Towards General Motion Recovery in Cone-Beam Computed Tomography

Wolfgang Wein and Alexander Ladikos

***Abstract*—Irreproducible motion of either the patient or the device during a cone-beam X-Ray scan remains a major issue limiting reconstruction quality in many practical applications. Computational approaches are starting to emerge, which allow to model general motion parameters during the reconstruction itself. Besides, intelligent image processing on the projection data may reveal clues about "what went wrong" during a scan. We present a novel algorithm which uses a combined analysis in projection and reconstruction space, to both detect and account for unknown motion. This allows not only for the detection of large-scale, non-periodic bulk motion, but also an automatic recovery of it, required for a reconstruction void of artifacts. Using the proposed method, we can restore the reconstruction of clinical head scans with severe unknown motion. Moreover, we evaluate our method on synthetic data with known motion trajectories in a radiotherapy scenario.**

***Index Terms*—cone-beam, computed tomography, motion.**

I. Introduction

The cone-beam reconstruction algorithms used for most X-Ray Computed Tomography (CT) imaging devices today, strongly rely on the assumption that the geometry of the X-Ray source-detector arrangement relative to the imaged subject is correctly known and modeled at all times. However this assumption is often violated in light of non-reproducible device inaccuracies (which cannot be modeled during offline geometric calibration), as well as patient motion. The latter is more often an issue in slowly rotating C-arm systems as opposed to gantry-based CT scanners. Loss of resolution, severe artifacts, and completely wrong structures in the reconstruction appear, when the individual rays are considered with the wrong geometry. Ultimately it does not even matter if the excess motion stems from the device or the patient.

Prior work dealing with patient motion has generally focused on particular clinical applications or anatomic regions. This allows to constrain the problem in terms of motion characteristics, such as periodicity and typical trajectories occurring with cardiac or respiratory motion. If a surrogate signal for the motion phase is available, a binned 4D-reconstruction (with corresponding loss of signal to noise ratio proportional to the number of volumes), followed by further refinement [1] is possible. Other practical solutions require a reference scan not affected by motion, e.g. using a breath-hold acquisition protocol [2]. A more holistic approach is to jointly reconstruct the target volume and a motion model [3]. This however requires a reasonably close initialization of the motion parameters, hence unknown large-scale motion cannot be addressed. General mathematical formulations of such a joint reconstruction have been described e.g. in [4] and [5], with only early results on abstract geometric data though.

We had previously proposed a practical self-calibration approach for refining uncertain geometric and radiometric parameters within an algebraic reconstruction framework [6]. Other related methods also repeatedly reconstruct with altered parameters, using a quality criterion such as the sharpness of the reconstruction volume or a single slice thereof [7], in analogy to a camera's auto-focus. Such methods are able to tune parameters, however again require close initialization (if the initial reconstruction is deteriorated, a local optimization would not converge).

To address unknown large-scale motion during a scan, we have recently developed a method to analyze successive X-Ray images in projection space using the epipolar geometry [8]. It is able to detect and approximately recover the three-dimensional motion between X-Ray images with a small baseline (such as successive projections from a cone-beam scan). However simply concatenating those incremental motion estimates would yield drift of the entire sequence.

In this work, we extend the motion detection method [8] by appropriate normalization of the transformation estimates, such that they actually can be used to improve the reconstruction. We then combine the results with the self-calibration approach [6], in order to fine-tune the motion estimates. This yields a powerful hybrid technique which can fully recover even large-scale bulk motion. The remainder of this manuscript is organized as follows. First, we review the basics of self-calibration and motion detection in projection space. Then the new transformation normalization scheme allowing to incorporate the results into the reconstruction is presented. The combined method which yields the final refinement is then explained. Results on both real clinical and synthetic data sets are shown, followed by a brief discussion.

II. Methods

A. Self-Calibration

Algebraic reconstruction techniques generally minimize the re-projection error between a volume estimate and the measured X-Ray projection data:

$$E = \arg\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{p}\| \tag{1}$$

where $\mathbf{A}$ is the system matrix, $\mathbf{x}$ the vector with all entries of the volume estimate, and $\mathbf{p}$ the X-Ray attenuation data. In a perfect scenario void of geometric errors, measurement noise, and with all details of the X-Ray physics modeled

ImFusion GmbH, Agnes-Pockels-Bogen 1, 80992 München, Germany. Corresponding author: Wolfgang Wein, E-mail: wein@imfusion.de.

in the forward- and back-projection steps, the residual re-projection error E would converge towards zero after sufficient iterations. It is therefore appropriate to minimize the same error E with respect to further unknown parameters (in addition to the individual attenuation values of the reconstruction volume). "Global" parameters which directly affect the entire reconstruction result, are optimized by re-computing a reconstruction, using equation 1 as cost function value to minimize. "Local" parameters such as additional drift or rotation of individual X-Ray projections, are sequentially optimized, after which the reconstruction is re-computed. The latter essentially comprises a 2D-3D registration algorithm of X-Ray projections to the reconstruction volume. A state of the art GPU implementation of an ordered subset simultaneous iterative reconstruction technique (OS-SIRT) is used, such that executing a non-linear optimization over the reconstruction as cost function is not computationally prohibitive [6].
For the "global" self-calibration, the number of parameters one can use is limited, both for increase in computation time and numerical stability (in terms of ambiguities and local optima). For the local sequential optimization of individual projection parameters, a close starting estimate is required, since otherwise the initial reconstruction would be deteriorated to start with. Therefore, these computational tools are in the current form rather suited for self-calibration of device parameters than for recovering large-scale patient motion.

B. Motion Detection

Using geometric calculations from stereo computer vision, it is possible to express the attenuation value $I(\mathbf{x})$ of an X-Ray image I at pixel location $\mathbf{x}$ as a linear combination of pixels along the epipolar line in the next X-Ray image J for a supposed geometric relationship between the images:

$$I(\mathbf{x}) = \sum_{k=1}^{n} \mathbf{w}_k J(\mathbf{x}_k) + \mathbf{w}_{n+1} \tag{2}$$

Here, $\mathbf{x}_k$ are a number of discrete sample locations along the epipolar line in image J, defined from preferred depth locations where the image structures are most prevalent. The unknown weights w_k are estimated locally within each pixel's neighborhood in a least-squares fashion. It is then possible to derive an image similarity measure for every image location from that:

$$S_x(I,J) = 1 - \frac{|\mathbf{i}_x - \mathbf{J}_x \mathbf{w}_x|^2}{Var(\mathbf{i}_x)} \tag{3}$$

where $\mathbf{i}_x$ is now a vector of pixel values within a neighborhood of I, $\mathbf{J}_x$ is a matrix with the number of pixels considered times the number of samples along the epipolar line, and $\mathbf{w}_x$ is the vector of weights computed around image location $\mathbf{x}$. Computed over entire pairs of successive X-Ray images, S constitutes a similarity measure which is sensitive to the supposed 3D-geometric relationship between the images. Varying the transformation between the images and hence modifying the epipolar line segments along which the image relationship is assessed, allows to detect if additional motion is present. In addition to the binary decision whether motion is present for every pair of X-Ray images, the relative motion between them can be approximately recovered by optimizing over the motion similarity [8].

C. Transformation Normalization

We describe the projection matrix of a cone-beam X-Ray frame i as

$$\mathbf{P}_i = \begin{bmatrix}\mathbf{K} & \mathbf{0}^\top\end{bmatrix} \mathbf{M}_i' \mathbf{R}_i \mathbf{M}_i \tag{4}$$

where the 4×4 matrix $\mathbf{R}_i$ describes the transformation from iso-center into detector coordinates (i.e. containing the rotation parameters of the cone-beam setup), and the 3×3 matrix $\mathbf{K}$ contains the intrinsic projection parameters. $\mathbf{M}_i'$ and $\mathbf{M}_i$ contain additional motion in detector and iso-center coordinates, respectively. Either of the latter ones can be optimized. For modeling patient motion, one would work in iso-center coordinates and describe the overall motion during a frame i as concatenation of incremental contributions from all previous frames:

$$\mathbf{M}_i = \prod_{k=1}^{i} \mathbf{T}_{i-k+1} \tag{5}$$

The relative transformations $\mathbf{T}_i$ computed by the motion detection algorithm for each single frame should add up to the identity matrix when closing the loop to the first frame (for a full 360° scan). In practice this is not the case due to drift and noise in the motion estimation. We therefore need a method for loop closing which takes the structure of the relative transformations into account but modifies them in such a way that they result in the identity transformation when concatenated.

More formally let us assume that we have n relative transformation matrices $\mathbf{T}_1, \mathbf{T}_2, \ldots, \mathbf{T}_n$ and a final relative transformation $\mathbf{C}$. We then want to find $\widehat{\mathbf{C}}_i$ so that

$$\begin{aligned}\mathbf{T} = \mathbf{C}\prod_{i=1}^{n}\mathbf{T}_{n-i+1} &= \left(\prod_{i=1}^{n}\widehat{\mathbf{C}}\right)\left(\prod_{i=1}^{n}\mathbf{T}_{n-i+1}\right) \\ &= \prod_{i=1}^{n}\widehat{\mathbf{C}}_{n-i+1}\mathbf{T}_{n-i+1}\end{aligned} \tag{6}$$

$\widehat{\mathbf{C}}$ is the n-th matrix root of $\mathbf{C}$. Assuming $\mathbf{C}$ is of the form

$$\mathbf{C} = \begin{bmatrix}\mathbf{A} & \mathbf{t} \\ \mathbf{0} & 1\end{bmatrix}; \quad \widehat{\mathbf{C}} = \begin{bmatrix}\widehat{\mathbf{A}} & \widehat{\mathbf{t}} \\ \mathbf{0} & 1\end{bmatrix} \tag{7}$$

and $\mathbf{A}$ is diagonalizable, we can compute the eigenvalue decomposition of $\mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1}$ and write $\widehat{\mathbf{C}}$ as above with

$$\widehat{\mathbf{A}} = \mathbf{V}\mathbf{D}^{\frac{1}{n}}\mathbf{V}^\top \tag{8}$$

$$\widehat{\mathbf{t}} = \left(\left(\sum_{i=1}^{n-1}\left(\prod_{j=1}^{i}\widehat{\mathbf{A}}\right)\right) + \mathbf{I}\right)^{-1}\mathbf{t} \tag{9}$$

Since $\mathbf{D}$ is a diagonal matrix $\mathbf{D}^{\frac{1}{n}}$ is obtained by simply taking the n-th root of each entry on the diagonal.

Finally, we need to move the $\widehat{\mathbf{C}}$ matrices into the coordinate frame of each transformation $\mathbf{T}_i$. The $\widehat{\mathbf{C}}_i$ in $\mathbf{T} = \prod_{i=1}^{n}\widehat{\mathbf{C}}_i\mathbf{T}_i$ resulting from this coordinate system change are given by

$$\widehat{\mathbf{C}}_i = \mathbf{T}_i\mathbf{G}^{-1}\widehat{\mathbf{C}}\mathbf{G}; \quad \mathbf{G} = \prod_{j=1}^{n-i+1}\mathbf{T}_{n-j+1} \tag{10}$$

For the transformation chain given in equation 5 the incremental transform $\mathbf{C}$ we correct is $\mathbf{C} = \mathbf{M}_i^{-1}$.

D. Hybrid Refinement

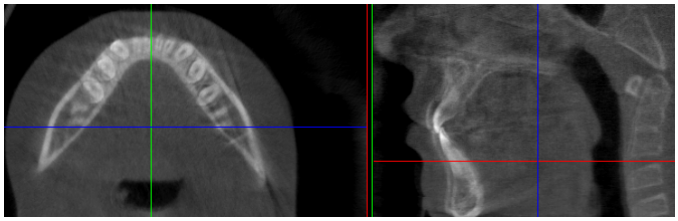
Our new overall algorithm operates as follows.
1) All projection images containing motion wrt. the previous frame are detected (including motion from the first frame to the last). This relative motion is successively found by optimizing over rigid transformation parameters for the detected frames. In the case of a full scan, the scheme described in section II-C is used to normalize the optimized transformations in order to avoid drift. A reconstruction after this step generally exhibits significantly reduced motion artifacts.
2) The main blocks of connected frames without motion are selected. Rigid parameters for the transformation of all blocks but the largest (which remains fixed) are fed into the global self-calibration algorithm. Before every evaluation of parameters, the transformations of all frames outside of blocks are normalized according to the interpolation scheme described in section II-C, in order to maintain the initial estimate from the motion compensation. This step usually yields the largest overall improvement, since the rough motion estimation on projection images is optimized using the residual error of the reconstruction, eliminating drift but keeping the overall structure of inter-frame motion.
3) A local self-calibration step (i.e. 2D-3D registration) is executed for all frames outside of blocks, hence refining transformation parameters in $\mathbf{M}_i'$ wrt. detector coordinates. An improved reconstruction is computed after the optimization of all frames. This step may be repeated once or twice, until the residual error does not improve over a selected ϵ threshold.
4) Optionally, a local self-calibration step on all frames is conducted. This would also ensure a final compensation of slight device inaccuracies.

Since both the motion similarity measure S and the self-calibration residual error E yield smooth cost function values, an optimization algorithm which internally approximates derivatives results in a significantly lower number of evaluations than other direct search methods. We therefore use the Bound Optimization by Quadratic Approximation (BOBYQA) method [9].

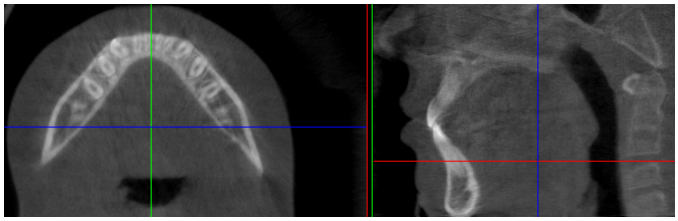
III. Results

A. Clinical Data

We have applied our algorithm on several orthodontic head scans with strong patient motion, however without ground truth information available. Figure 1 depicts axial and sagittal cross-sections of a reconstruction before and after our motion recovery method. In this example, the patient has moved his head several times during the sequence, with steady phases in between (which is addressed by the block optimization in step 2 of our method). The reduction of artifacts is clearly visible, in particular in the wrong "shadow" of the front teeth in the original reconstruction. This data set is a full scan with 450 frames and excentric detector motion for enlarged field of view. Five blocks of consistent successive frames have been found in this sequence, hence the block-based self-calibration in step 2 takes $(5-1)*6 = 24$ parameters. The non-linear optimization terminates after about 250 cost function evaluations, a single evaluation takes about 0.5 seconds with reconstruction volume dimension 256, on a NVIDIA GeForce GTX 670 GPU. All steps combined, therefore including the local self-calibration of all frames, yield a computation time in the order of five minutes. The following table shows the residual errors averaged over all frames after each step.



(a) original reconstruction



(b) after motion recovery

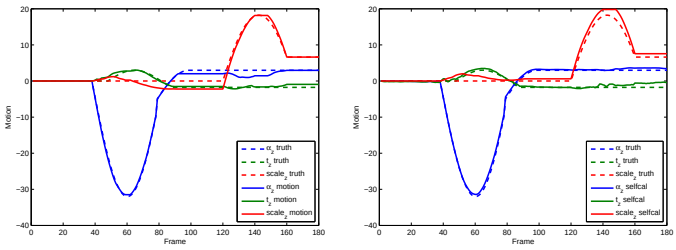
Fig. 1. Reconstruction result on a real patient before and after motion recovery

step	error	step	error
original	0.00983316	step 3	0.00900554
step 1	0.00926058	step 3 repeated	0.00899732
step 2	0.00903790	step 4	0.00888388

Those values are expressed in average absolute differences per pixel, with intensities normalized to $[0 \dots 1]$.

B. Synthetic Data

We also evaluated our method on a thorax CT scan with simulated patient motion, by generating 180 DRRs on a 360 degree trajectory around the patient. The artifical motion included rigid bulk motion as well as affine non-uniform scaling to resemble respiratory motion, with the resting states after motion being different than before (see dotted lines in figure 2). Running the projection-based motion optimization we were able to approximately recover the motion parameters (see figure 2a). Running the self-calibration step on these parameters further reduced the error particularly in frames between motion, as seen in figure 2b. Figure 3 shows an axial and a sagittal slice through the reconstructed patient volume (resolution 512^3) at different stages of the motion optimization together with the residual error of the reconstruction. The final reconstruction quality is close to the ground truth, both in terms of visual appearance and residual error. The average per-frame spatial error for the center point of the reconstruction volume and its direction were reduced from 6.53 ± 9.40mm to 0.66 ± 0.59mm in translation, and from $1.29 \pm 0.90^\circ$ to $0.55 \pm 0.34^\circ$ in rotation.



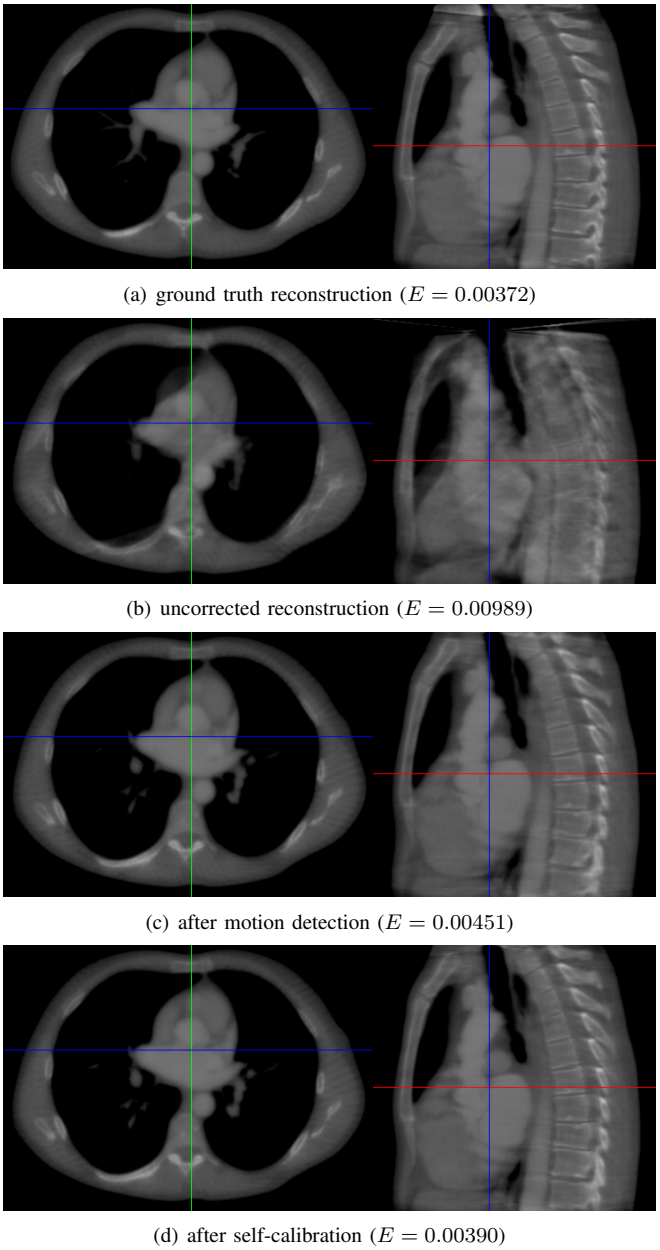
(a) after motion detection (step 1) (b) after-self calibration (step 4)

Fig. 2. Comparison of ground-truth and recovered patient motion

IV. Discussion

We have developed a new hybrid approach operating in both projection and reconstruction space to detect and recover patient or device motion in cone-beam CT. Detection of large motion is achieved using the projection based motion analysis, and approximately estimated with it. Only in combination with a self-calibration approach can the motion be precisely recovered however. The link between the intermediate results of relative motion between frames, and absolute motion parameters that are optimized using global block based self-calibration, is achieved by a normalization of the chain of transformation matrices.

We have shown that this method can restore severely compromised head scans, bootstrapping the motion information without any prior assumptions. To the best of our knowledge, this has not been achieved before. While these qualitative clinical results were obtained using a rigid parametrization, linear affine transformation matrices are supported as well (as demonstrated in section III-B), and should allow to deal with a majority of clinical scenarios involving respiratory motion. In order to tackle complex motion with significant local deformations (i.e. cardiac motion), adapted parametric motion models can be integrated with our method. For a given parameter configuration, it can then be evaluated simultaneously in projection and reconstruction space, how well it describes the actual motion. While this first approach suggested here is rather straightforward, we believe that such a joint analysis is a powerful foundation for dealing with motion in general. More complex motion models are the subject of future work, as well as thorough experiments on clinical data with ground truth (e.g. by tracking actual patient motion with auxiliary sensors). Further work is also required for a more theoretical understanding of the connection between the epipolar geometry in successive X-Ray images with intermediate results during reconstruction.



(a) ground truth reconstruction ($E = 0.00372$)

(b) uncorrected reconstruction ($E = 0.00989$)

(c) after motion detection ($E = 0.00451$)

(d) after self-calibration ($E = 0.00390$)

Fig. 3. Reconstruction results and residual errors without and with motion recovery on simulated data.

References

[1] C. Rohkohl, G. Lauritsch, L. Biller, and J. Hornegger, "ECG-gated interventional cardiac reconstruction for non-periodic motion," in *MICCAI 2010 Proceedings, Part I, LNCS 6361*, 2010, pp. 151–158.

[2] R. Zeng, J. Fessler, and J. Balter, "Estimating 3-D respiratory motion from orbiting views by tomographic image registration," *IEEE Trans. Med. Imag.*, vol. 26, pp. 153–163, Feb. 2007.

[3] E. Hansis, H. Schomberg, K. Erhard, O. Dössel, and M. Grass, "Four-dimensional cardiac reconstruction from rotational x-ray sequences - first results for 4D coronary angiography," in *SPIE Medical Imaging 2009 Conference*, vol. 7258, 2009.

[4] S. Brandt and V. Kolehmainen, "Motion without correspondence from tomographic projections by bayesion inversion theory," in *Computer Vision and Pattern Recognition (CVPR)*, vol. 1, 2004, pp. 582–587.

[5] M. Lyksborg, M. Hansen, and R. Larsen, "A statistical approach to motion compensated cone-beam reconstruction," in *IEEE International Symposium on Biomedical Imaging*, 2010, pp. 804–807.

[6] W. Wein, A. Ladikos, and A. Baumgartner, "Self-calibration of geometric and radiometric parameters for cone-beam computed tomography," in *Fully3D 2011 Proceedings*, Jul. 2011.

[7] Y. Kyriakou, R. Lapp, L. Hillebrand, D. Ertel, and W. Kalender, "Simultaneous misalignment correction for approximate circular cone-beam computed tomography," *Physics in Medicine and Biology*, vol. 53, pp. 6267–6289, 2008.

[8] W. Wein and A. Ladikos, "Detecting patient motion in projection space for cone-beam computed tomography," in *MICCAI 2011 Proceedings*, ser. Lecture Notes in Computer Science. Springer, Sep. 2011.

[9] M. J. D. Powell, "The BOBYQA algorithm for bound constrained optimization without derivatives," Department of Applied Mathematics and Theoretical Physics, Cambridge England, Tech. Rep. NA06, 2009.