A simplified motion model for estimating respiratory motion from orbiting views

Rongping Zeng[a], Jeffrey A. Fessler[a] and James M. Balter[b]

[a] EECS Dept., University of Michigan, Ann Arbor, MI48109-2122
[b] RadOnc Dept., University of Michigan, Ann Arbor, MI 48109-0010

ABSTRACT

We have shown previously that the internal motion caused by a patient's breathing can be estimated from a sequence of slowly rotating 2D cone-beam X-ray projection views and a static prior of of the patient's anatomy.[1,2] The estimator iteratively updates a parametric 3D motion model so that the modeled projection views of the deformed reference volume best match the measured projection views. Complicated motion models with many degrees of freedom may better describe the real motion, but the optimizations assiciated with them may overfit noise and may be easily trapped by local minima due to a large number of parameters. For the latter problem, we believe it can be solved by offering the optimization algorithm a good starting point within the valley containing the global minimum point. Therefore, we propose to start the motion estimation with a simplified motion model, in which we assume the displacement of each voxel at any time is proportional to the full movement of that voxel from extreme exhale to extreme inhale. We first obtain the full motion by registering two breath-hold CT volumes at end-expiration and end-inspiration. We then estimate a sequence of scalar displacement proportionality parameters. Thus the goal simplifies to finding a motion amplitude signal. This estimation problem can be solved quickly using the exhale reference volume and projection views with coarse (downsampled) resolution, while still providing acceptable estimation accuracy. The estimated simple motion then can be used to initialize a more complicated motion estimator.

Keywords: respiratory motion, simplified motion model, cone-beam CT scanner

1. INTRODUCTION

Geometric uncertainties caused by respiratory motion are always serious concerns for radiation therapy in the thorax and upper abdomen. To ensure more accurate X-ray dose delivery to the tumor area in those region, the anatomy movement during breathing should be considered. Therefore, intensive research work has been conducted to seek a whole picture of patients' anatomy during breathing.[3–8] We have worked toward this goal in recent publications.[1,2] In the papers we proposed to estimate respiratory motion from a sequence of slowly rotating cone-beam projection views, which we called deformation from orbiting views (DOV). DOV updates a B-spline based motion model to best match the modelled projection views to the actual cone-beam projection views. The modelled projection views are calculated from the deformed static anatomy prior of the patient. The DOV method yields good estimation accuracy. However, it involves heavy-loaded operations on data sets with very large size, such as the warping and forward projection of 3D volumes, plus a huge number of parameters associated with the estimation problem to be optimized. It is not surprising that very long computation time is required for solving the DOV problem, which may limit its usage in clinic.

To reduce the DOV computation time, we search ways to find a good starting point that can be used to initialize the optimization algorithm. The advantages of a good initialization include a reduced chance of the optimizations being trapped by local minima and fewer iterations for convergence hence speed up the estimation. This paper mainly discusses how we initialize the DOV estimation problem. We suggest a simple motion model and start DOV with a simplified motion estimation. The simplified motion model is based on a

motion proportionality assumption. We model the displacement of each voxel at any time as being proportional to the full movement of that voxel from end-exhalation to end-inhalation. Thus the estimation goal simplifies to finding a sequence of scalar motion proportionality parameters, given an estimate of the full deformation of the anatomy during breathing in. We obtain the full motion by registering two breath-hold CT volumes end-exhalation and end-inhalation. These two breath-hold CT volumes are usually available in clinic for treatment plans. The simplified motion estimation would be easily solved because the small number of parameters to be optimized. The estimated results can used to initialize the more complicated B-spline based motion estimator.

The paper is organized as follows. We first give an overview of the DOV method and state its limitation of long computation time. We then introduce an initialization method for the DOV optimization by a simple motion estimation. We present the simulation and phantom experiment afterwards.

2. OVERVIEW OF DOV

This simplified motion estimation was started for the purpose of easing the DOV optimization. So we first give a brief introduction on DOV. For a more detailed description, please refer to our previously published paper.[2]

2.1. DOV

The goal of DOV is to estimate respiratory motion from a sequence of X-ray cone-beam projection views for radiation therapy. It utilizes two data sets. One is a reference thorax CT volume obtained from a conventional fast CT scanner under breathhold conditions, denoted $f_{\mathrm{ref}}(\boldsymbol{x})$, $\boldsymbol{x} \in \mathbb{R}^3$. The other is a sequence of projection views of the same patient acquired under free breathing condition at treatment time using a slowly rotating cone-beam system, denoted $\boldsymbol{g}_m, m = 1, \cdots, M$, where M is the total number of projection views. Although the cone-beam scanners rotate slowly, the acquisition time of each projection view is short. For example, recently developed systems can acquire 15 frames per second, i.e., around 0.067 second per frame. We therefore assume the respiratory motion is negligible within each single projection view. Let the motion be denoted $\boldsymbol{\mathcal{T}}_{\boldsymbol{\theta}}(\boldsymbol{x}; t_m)$, a temporally variant deformable model controlled by parameters $\boldsymbol{\theta}$. Since the projection views and the reference volume are all from the same patient, a mathematical relationship between $\boldsymbol{g}_m$ and f_{ref} can be represented by the following equations, ignoring any noise and imaging artifacts,

$$\boldsymbol{g}_m = \mathcal{A}_{\phi_m} f_{t_m}, \tag{1}$$

$$f_{t_m}(\boldsymbol{x}) = f_{\mathrm{ref}}(\boldsymbol{\mathcal{T}}_{\boldsymbol{\theta}}(\boldsymbol{x}; t_m)), \tag{2}$$

where $\mathcal{A}_{\phi_m}$ denotes the projection operator from projection angle ϕ_m, and f_{t_m} is the deformed volume at time t_m. However, in practice the measured X-ray projection views $\boldsymbol{Y}_m$ are always degraded by noise, which we assume to be dominanted by Poisson effect,

$$Y_{m,n} \sim Poisson(I_{m,n}e^{-g_{m,n}} + S_{m,n}), \tag{3}$$

where $I_{m,n}$ is a constant related to the incident X-ray intensity, $S_{m,n}$ denotes the scatter contribution to $Y_{m,n}$ and n is the number of detector elements. The block diagram in Fig. 1 summarizes the system model described above, treating the reference volume f_{ref} as the input and the cone-beam projection views $\boldsymbol{g}_m$ as the output.
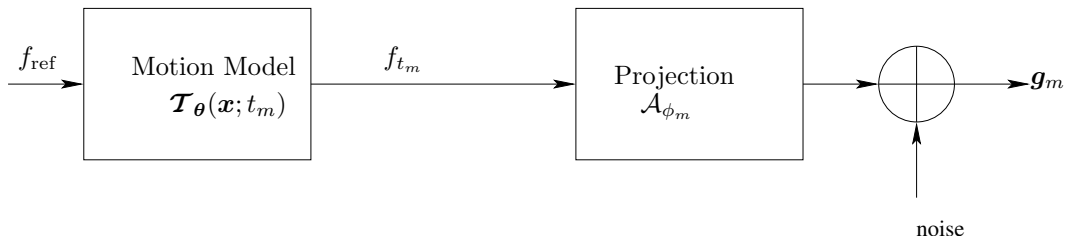
PSfrag replacements

Figure 1. Block diagram of the system model.

2

$\boldsymbol{\mathcal{T}_\theta}(\boldsymbol{x}; t)$ can be any suitable deformation models. In DOV we adopted a cubic B-spline based motion model as follows,

$$\boldsymbol{\mathcal{T}_\theta}(\boldsymbol{x}; t) = \boldsymbol{x} + \sum_j \sum_i \boldsymbol{\theta}_{j,i}\, \beta\left(\frac{t - \tau_j}{\Delta_\mathrm{t}}\right) \beta\left(\frac{\boldsymbol{x} - \boldsymbol{x_i}}{\Delta_\mathbf{x}}\right) \tag{4}$$

where $\beta(.)$ is the cubic B-spline function and $\beta(\boldsymbol{x})$ the tensor product of cubic B-spline functions, τ_j and $\boldsymbol{x_i}$ the spatial and temporal knot locations, $\Delta_\mathbf{x}$ and Δ_t control the width of the spatial and temporal basis functions respectively, and $\boldsymbol{\theta}$ the knot coefficients. There are two advantages using a cubic B-spline based model. One is that the small support of the cubic B-spline function eases the computation and optimization. The other is that the B-spline control grid can be flexibly adjusted to meet the approximation accuracy of continuous signals. Based on the motion model (4), the estimation goal is to find the motion parameters $\boldsymbol{\theta}$ from the projection views $\boldsymbol{g}_m$ and f_ref.

From the system model described in Fig. 1, we may use statistical estimation method to find the motion paramters $\boldsymbol{\theta}$. We construct a regularized estimator for this ill-posed problem. It contains three terms as follows,

$$\hat{\boldsymbol{\theta}} \quad = \quad \arg\min_{\boldsymbol{\theta}}(L(\{\boldsymbol{g}_m\}, \{\mathcal{A}_{\phi_m} f_\mathrm{ref}(\boldsymbol{\mathcal{T}_\theta}(\boldsymbol{x}; t_m))\}) + \beta_s R_s(\boldsymbol{\theta}) + \beta_t R_t(\boldsymbol{\theta})), \tag{5}$$

where $L(\cdot)$ is a data fidelity term, $R_s(\cdot)$ is a motion roughness penalty term, $R_t(\cdot)$ is a temporal motion aperiodicity penalty term, and β_s and β_t are scalars that control the trade-off between the three terms.

The data fidelity term $L(\cdot)$ measures the dissimilarity between the acquired and our estimated projection views. We select a correlation-based metric, Sum of squared differences (SSD) metric may be used. However, the X-ray energy spectra for imaging the static CT and for acquiring the cone-beam projection views often are not identical. Extra effort is needed to find the attenuation map between the two energy levels. So in practice the correlation-base metric may be preferable. The motion roughness penalty term $R_s(\cdot)$ discourages rapidly changing breathing motion estimates that would be unrealistic. It can be measured by the differences between the displacements of neighbouring voxels. We further simplify it to be the differences between the values of the neighbouring knot coefficients. With this simplification, the roughness penalty now takes the form of $R(\boldsymbol{\theta}) = \frac{1}{2}\|\boldsymbol{C}\boldsymbol{\theta}\|^2$, where $\boldsymbol{C}$ is a differencing matrix, with a typical row having the form $(\ldots, 0, -1, 1, 0, \ldots)$ for the first-order roughness penalty and $(\ldots, 0, -1, 2, 1, 0, \ldots)$ for the second-order roughness penalty. The aperiodicity penalty term $R_t(\cdot)$ encourages similarity between deformation estimates that correspond to similar respiratory phases. We add this penalty term to help overcome the limited gantry range for each breathing cycle. Current radiotherapy systems rotate $6°$ per second, spanning around $20 - 40°$ in a breathing cycle. The measured projection views for one breathing cycle in this limited angle interval may poorly reflect the motion along those projection directions. However they may better capture the motion along the perpendicular-to-projection directions. The aperiodicity term lends the motion information contained in the adjacent breathing cycles to help compensate for the angular limitation. We design this term as follows. If the temporal knots are evenly spaced in each breathing period and each breathing period contains the same number of knots, then the deformation similarity can be quantified by the closeness of the coefficients of knots located at the same relative position in the breathing periods, since those knots are belong to similar breathing phase bins. We need a respiratory marker to determine the interval of each breathing period in the scan duration. Without the use of extra respiratory monitering system, we extract a respiratory marker from the SI position of the diaphragm in the collected projection views. Further discussion of this aperiodicity penalty design can be found in our paper.[2]

We used the conjugate gradient (CG) algorithm for this minimization problem. The multi-resolution technique is also applied to reduce the chance being trapped by local minima.

2.2. Computation complexity

The implementation of this motion estimation method includes the following operators on 3D volumes: B-spline based image interpolation, B-spline based image deformation, and cone-beam forward projection. The computation complexity for the optimization algorithm is $O(CMVU)$, where C is the number of knots for describing the B-spline based motion, M is the number of projection views, V is the size of the static reference volume and U is the size of each projection views. The computation computation time is long. This may limit the usage of DOV in clinic. While we can optimize the computations themselves, we may find another way to speed up the convergence of the CG optimization algorithm, which we discuss in the next section.

3. A SIMPLIFIED MOTION ESTIMTION FOR INITIALIZATION

For nonconvex optimization problems with many number of parameters, there can be many local minima. However, if the iteration starts from some point located in the global minimum valley, the algorithm would be able to find a close-to-optimum solution. Besides, a good intialization can also speed up the convergence. Based on these motivations, we propose a simplified motion model and start the DOV with a simpler motion estimator that requires much less computation time.

The simplified motion model we designed is based on an motion proportionality estimation. We assume the displacement of each voxel at any time is proportional to the full movement $\mathcal{T}_{full}(\boldsymbol{x})$ of that voxel from the end-exhalation to end-inhalation. It be expressed as follows,

$$\mathcal{T}_\alpha(\boldsymbol{x};t) = \boldsymbol{x} + \alpha(t_m)\mathcal{T}_{full}(\boldsymbol{x}). \tag{6}$$

Given the same system model as in Fig. 1, the goal of the simplified motion estimation is to find a motion amplitude signal $\alpha(t_m), m = 1, \ldots, M$.

We determine the parameters $\boldsymbol{\alpha}$ by solving the following minimization problem,

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \left(L\big(\{\boldsymbol{g}_m\}, \{\mathcal{A}_{\phi_m} f_{ref}(\mathcal{T}_{\boldsymbol{\alpha}}(\boldsymbol{x};t_m))\}\big) + \lambda R(\boldsymbol{\alpha}) \right). \tag{7}$$

Similar to the previous cost function (5), $L(\cdot)$ in the cost function is a data fidelity term, measuring the dissimilarity between the actual projection views and the modelled projection views. $R(\cdot)$ is a roughness penalty term, taking the form of $||\boldsymbol{C}\boldsymbol{\alpha}||^2$, where $\boldsymbol{C}$ is a differencing matrix.

The simplified motion model (6) is build upon a prior of the full deformation from the beginning to the end of inhalation. This full deformation can be obtained by registering two reference CT volumes of the patient, one at end-exhalation and one at end-inhalation, denoted $f_{ex}(\boldsymbol{x})$ and $f_{in}(\boldsymbol{x})$ ($\boldsymbol{x} \in \mathbb{R}^3$) respectively. Usually these two reference volumes are clinically available for treatment plan. One of them can be used as the reference volume in the estimator (5) and (7. We used a B-spline based deformation model for registration. The deformation parameters for $\mathcal{T}_{full}(\boldsymbol{x})$ are estimated by minimizing a cost function containing a SSD similarity term and a Jacobian penalty term.[9] SSD can be used for this registration since the two reference volumes are with the same imaging modality. The Jacobian penalty term discourages irreversible deformation estimates.

The minimization problem of this simplified motion estimation is relatively easy to solve because of the small number of parameters involved. We used the CG algorithm for optimization. The solution $\hat{\boldsymbol{\alpha}}$ can be utilized twofold by DOV. One is for initializing DOV, as stated previously. We implement a least square fitting of $\hat{\boldsymbol{\alpha}}\mathcal{T}_{full}(\boldsymbol{x})$, the estimate of a simplfied respiratory motion, into the B-spline motion model. The fitted motion is then input to the DOV optimization algorithm. The second usage of $\hat{\boldsymbol{\alpha}}$ is on improving the aperidicity penalty design. $\hat{\boldsymbol{\alpha}}$ may provide us a better estimate of the patient's respiratory marker than the one extracted from the projection views based on the SI transition of the diaphgram. Thus it can guide us to place the temporal knots more correctly and the aperiodicity penalty term would be able to find more accurate phase correspondences among the knots.

4. EXPERIMENT RESULTS AND DISCUSSION

This section presents the simulation and phantom experiment results. We mainly compared the performances of DOV in two cases. One is DOV with "zero" initialization, in which we start DOV from a coarser resolution with all parameters initialized to be zeros. In the coarser resolution, we downsampled the references volumes and the projection views. The other case is DOV starting from the simplified motion estimates, in which we first run the simplifed motion estimation and then the B-spline based motion estimation with initialization of the simplified motion estimates.

4.1. Simulation

The simulation setup was based on three real clinical planning CTs (0%, 20%, 60% vital capacity above tidal exhale), with a voxel size of $2 \times 2 \times 5 \text{mm}^3$. We selected the 0% and 60% CT volumes as the end-exhalation and the end-inhalation reference volumes. We then simulated a cone-beam system and generated 70 dynamic projection views over 180° of the deformed 0% CT. The warping was based on a synthetic respiratory motion, obtained by temporally interpolating the deformations between 0% to 20% and 60% CT volumes. The simulated cone-beam system had a flat-panel detector of 180×200 elements of size $4 \times 4 \text{mm}^2$. The source to isocenter distance and the isocenter to detector distance were 1000mm and 500mm respectively. To simulate realistic projection views, we also contaminated the projection views with scatter effect and poisson noise.

We registered the 0% and the 60% CTs to obtain the full deformation. For the simple motion estimation, we downsampled the reference volume by 2 in the transaxial plane and correspondingly downsampled each projection view by 2. The optimization converged at about the 20th iteration, with each iteration taking about 30sec. Since the simple proportionality motion model has a small number of degrees of freedom, using the downsampled data would not degrade the accuracy of the simple motion estimates but would greatly reduce the computation time. This conjecture agrees with the results presented in Table 1. As seen in this table, the simple motion estimates themselves already attained acceptable accuracy. Fig. 2 shows the estimated motion proportionality parameters. Feeding the estimated simplified motion into DOV, it converged with considerable less iterations comparing to the DOV with "zero" initialization. as seen in the convergence plot Fig. 3. The estimation accuracy of these two cases are almost the same, as listed in Table 2. However the former one required fewer iterations.

Mean error/ STD (mm)	LR	AP	SI
Downsampled by 2	-0.05/0.19	0.03/0.37	-0.09/0.88
W/O downsampling	-0.05/0.18	0.03/0.35	-0.09/0.84

Table 1. Estimation accuracy for the simple motion estimation on the simulation data. The table shows the mean and the standard deviation (STD) of the errors over the entire volume through time. The estimation accuracy with data downsampled by 2 and without data downsampling attained very simular performance.

Mean error/ STD	LR	AP	SI
DOV with "zero" initialization	0.11/0.60	0.06/0.76	0.20/1.91
DOV starting from simplified motion estimates	-0.02/0.22	0.04/0.43	-0.03/0.90

Table 2. Estimation accuracy of DOV with "zero" initialization and DOV starting from simplified motion estimates

4.2. Phantom experiment

We also tested the simple motion estimation on a set of phantom data. The data set is the same to the one we collected for our previous paper.[2] Readers may refer to that paper for the detail on the experiment setup and data collection. We only give a brief description on the phantom data in the follwing text. The phantom was composed of a rigid frame and a compressible foam compartment inside, with some balls inserted. A rigid, flat plastic board with one side connected to a linear actuator was placed at the bottom of the phantom and could compress and deform the material inside by moving back and forth. Fig 4 displays the central cut of the phantom in three directions. We collected 668 cone-beam projection views the moving phantom with 360° rotation. Fig. 5 displays a few collected projection views. During the scan the plate moved in a higher power of sinosoidal pattern with varying amplutudes and periods for each cycle. The motion profie was saved in a file and can be used later to evaluate our simple motion estimates. Other than the cone-beam projection views, we also obtained three static CT volumes of the phantom (CT0, CT1, CT2) using conventional scanners.

For the purpose of motion estimation, we only used the projection views in the first 180° interval and downsampled them by 4 in the temporal axis. So there were about 80 views spanned over 180° used by DOV. CT0 and CT2 were selected as the reference volumes. To evaluate the performance, we established a "ground truth" using the landmark method. We identified five landmarks (such as the centers of balls in the phantom) and obtained their displacements at three motion phases by registering the three static CT volumes. We assumed
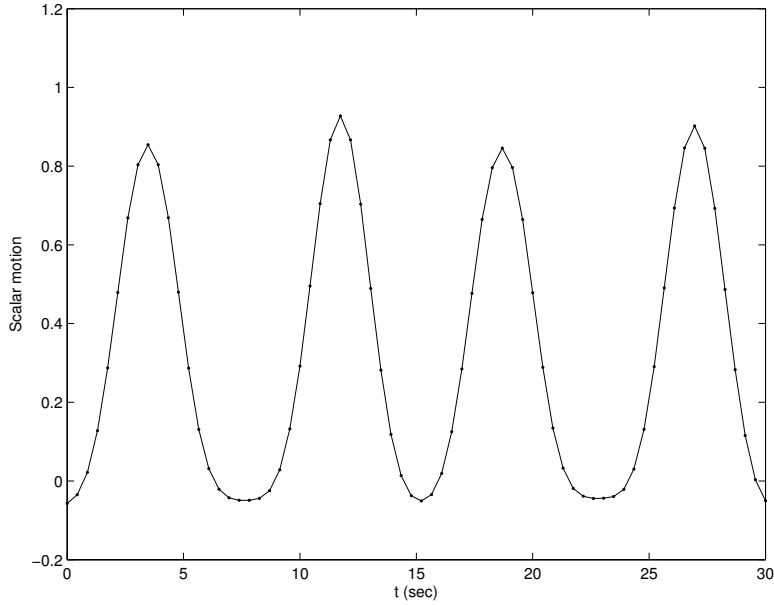
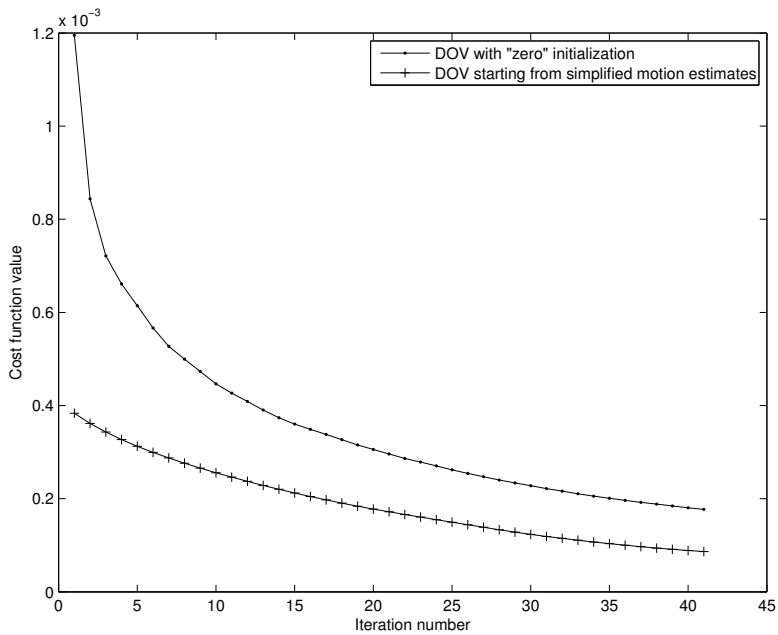**Figure 2.** The estimated motion proportionality parameters $\hat{\boldsymbol{\alpha}}$.

Figure 3. Convergence curves of the DOV with "zero" initialization and DOV starting from simplified motion estimates

the registration results to be the truth and compared the estimated motion of the landmarks at the corresponding phases to the "truth". The motion phase associated with each cone-beam projection view can be decided by the motion profile.

The simplified motion estimator yielded a sequence of proportionality parameters that resembled the motion profile closely, as shown in Fig. 6. Again, by initializing with the simplified motion estimates, DOV achieved faster convergence. The estimation accuracy of the simplified motion estimates and the motion estimates of the two DOV cases are listed in Table 3. As can be seen in the this table, DOV starting from simplified motion estimates
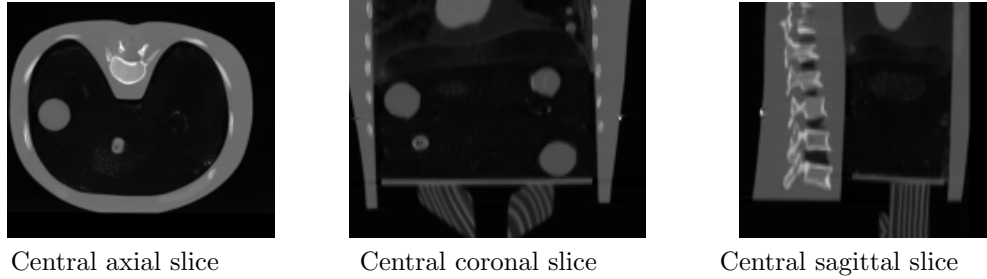
Central axial slice　　　　Central coronal slice　　　　Central sagittal slice

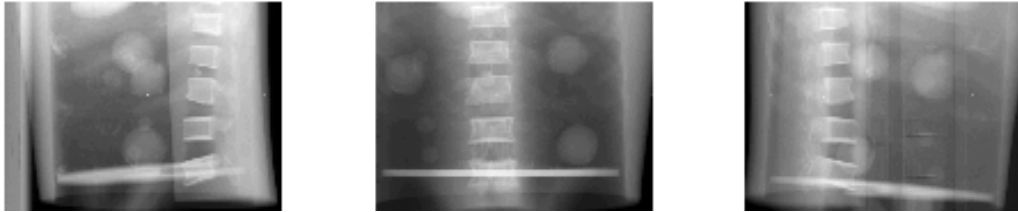**Figure 4.** The central transaxial, coronal and saggital slices of the phantom.

Figure 5. Collected cone-beam projection views at angles $181.7°, 95.7°, 29.0°$.

improved the estimation accuracy in the LR and AP directions, comparing to those with "zero" initialization. Table 3 also indicates that the optimization algorithm for B-spline based motion esitmation actually ran further away from the truth comparing to the simplified motion estimates. This phenomenan may be explained as follows. The collected projection views contained severe scatter effects and table artifact, as seen in Fig. 5. Our estimator treated the useful data information and those artifacts equally. Although a more complex motion model could characterize local deformations better, it might overfit the noise artifacts on the other hand. In addition,the actual phantom motion may agree with the motion proportionality assumption. Thus the estimator with the simple motion model could find closer-to-truth solutions, while the one with B-spline motion model performed worse. This phenomenan suggests that a posterior scatter correction may improve the DOV motion estimation accuracy.
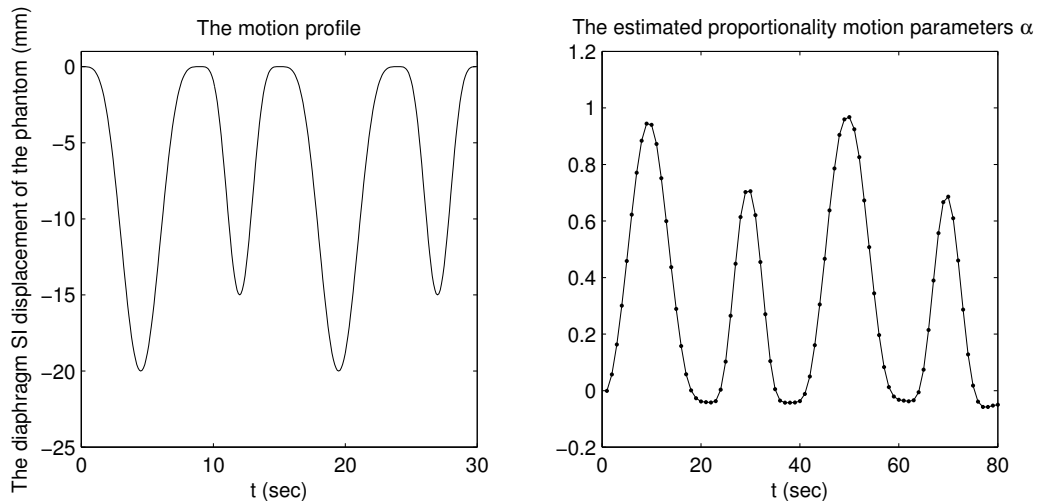

Figure 6. The motion profile and our estimated scalar motion proportionality parameters $\hat{\alpha}$. The estimates resemble the pattern of the motion profile.

Mean error/ STD(mm)	LR	AP	SI
Simplified motion estimator	-0.02/0.06	0.01/0.07	0.47/0.38
DOV with "zero" initialization	0.42/2.49	1.08/2.24	0.24/1.72
DOV starting from simplified motion estimates	0.41/0.68	1.03/0.91	-0.04/2.19

Table 3. Estimation accuracy of the simplified motion estimator, DOV with "zero" initialization and DOV starting from simplified motion estimates. The table shows the mean and the STD of the errors over the landmarks.

5. CONCLUSION

This paper discusses a method to speed up the convergence of our DOV method by providing it an initialization through a simplified motion estimation. The simplified motion estimator uses a proportionality motion model. It can be solved fast by computer because of a small number of parameters contained in the simplified motion estimator. The simplified motion estimates themselves attained good accuracy in both the simulation and phantom examples, hence maybe used alone. Starting from the simple motion estimates, DOV converges with fewer iterations, greatly reducing the DOV computation time.

ACKNOWLEDGMENTS

REFERENCES

1. R. Zeng, J. A. Fessler, and J. M. Balter, "Respiratory motion estimation from slowly rotating X-ray projections: Theory and simulation," *Med. Phys.* **32**, pp. 984–91, Apr. 2005.
2. R. Zeng, J. A. Fessler, and J. M. Balter, "Estimating 3D respiratory motion from orbiting views by tomographic image registration," *IEEE Trans. Med. Imag.* , 2006. To appear.
3. D. A. Low, M. Nystrom, E. Kalinin, P. Parikh, J. F. Dempsey, J. D. Bradley, S. Mutic, S. H. Wahab, T. Islam, G. Christensen, D. G. Politte, and B. R. Whiting, "A method for the reconstruction of four-dimensional synchronized CT scans acquired during free breathing," *Med. Phys.* **30**, pp. 1254–63, June 2003.
4. E. Rietzel, T. Pan, and G. T. Y. Chen, "Four-dimensional computed tomography: Image formation and clinical protocol," *Med. Phys.* **32**, pp. 874–89, Apr. 2005.
5. T. Pan, T.-Y. Lee, E. Rietzel, and G. T. Y. Chen, "4D-CT imaging of a volume influenced by respiratory motion on multi-slice CT," *Med. Phys.* **31**, pp. 333–40, Feb. 2004.
6. T. Kleshneva, J. Muzik, and M. Alber, "An algorithm for automatic determination of the respiratory phases in four-dimensional computed tomography," *Phys. Med. Biol.* **51**, pp. N269–76, Aug. 2006.
7. J.-J. Sonke, L. Zijp, P. Remeijer, and M. . Herk, "Respiratory correlated cone beam CT," *Med. Phys.* **32**, pp. 1176–86, Apr. 2005.
8. T. Li, E. Schreibmann, Y. Yang, and L. Xing, "Motion correction for improved target localization with on-board cone-beam computed tomography," *Phys. Med. Biol.* **51**, pp. 253–68, Jan. 2006.
9. J. Kim and J. A. Fessler, "Nonrigid image registration using constrained optimization," in *SIAM Conf. Imaging Sci., Abstract Book*, 2004.