Kingston, A Sakellariou, T Varslot, G Myers and A Sheppard, "Reliable automatic alignment of tomographic projection data by passive auto-focus," Med. Phys. **38**, 493445 (2011).
[16] Z Yu, J-B Thibault, J Wang, C A Bouman and K D Sauer, "Kinetic parameter reconstruction for motion compensation in transmission tomography," Proc. SPIE 7873, Computational Imaging IX **7873**, 7 (2011).
[17] C Rohkohl, H Bruder, K Stierstorfer and T Flohr, "Improving best-phase image quality in cardiac CT by motion correction with MAM optimization," Med. Phys. **40**, 031901 (2013).
[18] S Ens, J Mller, B Kratz and T M Buzug, "Sinogram-Based Motion Detection in Transmission Computed Tomography," 4th European Conference of the International Federation for Medical and Biological Engineering: ECIFMBE, 5058 (2009).
[19] G T Gullberg, B M W Tsui, C R Craford, B J Glen and J T Hagius, "Estimation of Geometrical Parameters for Cone Beam Tomography," Med. Phys. **17**, 26472 (1989).
[20] H M Hudson and R S Larkin, "Ordered Subsets of Projection Data," IEEE Trans. Med. Imaging **13**, 6019 (1994).
[21] J Nuyts, B De Man, P Dupont, M Defrise, P Suetens and L Mortelmans, "Iterative reconstruction for helical CT: a simulation study," Phys. Med. Biol. **43**, 72937 (1998).

[22] B De Man and S Basu, "Distance-driven projection and backprojection in three dimensions," Phys. Med. Biol. **49**, 246375 (2004).
[23] A Savitzky and M J E Golay, "Smoothing and Differentiation of Data by Simplified Least Squares Procedures," Anal. Chem. **36**, 162739 (1964).
[24] Y Nesterov, "Smooth minimization of non-smooth functions," Math. Program. **103**, 12752 (2005).
[25] M J Ackerman, "The visible human project," Proc. IEEE **86**, 50411 (1998).
[26] Z Wang, A C Bovik, H R Sheikh and E P Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Trans. Image Process. **13**, 60012 (2004).