

Optical Flow Motion Estimation for Approximate Motion Compensation in Cone-Beam CT

Colas Schretter, Fabian Pilatus, Georg Rose, Til Aach and Matthias Bertram

Abstract—Image quality of volumetric image reconstructions is often degraded by residual patient motion when using slowly rotating C-arm systems or radiotherapy linear accelerators. If the patient does not manage to hold his or her movements during data acquisition, the image resolution is impaired by strong motion blur artifacts. This work proposes a method to detect and to estimate arbitrary patient motion. The motion information is used within a motion-compensated variant of the FDK algorithm to improve the image quality. Results of two related experiments are shown. First, the motion is estimated in projection space by optical flow-based elastic image registration, using reference projections of a known static image. The motion-compensated FDK reconstruction from a simulated free breathing acquisition shows very good agreement with the objective image. Second, the motion is estimated from approximate reference projections that are computed by the SART method, using solely the acquired data.

Index Terms—X-ray tomography, image registration, motion estimation, motion segmentation, motion compensation.

I. INTRODUCTION

MOTION ESTIMATION is an ubiquitous problem in medical applications of computed tomography (CT). Each acquired projection is a sharp snapshot of the anatomy. However, the whole acquired dataset might be inconsistent if organ motion occurred during the acquisition. Data inconsistencies introduce motion blur artifacts in the reconstructed image. Those artifacts are even more prominent when using a slowly rotating gantry CT scanner, such as interventional C-arm or radiotherapy systems. On these systems, a typical acquisition last for about 8 s to 20 s and often, the patient is not able to hold the breath constantly. Furthermore, other unwanted non-periodic movements such as nervous shaking or bowel contractions can corrupt the data.

Optical flow has been applied with success for motion estimation in medical imaging. Gilland *et al.* used a variant of the seminal paper of Horn and Schunck [1] for estimating cardiac motion in emission tomography [2]. The present work aims to improve the image quality of soft tissue volumetric imaging on C-arm systems and relies on the optical flow-based elastic image registration algorithm proposed by Barber *et al.* [3]. His algorithm provides an automatic regularization of the deformation grid and is both fast and robust.

Our motion estimation and compensation approach has been first developed in the context of parallel-beam tomography [4], using the exact reconstruction algorithm of Desbat *et al.* [5]. The present work extends the method to the cone-beam geometry and validates it on physiologically plausible motion. The motion-compensated reconstruction relies on the heuristic algorithm of Schäfer *et al.* [6]. Taguchi and Kudo analyzed the properties of this algorithm [7] and concluded that “the fan-beam and cone-beam versions of Schäfer’s method are a very good approximation of the exact method.” This statement is confirmed by our experiments.

This work received financial support from the Marie Curie program of the European Commission under contract MEST-CT-2005-020424.

C. Schretter and M. Bertram are with Philips Research Europe, Weisshausstr. 2, 52066 Aachen, Germany (e-mail: colas.schretter@philips.com).

F. Pilatus and T. Aach are with the Institute of Imaging and Computer Vision, RWTH Aachen University, 52062 Aachen, Germany.

G. Rose is with the Otto-von-Guericke University, Universitätsplatz 2, 39016 Magdeburg, Germany.

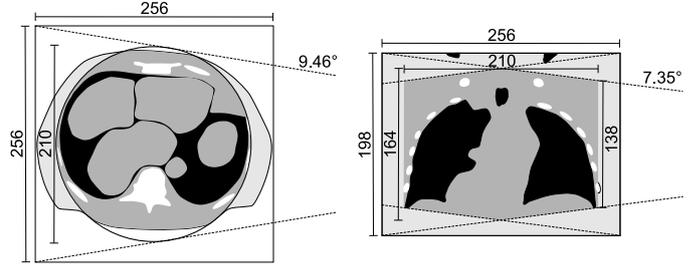


Figure 1. Illustration of the reconstructed region of interest (ROI) observed by all projections, in cone-beam geometry. The ROI is a truncated cylinder capped by two *Chinese hats* and is smaller than the whole field of view (FOV). The fan angle is 18.9° . The maximum cone angle is 14.7° . The definition of reconstructed images is $256 \times 256 \times 198$ voxels.

The remaining of this paper is organized as follows. Notations and the approximate motion-compensated Feldkamp Davis Kress (FDK) algorithm are defined in section II. Section III outlines the motion estimation and segmentation method. Results of experiments conducted on a realistic respiratory phantom are discussed in section IV. Finally, conclusions are drawn in section V.

II. BACKGROUND

With circular trajectory and cone-beam geometry, divergent X-rays are emitted from a point source and attenuated intensities are measured by a planar detector. The point source is located at distance R from the rotation axis and the detector is positioned at distance $R + D$ from the point source such that the radius of the cylindrical field of view (FOV) is equal to D . The region of interest (ROI) that should be reconstructed is composed of image voxels that are observed in every projection. The geometry of a realistic C-arm system illustrated in Fig. 1 is simulated by choosing $R = 5D$.

When the arm is oriented at angle $\alpha \in [0, 2\pi)$, the position of the point source is Rd_α where the vector $d_\alpha = (\cos \alpha, \sin \alpha, 0)$ is normal to the detector plane. The perspective projection operator

$$P_\alpha(x, y, z) = (y \cos \alpha - x \sin \alpha, z) \frac{(R + D)}{U} \quad (1)$$

maps a point $(x, y, z) \in \mathbb{R}^3$ defined in object space to a point $(u, v) \in \mathbb{R}^2$ defined in projection space. The denominator in (1) is the *perspective factor*

$$U = U_\alpha(x, y) = R + x \cos \alpha + y \sin \alpha \quad (2)$$

which is equal to the distance between the source and the orthogonal projection of the voxel position on the central plane. The central plane contains the source point and is orthogonal to the rotation axis.

The adjoint of the perspective projection operator

$$P'_\alpha(u, v) = (-u \sin \alpha, u \cos \alpha, v) \frac{R}{(R + D)} \quad (3)$$

maps a point $(u, v) \in \mathbb{R}^2$ defined in projection space to a point on a *virtual detector* defined in object space. The constant ratio $R/(R + D)$ is called the *magnification factor*. The virtual detector

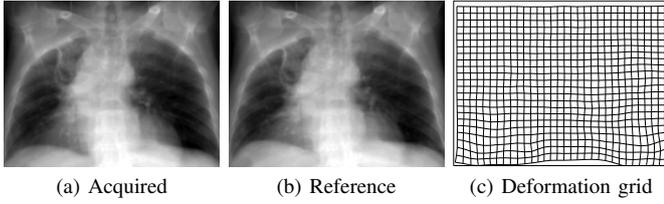


Figure 2. A specific deformation from motion estimation in projection space. Each acquired projection (a) is deformed on the corresponding reference projection (b) by elastic image registration based on optical flow. The resulting reverse deformation grid (c) is an approximation of the perceived motion. Independent registrations are performed for every pairs of projection.

shares the same orientation with the real one, but is smaller, proportionally to the magnification factor and centered on the origin of the FOV. Note the relation

$$P_\alpha (P'_\alpha (u, v)) = (u, v) \quad \forall (u, v) \in \mathbb{R}^2, \quad (4)$$

however (3) is not the inverse of (1) since $P'_\alpha (P_\alpha (x, y, z)) = (x, y, z)$ is true only for points that lie on the virtual detector, when $U = R$ and thus, when $x \cos \alpha + y \sin \alpha = 0$.

Let $f_t(x, y, z) \rightarrow \mathbb{R}$, a dynamic volumetric image where $(x, y, z) \in \mathbb{R}^3$ are Cartesian coordinates in image space and the subscript $t \in [0, 1]$ is a normalized time variable. The function f_t is compactly supported in the cylindrical FOV such that $f_t(x, y, z) = 0$ when $\sqrt{x^2 + y^2} > D$. Let $g_t(u, v) \rightarrow \mathbb{R}$, the line integrals of f_t where $(u, v) \in \mathbb{R}^2$ are Cartesian coordinates in projection space. The values of acquired line integrals are equal to

$$g_t(u, v) = \int f_t((1-s)Rd_\alpha + sP'_\alpha(u, v)) ds, \quad (5)$$

with $\alpha = 2\pi t$. Points of the integrated line segment connecting the X-ray source to a pixel of the detector are selected by varying the integration parameter $s > 0$.

A. Motion Model

The motion model can be seen as an extension of the admissible class of motion suggested by Desbat *et al.* [5] and is represented by a dynamic displacement vector field in projection space $D_t(u, v) \rightarrow \mathbb{R}^2$ and a normalized scalar field in image space $M(x, y, z) \in [0, 1]$. The associated image M indicates for each voxel if some motion occurred during the acquisition. This segmentation information is a key for successful local motion compensation.

The displacement of image elements in object space is modeled by backprojecting the displacement field D_t and weighting vectors with the scalar field M . Displacements expressed by

$$\Delta_t(x, y, z) = P'_\alpha(D_t(P_\alpha(x, y, z)))M(x, y, z), \quad (6)$$

with $\alpha = 2\pi t$. The trajectories of image elements along time are given by applying the displacement in (6), relatively to the initial position of image elements. Trajectories are expressed by

$$\Gamma_t(x, y, z) = (x, y, z) + \Delta_t(x, y, z), \quad (7)$$

where the displacement vectors of Δ_t lie on the detector plane and thus represent only the component orthogonal to the projection ray.

A particularity of the model is that neither periodicity nor spatial or temporal smoothness of the underlying motion is assumed. Therefore it could capture unwanted sudden patient motion such as hiccups, breath-hold failures, or bowel movements, for example. Although the model is approximate, our experiments demonstrate that it has the potential to capture local motion very well. The success of motion compensation mainly depends on the accuracy of motion estimation.

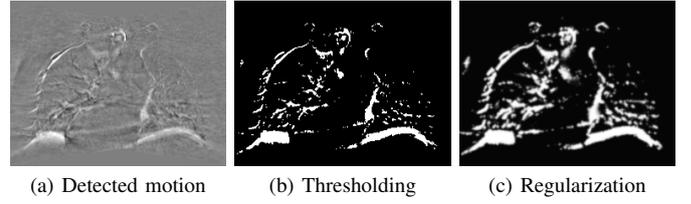


Figure 3. A specific coronal slice from motion segmentation in image space. The motion is first detected by reconstruction of absolute differences between acquired and reference projections (a). Then, a segmentation is obtained by thresholding (b). This binary map is regularized with a non-linear max filter and edges are smoothed by Gaussian filtering to produce the motion map (c).

B. Motion-Compensated Image Reconstruction

For motion compensation, a modification is introduced within the FDK algorithm by displacing the projected position of voxels before fetching the pre-weighted filtered line integrals, denoted here by g_t^* . The motion model defined in (7) is inserted in the back-projection of FDK and the reconstruction of a static volumetric image $f(x, y, z) \rightarrow \mathbb{R}$ is computed as follows:

$$f(x, y, z) = \int_0^1 \frac{R^2}{U^2} g_t^*(P_\alpha^\Gamma(x, y, z)) dt, \quad (8)$$

with $\alpha = 2\pi t$ and

$$\begin{aligned} P_\alpha^\Gamma(x, y, z) &= P_\alpha(\Gamma_t(x, y, z)) \\ &= P_\alpha((x, y, z) + P'_\alpha(D_t(P_\alpha(x, y, z))))M(x, y, z) \\ &= P_\alpha(x, y, z) + D_t(P_\alpha(x, y, z))M(x, y, z). \end{aligned} \quad (9)$$

Using the property pointed out in (4), the motion-compensated FDK algorithm can be implemented in projection space by displacing the projected position of the current image element, as shown by the last expression in (9).

III. METHOD

The motion estimation method is split into three sequential steps which are solved by standard algorithms from the image reconstruction and image processing communities. First, a sequence of reference projections is synthesized from the acquired projections, using a pilot image reconstruction. Second, a 2D displacement vector field that maps every acquired projection on its corresponding reference projection is computed using optical flow-based elastic image registration, as shown in Fig. 2. Third, a binary segmentation is obtained from the reconstruction of absolute difference between acquired and reference projections, as shown in Fig. 3. The motion segmentation can be executed in parallel with motion estimation. Finally, a static image is reconstructed with the modified FDK algorithm in (8) that uses both the motion estimate and the segmentation to apply locally a compensation during the back-projection step. The three steps of the method will be described in more details in the following sections.

A. Computing Reference Projections

Motion estimation requires a reference static image $\hat{f}(x, y, z) \rightarrow \mathbb{R}$ from which a set of reference projections $\hat{g}_\alpha(u, v)$ is forward projected, with one projection per acquisition angle $\alpha \in [0, 2\pi)$. The simultaneous algebraic reconstruction technique (SART) proposed by Andersen and Kak [8] is used to reconstruct a pilot image from the available acquired projections. By principle, iterative image reconstruction techniques try to estimate an image such that the error between forward projection and input projections is minimized. A corollary is that the forward projection of the image will be more

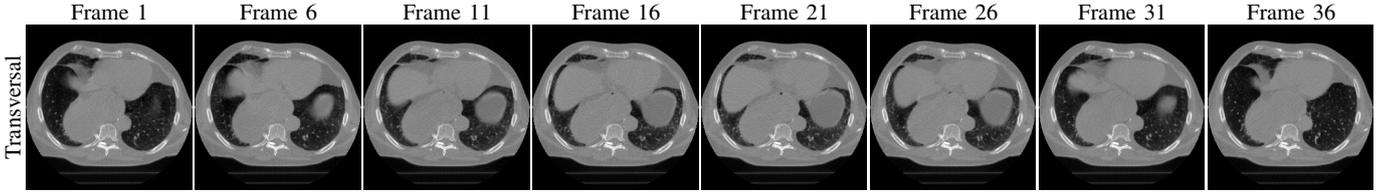


Figure 4. Transversal views of frames 1, 6, 11, 16, 21, 26, 31, and 36 of the breathing dynamic phantom from a respiratory-gated helical CT acquisition. The first four and last four frames correspond to exhalation and inhalation, respectively. Pulled by contraction of the thoracic diaphragm, organs leave the selected transversal slice during inhalation. By relaxing the contraction of the diaphragm, the air is expelled naturally during exhalation.

robust to possible projection truncations that often arise in cone-beam geometry, see Fig. 1.

It is known that, if data are consistent, the reference projections will match the acquired projections in the weighted least square sense when using SART. However, if acquired data are corrupted by unwanted patient motion, the iterative reconstruction will never converge and the image will contain motion blur artifacts. This pilot image is nevertheless valuable for sampling approximate reference projections. Experiments demonstrate that even if the resolution of those approximate reference projections is limited by motion blurring, their quality is sufficient for motion estimation in projection space. The SART method is used because of its fast convergence rate but another reconstruction algorithm could have been used as well.

B. Motion Estimation in Projection Space

The perceived deformation between each pair of acquired projections g_t and corresponding reference projections \hat{g}_α is represented in projection space by the bijective mapping function $D_t(u, v) \rightarrow \mathbb{R}^2$ that map $g_t(u, v)$ on $\hat{g}_\alpha(u, v)$, with $\alpha = 2\pi t$:

$$\hat{g}_\alpha(u, v) = g_t((u, v) + D_t(u, v)), \forall (u, v) \in \mathbb{R}^2. \quad (10)$$

In this paper, the displacement vectors D_t are computed using the optical flow principle introduced by Horn and Schunck in 1981 [1]. The brightness of a particular point of the projection g_t is assumed to remain constant over time, so that

$$\frac{\partial g_t}{\partial t} = -D_t \cdot \left(\frac{\partial g_t}{\partial u}, \frac{\partial g_t}{\partial v} \right). \quad (11)$$

Provided that the displacement vectors in (11) are small, the chain rule for differentiation gives

$$\hat{g}_\alpha - g_t = D_t \cdot G_t, \quad (12)$$

where G_t contains gradient vectors that capture the direction and the amplitude of the intensity change between the two images \hat{g}_α and g_t . Partial derivatives are approximated by finite differences.

For robustness, the components of gradient vectors are computed for both the source and target images and then averaged such that

$$G_t = \frac{1}{2} \left(\frac{\partial \hat{g}_\alpha}{\partial u} + \frac{\partial g_t}{\partial u}, \frac{\partial \hat{g}_\alpha}{\partial v} + \frac{\partial g_t}{\partial v} \right). \quad (13)$$

In the implementation, vectors of (13) are averaged at a coarse grid of control points and then interpolated back for every pixel. This procedure improves the numerical stability but also reduces the resolution of the deformation vector field. The optimal deformation D_t is iteratively estimated by the conjugate gradient descent implementation of Barber *et al.* [3]. The objective function minimizes the sum of squared differences between \hat{g}_α and g_t while a regularization term penalizes the updates of D_t to avoid irregularities in the final deformation. The penalization is proportional to the Laplacian of the deformation vector field computed in the previous iteration. For the

first iteration, the initial deformation is a zero vector field and no regularization term is used.

C. Motion Segmentation in Image Space

For detecting motion in image space, data inconsistencies are first identified in projection space by computing the absolute differences $d_t = |\hat{g}_\alpha - g_t|$, $\alpha = 2\pi t$. The differences d_t are evaluated independently for each acquired projection and are reconstructed to detect the voxels strongly affected by motion. The reconstruction relies on the regular FDK algorithm. Finally, a threshold is applied to yield a binary image that is regularized to give the final motion map.

In experiments the reconstruction of differences is post-processed with the following procedure. First, a binary mask is created by thresholding the image at 100 Hounsfield units (HU). Then, a non-linear max filter with a spherical structural element of radius 2 is applied to grow the segmented region. Finally, edges between moving and non-moving regions are smoothed by convolving the binary image with a volumetric Gaussian kernel. Note that the end result is the same if the threshold is applied after the max filter. Alternative post-processing pipelines may equally fulfill the regularization task.

IV. RESULTS

In this paper, all slices are extracted from volumetric images represented by a Cartesian grid of $256 \times 256 \times 198$ isotropic voxels of size equal to 1.36 mm. Gray is set to the attenuation of water and the window width equals 1000 HU such that black corresponds to the attenuation of air. To simulate a motion-corrupted acquisition, a set of 360 projections for one full circular rotation was forward-projected from the dynamic free-breathing phantom shown in Fig. 4.

The dynamic phantom is a sequence of 40 frames reconstructed from a respiratory-gated helical CT acquisition. The speed of the CT gantry was 0.444 s per rotation. The duration of the acquisition was 93.3 s during which 26 breathing cycles were observed. The mean duration of breathing cycles was 3.6 s. Images still contain some blurring, due to helical artifacts and residual motion within each gate of the breathing cycle. The simulated acquisition time was 12 s, matching the typical rotation speed of a C-arm system when used for soft tissue imaging. The experiments validate the method on a challenging scenario, when the patient is breathing freely.

For the first experiment, reference projections have been computed by the SART method while for the second experiment, reference projections have been simulated from a chosen reference frame of the dynamic phantom. Using a ground truth static reference image, it is possible to assess the accuracy of the motion estimation and compensation method by measuring the similarity of the reconstructed image with the reference image. Quantitative analyses consider only voxels where motion compensation is applied. The mean absolute error (MAE) equals 151 HU without motion compensation and 61 HU with reference motion estimation.

Results of image reconstructions are shown in Fig. 5. Without compensation, the border of the myocardium and vessels appear

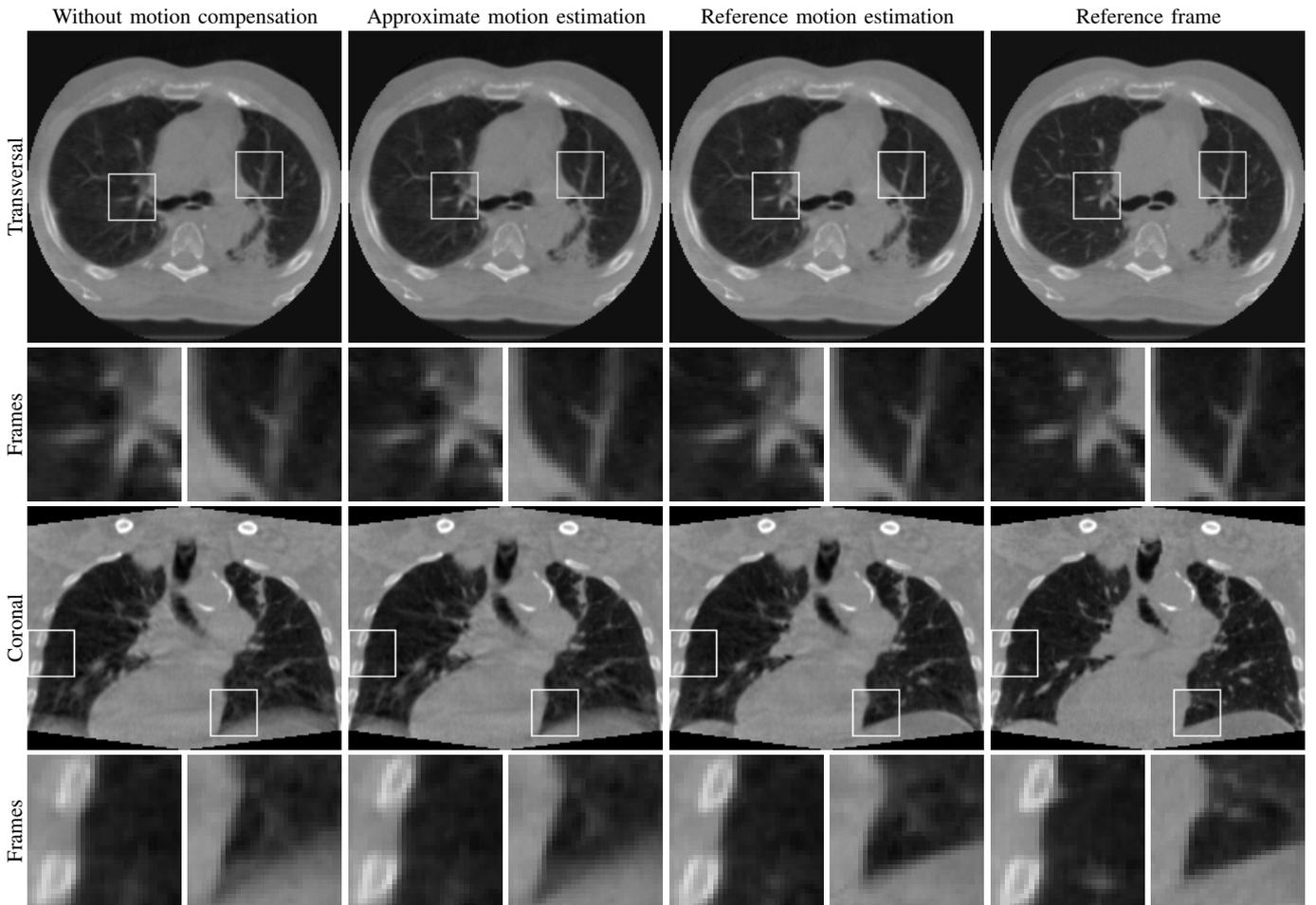


Figure 5. Results of image reconstructions for the two motion estimation experiments, compared to a static reference frame. Two close-up views of 32×32 pixels are selected per image and marked by the white frames. The border of the myocardium and the ribs contours get sharper. However, motion estimation from approximate reference projections did not succeed to suppress the blurring observed at the border of the respiratory diaphragm.

blurred in transversal views and the border of the diaphragm is very fuzzy too in the coronal view. The second column shows that the image gets sharper with motion estimation and segmentation using approximate reference projections. The third column contains slices of the reconstructed image when using reference projections from a known static image for motion estimation. The motion-compensated image is very close to the reference image shown in the last column.

V. CONCLUSIONS

This work proposes a practical technique to improve image quality when acquired data are corrupted by arbitrary patient motion. First an iterative reconstruction is performed to produce a set of reference projections. Then, the perceived motion is estimated in projection space by elastic image registration and segmented in image space by reconstructing the absolute differences between acquired and reference projections. Finally, a motion-compensated image is reconstructed by a slightly modified FDK implementation. Motion blur artifacts are locally reduced with motion compensation.

The method has been validated on experiments using a dynamic phantom reconstructed from clinical patient data on experiments. Results demonstrate a great potential to estimate and compensate breathing motion. Since the underlying motion model does not assume periodicity, our technique could capture arbitrary residual patient motion that could corrupt the data in breath-hold acquisitions.

ACKNOWLEDGMENT

The authors gratefully thank David Barber for sharing the Sheffield image registration toolkit (ShIRT) that was used for computing optical flows. They also thank Jens Wiegert for his advices about image reconstruction and Thomas Köhler for sharing the breathing phantom.

REFERENCES

- [1] B. Horn and B. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, no. 1, pp. 185–203, 1981.
- [2] D. Gilland, B. Mair, and J. Parker, "Motion estimation for cardiac emission tomography by optical flow methods," *Physics in Medicine and Biology*, vol. 53, no. 11, pp. 2991–3006, 2008.
- [3] D. Barber, E. Oubel, A. Frangi, and D. Hose, "Efficient computational fluid dynamics mesh generation by image registration," *Medical Image Analysis*, vol. 11, no. 6, pp. 648–662, 2007.
- [4] C. Schretter, C. Neukirchen, M. Bertram, and G. Rose, "Correction of some time-dependent deformations in parallel-beam computed tomography," *IEEE ISBI Conference*, pp. 764–767, 2008.
- [5] L. Desbat, S. Roux, and P. Grangeat, "Compensation of some time dependent deformations in tomography," *IEEE Trans. Medical Imaging*, vol. 26, no. 2, pp. 261–269, 2007.
- [6] D. Schäfer *et al.*, "Motion-compensated and gated cone beam filtered back-projection for 3-D rotational x-ray angiography," *IEEE Trans. Medical Imaging*, vol. 25, no. 7, pp. 898–906, 2006.
- [7] K. Taguchi and H. Kudo, "A simple motion tracking backprojection for a class of affine transformation," *SPIE Medical Imaging*, vol. 6913, 2008.
- [8] A. Andersen and A. Kak, "Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm," *Ultrasonic imaging*, vol. 6, no. 1, pp. 81–94, 1984.