Gradient-Based Differential Approach for 3-D Motion Compensation in Interventional 2-D/3-D Image Fusion

Jian Wang*†, Anja Borsdorf†, Benno Heigl†, Thomas Köhler*‡, Joachim Hornegger*‡
**Pattern Recognition Lab, Friedrich-Alexander-University Erlangen-Nuremberg*
Erlangen, Germany
† *Siemens Healthcare*
Forchheim, Germany
‡*Erlangen Graduate School in Advanced Optical Technologies (SAOT)*
Erlangen, Germany

***Abstract*—In interventional radiology, preoperative 3-D volumes can be fused with intra-operative 2-D fluoroscopic images. Since the accuracy is crucial to the clinical usability of image fusion, patient motion resulting in misalignments has to be corrected during the procedure. In this paper, a novel gradient-based differential approach is proposed to estimate the 3-D rigid motion from the 2-D tracking of contour points. The mathematical relationship between the 3-D differential motion and the 2-D motion is derived using the 3-D gradient, based on which a tracking-based motion compensation pipeline is introduced. Given the initial registration, the contour points are extracted and tracked along 2-D frames. The 3-D rigid motion is estimated using the iteratively re-weighted least square minimization to enhance the robustness. Our novel approach is evaluated on 10 datasets consisting of 1010 monoplane fluoroscopic images of a thorax phantom with 3-D rigid motion. Over all datasets, the maximum structure shift in the 2-D projection caused by the 3-D motion varies from 17.3 mm to 33.2 mm. Our approach reduces the 2-D structure shift to the range of 1.93 mm to 6.52 mm. For the most challenging longitudinal off-plane rotation, our approach achieves an average coverage of 79.9% regarding to the ground truth.**

***Keywords*-2-D/3-D image fusion; motion compensation; interventional radiology; gradient-based approach**

I. Introduction

In current clinical practice, interventional radiology becomes a standard routine for image-guided minimally invasive procedures, e. g. endovascular aneurysm coiling, balloon angioplasty and stenting. Interventional C-arm systems provide two-dimensional (2-D) fluoroscopic X-ray images for real-time guidance. Preoperative three-dimensional (3-D) volumes (e.g. X-ray computed tomography) can be fused onto the live 2-D X-ray images. 2-D/3-D image fusion provides complementary information from the 3-D volume that is not directly available in 2-D fluoroscopic images, e. g. preoperative planning information, depth information or vascular structures without contrast media.

The accuracy of image fusion is critical to the clinical usability, which leads to the topic of 2-D/3-D image registration [1]. In clinical practice, 2-D/3-D registration is performed at the starting point to ensure the overlay accuracy. However, the patient motion during the intervention must be corrected as well. In recent years, efforts have been made for cardiac and breathing motion compensation. Brost et al. [2] proposed a breathing motion compensation approach for electrophysiology (EP) procedures, where 2-D motion is estimated and extracted by catheter tracking. Hadida et al. [3] proposed an approach for 3-D motion compensation of coronary arteries, where a stochastic model of the cardiac cycle is used for the deformation estimation. However, there are still challenges for general patient movement compensation: 1) breathing/cardiac model-based methods are restricted to the model cycles and the patient movement usually leads to the failure of these approaches; 2) device-tracking-based approaches are restricted to certain procedures; 3) pure 2-D motion compensation is not fundamentally correct and estimation of 3-D motion from a monoplane system is challenging [2]. Currently, patient movement correction is done by manually triggering the registration procedure when misalignment is detected.

The image gradient contains important structural information, thus 2-D and 3-D gradient-based registration methods have been widely investigated [4], [5], [6], [7]. However, most of the methods are not designed for real-time motion compensation and it is usually computationally expensive to perform the registration procedure each frame. On the contrary, tracking-based approaches are investigated for real-time motion compensation [2], [8], [9]. However, these approaches are limited to 2-D motion estimation. The goal of our work is to build a tracking-based motion compensation framework, where 3-D motion is recovered from 2-D motion using more image information (e. g. gradient).

Recently, Wang et al. [10] proposed a depth-layer-based tracking approach for patient motion compensation. Since initial registration is usually available, the structures from 2-D and 3-D are well aligned. Instead of a whole volume rendering, the depth-layer images are rendered from different depth intervals. The depth information of 2-D feature points can be estimated by performing a patch-wise local similarity measure (e. g. normalized cross correlation) between the initial frame and the depth-layer images. By 2-D tracking of the extracted points, 3-D rigid motion is estimated from the point correspondences. However, this approach is not optimal for X-ray images (comparing to the optical images), where ideal point correspondences are difficult to obtain.

In this paper, we propose a tracking-based differential approach for 3-D motion compensation using image gradient. Since only small motion occurs between neighboring frames during continuous acquisition, the differential approximation of the 3-D rigid motion is employed. A mathematical relationship between the differential 3-D motion and 2-D motion is derived using the 3-D gradient, based on which a tracking-based motion compensation pipeline is developed. Under the initial registration, occluding contour points are initialized based on 3-D gradient analysis and 2-D/3-D selection criteria. By tracking of the contour points in 2-D X-ray images, the 3-D rigid motion is estimated frame by frame using an iteratively re-weighted least square estimation scheme.

The remainder of the paper is structured as follows: after a brief problem description (Sec. II-A), the mathematical derivation of the relationship between 2-D and 3-D differential motion is presented in Sec. II-B. Then, the tracking-based motion compensation pipeline is introduced in Sec. II-C. Experiments are explained and the evaluation is discussed in Sec. III. At last, Sec. IV concludes the paper with an outlook.

II. Method

A. Problem Description

As illustrated in Fig. 1(a), an interventional C-arm system can be modeled as a pinhole camera [11], where the camera center is the X-ray source and the imaging plane is the detector. The projection geometry is described by the projection matrix

$$\mathbf{P} = \mathbf{K}[\mathbf{R}_e|\mathbf{t}_e] \in \mathbb{R}^{3\times4} \ , \tag{1}$$

where $\mathbf{K} \in \mathbb{R}^{3\times3}$ is the camera matrix, $\mathbf{R}_e \in \mathbb{R}^{3\times3}$ and $\mathbf{t}_e \in \mathbb{R}^3$ are respectively rotation and translation of the camera center in the world coordinate system [12]. Without loss of generality, we use the camera coordinate system for derivation, where the projection matrix is simplified as

$$\mathbf{P}_e = \mathbf{K}[\mathbf{I}|\mathbf{0}] \ . \tag{2}$$

Given the projection parameters, the 3-D volume can be rendered as imaged from the X-ray focus. The image fusion is done by blending the resulting image to the live fluoroscopic image [11]. After the initial registration, the corresponding structures of the patient's anatomy are properly aligned in the fused image. Our task is to estimate the 3-D motion out of the fluoroscopic image sequence and apply it to the 3-D volume rendering, such that the alignment is maintained under the patient movement. Since bone structures are usually well observed in X-ray images and directly present the patient movement, they are employed as the anatomical target for motion estimation. Therefore, the target movement is modeled as 3-D rigid motion since bones almost stay rigid during the intervention.

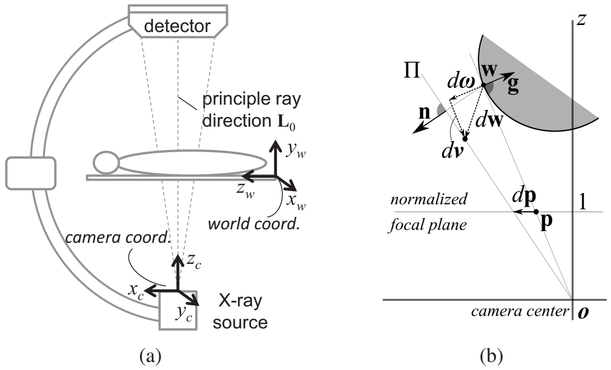


Figure 1. (a) Illustration of the projection model of a C-arm system and the coordinate systems; (b) Sketch for the mathematical relationship in gradient-based differential motion estimation.

B. Mathematical Model for Differential 3-D Motion given 2-D Motion and 3-D Gradient

During the continuous 2-D acquisition, the difference between two neighboring frames gives a good impression how the object moves in 3-D. However, a mathematical model is required to apply this impression for motion estimation. In this section, we derive the mathematical relationship between 2-D motion to 3-D differential motion using gradient information.

3-D points on occluding contours have strong gradient in the 3-D volume and also appear in the projection image as the border of a region having remarkably different attenuation values in relation to its neighborhood. This is the case, e. g. for bone structures surrounded by water-like environment or soft tissue. Thus, the occluding contour points are chosen as the target for our motion estimation approach.

In the X-ray image sequence, a movement of a point $\mathbf{x} \in \mathbb{R}^2$ can be determined only in the direction of the image gradient vector $\nabla I(\mathbf{x})$, where I is the 2-D image function of the projection image. Similarly, within a vicinity of a point $\mathbf{w} \in \mathbb{R}^3$ in 3-D, a small movement only causes a change of intensity values if the movement has a component in the direction of its 3-D gradient vector $\mathbf{g} = \nabla f(\mathbf{w}) \in \mathbb{R}^3$ (Fig 1(b)), which can be determined for each point $\mathbf{w}$ from the volume data described by the 3-D image function f. All movements orthogonal to $\mathbf{g}$ do not change the intensity values in the vicinity of $\mathbf{w}$.

In this case, the 2-D motion vector $d\mathbf{x}$ and 3-D gradient vector $\mathbf{g}$ are co-planar as both are related to the same occluding contour. Therefore, we assume that only movements in the direction of the 2-D and 3-D image gradients are observable, which is the intuition behind the derivation.

As illustrated in Fig. 1(b), a 2-D point $\mathbf{x}$ is normalized as $\mathbf{p} \in \mathbb{R}^3$ in homogeneous coordinate with respect to the focal length, such that the normalized point $\mathbf{p}$ is treated as a 3-D point lying on the plane $z = 1$ in the camera coordinate system. Under the convention in Eq. (2), given the image coordinate $\mathbf{x} = (u, v)$ of the projection of $\mathbf{w}$, the normalized image coordinate is formulated as the back-projection

$$\mathbf{p} = \left| \mathbf{K}^{-1} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \right|_H \quad , \tag{3}$$

where $|\cdot|_H$ denotes the homogeneous normalization, such that the last element of $\mathbf{p}$ is normalized as 1.

Under the small motion assumption, the 3-D motion vector $d\mathbf{w} \in \mathbb{R}^3$ of the 3-D point $\mathbf{w}$ between two neighboring frames is represented by the differential form as

$$d\mathbf{w} = d\boldsymbol{\omega} \times \mathbf{w} + d\boldsymbol{v} \quad , \tag{4}$$

where $d\boldsymbol{\omega} \in \mathbb{R}^3$ denotes the differential rotation around the origin[1] and $d\boldsymbol{v} \in \mathbb{R}^3$ denotes the differential translation. As shown in Fig 1(b), $d\mathbf{w}$, $\boldsymbol{\omega}$ and $\boldsymbol{v}$ are drawn as dotted lines to show that they can lie outside of the drawing plane described by $\mathbf{o}$, $\mathbf{w}$, and the 3-D gradient vector $\mathbf{g}$.

Based on the above considerations, the 2-D and 3-D movements are now related to each other by a plane Π containing the target point of the movement $\mathbf{w} + d\mathbf{w}$, the coordinate origin $\mathbf{o}$, and the moved image point $\mathbf{p} + d\mathbf{p}$. The normal $\mathbf{n} \in \mathbb{R}^3$ of this plane can be computed by

$$\mathbf{n} = \frac{(\mathbf{w} \times \mathbf{g}) \times (\mathbf{p} + d\mathbf{p})}{\|(\mathbf{w} \times \mathbf{g}) \times (\mathbf{p} + d\mathbf{p})\|} \quad . \tag{5}$$

Now, the plane Π is defined by the normal $\mathbf{n}$ and the coordinate origin $\mathbf{o}$. Therefore, a point $\mathbf{a}$ lies on Π if $\mathbf{n}^{\mathrm{T}}(\mathbf{a} - \mathbf{o}) = 0$, i.e. $\mathbf{n}^{\mathrm{T}}\mathbf{a} = 0$. Therefore, the following equation holds for $\mathbf{w} + d\mathbf{w}$ as

$$\mathbf{n}^{\mathrm{T}}(\mathbf{w} + d\mathbf{w}) = 0 \quad . \tag{6}$$

By substituting $d\mathbf{w}$ in Eq. (4) to Eq. (6), it yields

$$\mathbf{n}^{\mathrm{T}}(d\boldsymbol{\omega} \times \mathbf{w}) + \mathbf{n}^{\mathrm{T}} d\boldsymbol{v} + \mathbf{n}^{\mathrm{T}}\mathbf{w} = 0 \quad . \tag{7}$$

After reformulating the equation in the way that the unknown motion vector is on one side, the linear constraint between the differential motion and a contour point (2-D/3-D positions, i.e. $\mathbf{p}$, $\mathbf{w}$ and gradient $\mathbf{g}$) can be formulated as

$$\begin{pmatrix} \mathbf{n} \times \mathbf{w} \\ -\mathbf{n} \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} d\boldsymbol{\omega} \\ d\boldsymbol{v} \end{pmatrix} = \mathbf{n}^{\mathrm{T}}\mathbf{w} \quad . \tag{8}$$

A system of linear equations can be assembled by combining a set of occluding contour points $\{\mathbf{w}_i\}$ with their normalized projections $\{\mathbf{p}_i\}$ $(i = 1, ..., N)$ as

$$\mathbf{A}\delta\mathbf{v} = \mathbf{b} \quad , \tag{9}$$

where $\delta\mathbf{v} = \begin{pmatrix} d\boldsymbol{\omega}^{\mathrm{T}} & d\boldsymbol{v}^{\mathrm{T}} \end{pmatrix}^{\mathrm{T}} \in \mathbb{R}^6$ denotes the differential motion vector, $\mathbf{a}_i^{\mathrm{T}} = \begin{pmatrix} \mathbf{n}_i \times \mathbf{w}_i \\ -\mathbf{n}_i \end{pmatrix}^{\mathrm{T}} \in \mathbb{R}^6$ is the i-th row of matrix $\mathbf{A} \in \mathbb{R}^{N \times 6}$, $b_i = \mathbf{n}_i^{\mathrm{T}}\mathbf{w}_i$ is the i-th entry of vector $\mathbf{b} \in \mathbb{R}^N$ and N is the number of points. At least $N = 6$ points are acquired for the closed-form solution of the motion vector $\delta\mathbf{v}$ using the pseudo-inverse as

$$\widehat{\delta\mathbf{v}} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1} \mathbf{A}^{\mathrm{T}}\mathbf{b} \quad . \tag{10}$$

Given the differential motion vector $\delta\mathbf{v}$, the rotation matrix $\delta\mathbf{R} \in \mathbb{R}^{3\times3}$ can be calculated from $d\boldsymbol{\omega}$ as

$$\delta\mathbf{R} = \cos\theta\mathbf{I} + (1 - \cos\theta)\mathbf{r}\mathbf{r}^{\mathrm{T}} + \sin\theta[\mathbf{r}]_\times \quad , \tag{11}$$

where $\theta = \|d\boldsymbol{\omega}\|$, $\mathbf{r} = d\boldsymbol{\omega}/\|d\boldsymbol{\omega}\|$ and

$$[\mathbf{r}]_\times = \begin{bmatrix} 0 & -r_z & r_y \\ r_z & 0 & -r_x \\ -r_y & r_x & 0 \end{bmatrix} \quad . \tag{12}$$

The 3-D rigid motion $\mathbf{T}_k \in \mathbb{R}^{4\times4}$ is then updated from frame k to $k+1$ as

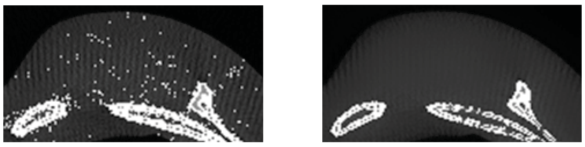
$$\mathbf{T}_{k+1} = \begin{bmatrix} \delta\mathbf{R} & d\boldsymbol{v} \\ \mathbf{0} & 1 \end{bmatrix} \mathbf{T}_k \quad . \tag{13}$$

C. Tracking-Based Differential Motion Compensation

Section II-A presents the mathematical relationship between the 3-D differential motion and 2-D motion given the 3-D gradient. Based on this model, a tracking-based differential motion compensation pipeline is developed. It consists of three main steps: 1) 3-D volume gradient analysis for pre-selection of occluding contour points; 2) Occluding contour point extraction with 2-D/3-D correspondences; 3) 2-D tracking and motion estimation. Step 1) and 2) are performed once as initialization. Then, step 3) is performed for each frame. Details are described as follows.

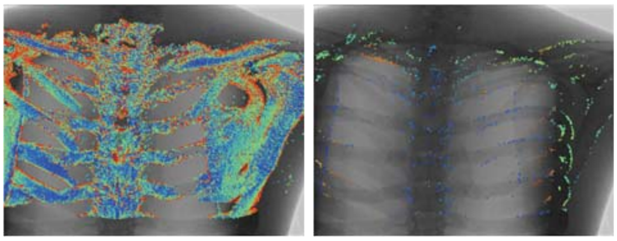
1) 3-D gradient analysis for pre-selection of candidates: Since occluding contour points have strong gradient magnitude, volume gradient analysis is performed. He et al. [13] recently proposed the guided image filter (GIF), which is an effective edge and gradient preserving image filter. We extend the GIF to 3-D for preprocessing of the volume to reduce noise and artifacts. First, window thresholding is applied to the intensity values in the way, that only the voxels representing the structures of interest (e.g. bone structures) remain for analysis. The 3-D gradient is computed for all remaining voxels using the error function of third degree (3EF) suggested by Rheingans and Ebert in [14]. Thresholding on 3-D gradient magnitude is also performed to leave out the weak contour points. The result is a set of voxels that potentially lie on occluding contours depending on the viewing direction. Figure 2 shows the

[1] The differential rotation is the estimate from the Rodrigues' rotation formula $\mathbf{w}' = \mathbf{w} + \sin\theta(\boldsymbol{\omega} \times \mathbf{w}) + (1 - \cos\theta)\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{w})$



(a) (b)

Figure 2. Pre-selected candidates (white dots) in one slice of (a) the original volume and (b) the filtered volume. The volume filtering reduces the number of wrong candidates due to noises and artifacts comparing to the original volume.



(a) (b)

Figure 3. (a) shows the pre-selection after volume gradient analysis, where the color encodes the perpendicularity between the 3-D gradient and the viewing direction (red: high perpendicularity, blue: low perpendicularity and green: between); (b) shows the final selection of occluding contour points, where the color encodes depth (red: near, blue: far and green: in between).

pre-selection results of an example volume slice. Since artifacts and noise disturb the selection, the volume filtering significantly reduces wrong candidates (Fig. 2(b)) compared to the results on the original volume (Fig. 2(a)). An example of the projected candidates is shown in Fig. 3(a).

2) Occluding contour extraction with 2-D/3-D correspondence: After the pre-selection of feature candidates, occluding contour points are firstly selected in 3-D using the projection parameters of the initial registration. The perpendicularity α between the 3-D gradient $\mathbf{g}$ and the viewing direction $\mathbf{w}$ (in camera coordinate system) of each candidate is checked by $\alpha = \arccos\left(\left|\left(\mathbf{g}\cdot\mathbf{w}\right)/\left(\|\mathbf{g}\|\cdot\|\mathbf{w}\|\right)\right|\right)$. In 3-D, the points with the perpendicularity $\alpha \geq \alpha_T$ are chosen to be the occluding contour points, where α_T is the threshold controlling contour thickness.

After the 3-D perpendicularity check, the set of 3-D points $\{\mathbf{w}_i\}$ are projected onto the 2-D fluoroscopic image using the projection geometry from the initial registration. Two criteria are then applied to further selection of the occluding contour points from 2-D image: 1) 2-D gradient magnitude and 2) depth similarity maps. Occluding contour points with high 2-D gradient magnitude are chosen, which are relatively easy to track in the following frames; Similarity maps are generated between the initial frame and the depth-layer images rendered from different depth intervals as described in [10], where patch-wise local normalized cross correlation is employed as the similarity measure. Feature points within the corresponding depth interval with high similarity are chosen to neglect the points that are not observed as the corresponding structure in the 2-D projection image. After the feature point selection both in 3-D and 2-D, a set of occluding contour points with 2-D/3-D correspondence $\{\mathbf{w}_i, \mathbf{p}_i\}_{\text{sel}}$ are then used for motion estimation. Figure 3(b) shows the contour points selected from the candidates shown in Fig. 3(a).

3) 2-D tracking and robust motion estimation: Kanade-Lucas-Tomasi (KLT)-based optical flow [15] is used to measure the 2-D motion fields of the feature points along frames. To achieve a better tracking performance, the gradient magnitude map of the k-th filtered X-ray frame is used, i.e.$\|\nabla h(I_k)\|$, where h is the 2-D GIF [13].

After measuring the 2-D offsets, we have all the necessary information $\{\mathbf{w}_i, \mathbf{g}_i, \mathbf{p}_i, d\mathbf{p}_i\}_{\text{sel}}$ of the feature points to set up the linear system of equations according to Eq. (9). A closed-form solution is available as shown in Eq. (10) when more than 6 points are chosen. However, usually over 1,000 candidates are available and each candidate may have a different confidence. Therefore, instead of using direct pseudo-inverse (Eq. (10)), the linear equation system (Eq. (9)) is converted to a weighted least square problem as

$$\widehat{\delta\mathbf{v}} = \arg\min_{\delta\mathbf{v}} \sum_{i}^{N} \beta_i \left(\mathbf{a}_i^{\mathrm{T}}\delta\mathbf{v} - b_i\right)^2 \quad . \tag{14}$$

The above scheme allows to assign the confidence of the i-th feature point as its weight β_i. The least square optimization can be solved using the Levenberg-Marquardt optimization [12]. For the purpose of enhancing the robustness, an outlier-aware iteratively re-weighted estimation scheme [16], [17] is employed. For the i-th feature point, the confidence β_i is formulated as

$$\beta_i = \beta_{z,i} \cdot \beta_{r,i} \quad , \tag{15}$$

where $\beta_{z,i}$ is the observation confidence and $\beta_{r,i}$ is the estimation confidence.

The observation confidence $\beta_{z,i}$ is determined by two observation measurements as

$$\beta_{z,i} = \mathcal{N}_{(0.5,1)}\left(r_{\alpha,i}/r_{klt,i}\right) \quad , \tag{16}$$

where $r_{klt,i}$ is the tracking residual from KLT tracker and $r_{\alpha,i} = 1 - |\cos\alpha_i|$ is the perpendicularity term in 3-D. The normalization operation

$$\mathcal{N}_{(a,b)}\left(\beta_i\right) = a + (b-a)\frac{\beta_i}{\max(\boldsymbol{\beta})} \tag{17}$$

is the normalization of all β_i to the range of $[a,b]$ to avoid the dramatic influence of each factor on the weights.

Furthermore, the estimation confidence $\beta_{r,i}$ is used to suppress outliers. It is initialized as $\beta_{r,i}^{(0)} = 1$, where $i = 1, ..., N$. For further iterations, $\beta_{r,i}^{(t)}$ is determined by the corresponding residual term as

$$\beta_{r,i}^{(t)} = \mathcal{N}_{(0.5,1)}\left(1/r(\delta\mathbf{v}^{(t-1)})\right) \quad , \tag{18}$$

1. Initialize $t = 0$ and the estimation confidence $\beta_{r,i}^{(0)} = 1$ ($i = 1, ..., N$);
2. Determine the observation confidence $\beta_{z,i}$ from measurements using Eq. (16);
3. Determine the residual term $r(\delta\mathbf{v}^{(t-1)})$ using Eq. (19);
4. Determine the confidence $\beta_i^{(t)}$ with $\beta_{r,i}^{(t)}$ using Eq. (18) and $\beta_{z,i}$;
5. Solve for $\delta\mathbf{v}^{(t)}$ according to Eq. (14) using Levenberg-Marquardt optimization;
6. Set $t \leftarrow t + 1$ and repeat step 3 to 5 until convergence.

Table I
THE ITERATIVELY RE-WEIGTHED LEAST SQUARES(IRLS) SCHEME FOR MOTION ESTIMATION

where the residual term at t-th iteration is

$$r(\delta\mathbf{v}^{(t-1)}) = \mathbf{a}_i^{\mathrm{T}}\widehat{\delta\mathbf{v}} - b_i \qquad (19)$$

For each frame k, the motion estimation is done using the iteratively re-weigthed least squares (IRLS) (Tab. I).

III. EXPERIMENTS AND EVALUATION

A. Experimental Materials

For experimental evaluation, an interventional C-arm system was used for image acquisition. Figure 4 shows the experimental imaging environment. A thorax phantom was used to acquire both 3-D volume and 2-D image sequence (fluoroscopy) by the calibrated C-arm system. Initial 2-D/3-D alignment was accurate, since no position change raised between the 3-D acquisition and the initial frames of 2-D fluoroscopic images. The motion of the thorax phantom was triggered manually by pulling a belt that was attached to the phantom during the 2-D acquisition. The 3-D optical camera (OptoTrak motion capture system) was employed to capture the motion of the markers attached on the phantom, which is used as the ground truth for evaluation. The markers were attached on the borders of the phantom, such that no markers were visible in the X-ray image sequences. Ten sequences under 3-D rigid motion consisting of in-/off-plane translation/rotation were acquired.

B. Ground-Truth Motion Acquisition

The OptoTrak motion capture system was employed to acquire the ground-truth motion (Fig. 4). To make the OptoTrak data available for evaluation, two main issues are considered: 1) transforming the OptoTrak motion sequence $\{\mathbf{T}_{k'}^{opt}\}$ from the OptoTrak coordinate system to the world coordinate system (i.e. determining the transformation $\mathbf{T}_{opt2wld}$); 2) synchronizing the OptoTrak motion sequence with the 2-D fluoroscopic acquisition.

As shown in Fig. 4, the markers are tracked under the OptoTrak coordinate system as $(x_{opt}, y_{opt}, z_{opt})$. The principle of determining the transformation $\mathbf{T}_{opt2wld}$ is to determine the position of the markers in the world coordinate system (x_w, y_w, z_w). Since the markers are also included in the 3-D volume, the markers can be segmented from the 3-D volume. Using the projection parameters of the calibrated C-arm system, the position of the markers in the world coordinate system are determined. Then, the rigid transformation matrix is directly estimated from the 3-D point correspondences under two coordinate systems. The closed-form solution of absolute orientation [18] is employed to compute the orientation (rotation) and position (translation). Due to the limited field of view, multiple 3-D scans were taken under different table positions to acquire the positions of more markers.

The synchronization of the OptoTrak data to the fluoroscopy image sequence was done manually. The frame rate of OptoTrak acquisition was 100 fps, compared to 10 fps of the fluoroscopy. There are two criteria used for the synchronization: 1) the 2-D/3-D alignment under the synchronized ground-truth motion along all frames and 2) the estimated motion curves should be well aligned to the ground-truth motion, especially the steep shifts should be synchronized. The motion curves are later explained in the results (Fig. 5).

C. Motion Compensation Results

We assess the quality of motion compensation as follows: 1) comparing the 3-D motion with the corresponding ground-truth motion and 2) comparing the projection error of the structures of interest under the estimated motion and under the ground-truth motion. The depth-layer-based (DL) motion compensation [10] is used as our baseline approach.

1) The Estimated 3-D Motion: 3-D rigid motion can be decomposed into 6 components, i.e. 3 rotation angles (R_x, R_y and R_z) and 3 translations (t_x, t_y and t_z) about the corresponding axes. Thus, we compare the estimated motion components using both our new differential approach and the baseline approach with the ground truth in the world coordinate system (Fig. 4).

The motion curves: Figure. 5 shows the comparison between the estimated motion curves and the ground-truth motion curves of Seq. 6 and 10. The longitudinal rotation can be considered as the rolling motion of the patient, which is described by R_z together with t_x (shift of the rotation center). This off-plane motion is difficult to estimate for monoplanar approaches. The longitudinal rotation is the major motion in both of the sequences. Seq. 10 also has a significant in-plane rotation R_y. Comparing to the baseline approach, our approach gives a significant improvement in the longitudinal rotation.

The statistics of the estimation error: the means and standard deviations of the estimation errors over all frames of each sequence are calculated, i.e. $\epsilon(R_x)$, $\epsilon(R_y)$ and $\epsilon(R_z)$ in rotation and $\epsilon(t_x)$, $\epsilon(t_y)$ and $\epsilon(t_z)$ in translation. Table. II shows the estimation error statics. For each sequence, the maximum magnitudes of the corresponding ground-truth motion components are given as reference to

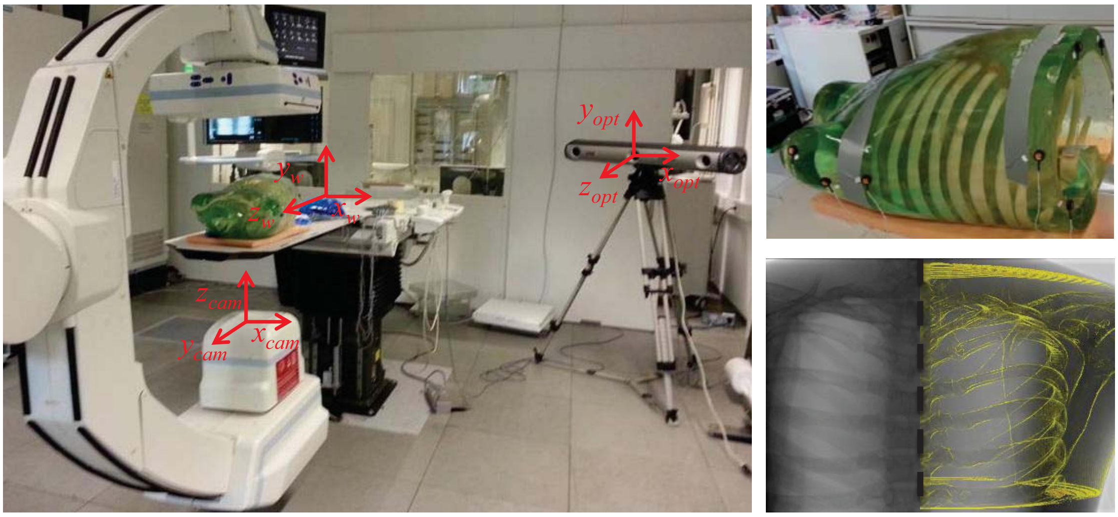
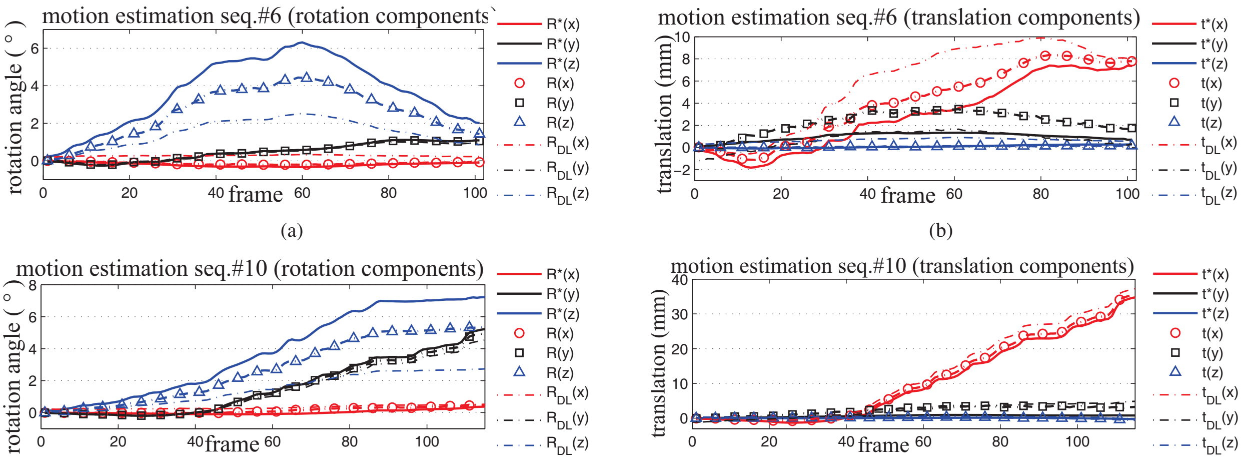


Figure 4. An interventional C-arm system is used to acquire both 2-D image sequence and 3-D C-arm CT volume of a thorax phantom. Motion of the thorax phantom is triggered manually. The OptoTrak motion capture system is used to acquire the ground truth.



(a)

(b)

(c)

(d)

Figure 5. The motion curves. The rotation and translation components are plotted for Seq. 6 in (a) and (b), and Seq. 10 in (c) and (d). The ground-truth curves are the solid plots (($R^*(x)$, $R^*(y)$, $^*R(z)$ in rotation and $t^*(x)$, $t^*(y)$, $t^*(z)$ in translation)). Our results are dotted plots with markers, and the results of the baseline approach (Wang et al. [10]) are the dotted plots without markers.

the estimation error. We focus on the estimation of the major motion components (bold blue in Tab. II), which have significant maximum magnitudes. To describe how well the motion is estimated, the *motion recovery rate* is defined as $r(m) = 1 - \epsilon(m)/\max|m^*|$, where $\epsilon(m)$ is the estimation error of the motion component m and $\max|m^*|$ is the maximum ground-truth magnitude. Our approach achieves a mean recovery rate of $97.4 \pm 0.8\%$, $95.4 \pm 2.9\%$ and 96.7% (Seq. 5) for the major inplane rotation R_y and translations t_x and t_z. For the longitudinal rotation, our approach achieves a mean recovery of $79.9 \pm 3.5\%$. As a general problem of monoplanar approaches, translations in depth (t_y) appear the least stable, which does not influence too much of the 2-D/3-D fusion accuracy.

Both Fig. 5 and Tab. II show our approach is capable of estimating the 3-D rigid motion from tracking 2-D occluding contour points.

2) The 2-D Projection Error: Another perspective of assessing the motion compensation is to evaluate the misalignment observed in the fused view. For a 3-D point, the distance between the projection under the estimated motion and the ground-truth motion is noted as the *projection error*. Given a set of points, the mean and the standard deviation of the projection errors represent the amount and the distribution of misalignment. The pre-selected 3-D points (Fig. 2) are used for evaluation, because they well

seq.	# fr.	max $\|R_x^*\|(^\circ)$	$\epsilon(R_x)(^\circ)$	max $\|R_y^*\|(^\circ)$	$\epsilon(R_y)(^\circ)$	max $\|R_z^*\|(^\circ)$	$\epsilon(R_z)(^\circ)$
1	33	0.18	0.04 ± 0.02	0.97	0.05 ± 0.03	**4.48**	**0.93 ± 0.43**
2	93	0.80	0.43 ± 0.26	0.52	0.05 ± 0.03	**11.93**	**2.82 ± 1.55**
3	111	0.63	0.40 ± 0.18	0.38	0.01 ± 0.01	**10.7**	**2.59 ± 1.04**
4	111	0.27	0.12 ± 0.11	0.56	0.03 ± 0.02	**8.31**	**1.51 ± 0.94**
5*	110	0.06	0.07 ± 0.03	**10.0**	**0.37 ± 0.16**	0.04	0.18 ± 0.08
6	101	0.338	0.0682 ± 0.036	1.119	0.0245 ± 0.0123	**6.308**	1.081 ± 0.48
7	105	0.23	0.36 ± 0.26	**4.43**	**0.08 ± 0.07**	**6.82**	**1.54 ± 0.90**
8	117	0.26	0.32 ± 0.18	1.47	0.04 ± 0.03	**7.93**	**1.32 ± 0.83**
9	114	0.18	0.10 ± 0.06	**4.57**	**0.15 ± 0.08**	**8.13**	**1.89 ± 0.92**
10	115	0.371	0.12 ± 0.078	**5.224**	0.0897 ± 0.081	**7.225**	1.064 ± 0.66

seq.	# fr.	max $\|t_x^*\|$(mm)	$\epsilon(t_x)$(mm)	max $\|t_y^*\|$(mm)	$\epsilon(t_y)$(mm)	max $\|t_z^*\|$(mm)	$\epsilon(t_z)$(mm)
1	33	2.76	1.13 ± 0.67	0.99	2.27 ± 1.66	0.16	0.09 ± 0.07
2	93	4.58	2.96 ± 1.63	0.71	2.06 ± 1.01	0.61	0.64 ± 0.39
3	111	4.19	3.11 ± 1.32	0.66	3.41 ± 1.37	0.65	0.53 ± 0.24
4	111	5.72	2.12 ± 1.26	1.40	1.39 ± 0.56	0.94	0.16 ± 0.13
5*	110	**69.3**	**2.29 ± 0.93**	0.45	0.91 ± 0.37	**6.31**	**0.22 ± 0.10**
6	101	7.414	1.159 ± 0.49	1.341	1.365 ± 0.59	0.271	0.0629 ± 0.035
7	105	**30.0**	**2.62 ± 1.36**	0.58	1.27 ± 1.06	0.44	0.59 ± 0.32
8	117	9.81	1.68 ± 0.94	1.24	1.13 ± 0.69	0.15	0.63 ± 0.36
9	114	**30.3**	**1.17 ± 0.57**	1.69	3.56 ± 1.66	0.11	0.16 ± 0.14
10	115	**34.763**	0.802 ± 0.38	0.96	1.709 ± 0.91	0.247	0.107 ± 0.059

Table II

THE MOTION ESTIMATION ERRORS: FOR EACH SEQUENCE, THE TOTAL NUMBER OF FRAMES IS SHOWN IN COLUMN "# FR.". THE max $|\cdot|$ COLUMN SHOWS THE MAXIMUM AMOUNT OF THE GROUND TRUTH MOTION COMPONENT, WHERE "*" MARKS THE GROUND TRUTH. $\epsilon(\cdot)$ SHOWS THE MEAN AND STANDARD DEVIATION OF THE ESTIMATION ERROR OVER ALL FRAMES OF EACH CASE. THE MAJOR MOTION COMPONENTS ARE BOLD BLUE. SEQ. 5 IS HIGHLIGHTED AS THE ONLY IN-PLANE MOTION CASE.

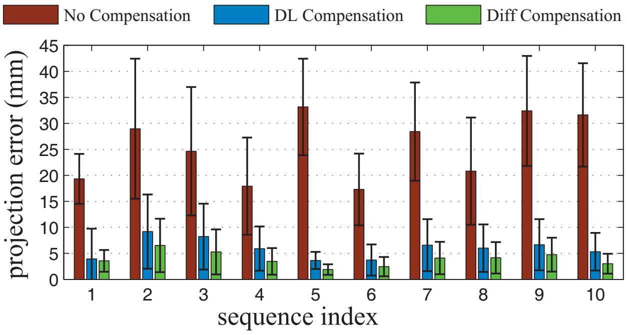
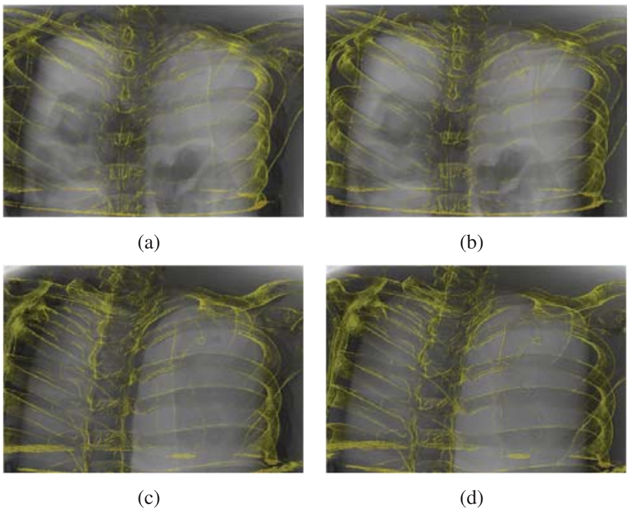


Figure 6. The projection error using our approach ("Diff Compensation") and the baseline approach ("DL compensation") compared to the maximum projection shifts ("No Compensation").

present the structures of interest. As reference, the distance between the projection under the ground-truth motion and the original projection is noted as the *projection shift* caused by the motion. For each sequence, we select the frame with the largest projection shift for evaluating the motion compensation capability of our approach.

Figure 6 shows the maximum projection shifts and the projection errors using our approach and the baseline approach. The performance of our approach outmatches the baseline approach in all cases. Except for Seq. 2 and 3, we have kept all the other (80%) cases below the failure threshold (5 mm) used in [19] along all frames. Figure 7 also directly shows two example frames without and with our motion compensation.


(a)

(b)

(c)

(d)

Figure 7. (a) and (b) show respectively the 2-D/3-D fusion of Seq. 4 at frame 65 (maximum motion) without and with motion compensation. Bags filled with contrast agent were put in the phantom during fluoroscopy, which were not in the 3-D scan. (c) and (d) show respectively the 2-D/3-D fusion of Seq. 7 at frame 50 without and with motion compensation.

IV. CONCLUSION AND OUTLOOK

In this paper, a novel gradient-based differential approach is proposed, where 3-D rigid motion is estimated from 2-D tracking for motion compensation in 2-D/3-D image fusion. The occluding contour points are used for tracking, and a mathematical relationship from 2-D motion to 3-D differential motion is derived by considering the 3-D gradient. An iterative re-weighted least square minimization scheme is used for robust motion estimation. The evaluation is performed on 10 image sequences with 1010 monoplane fluoroscopic images. The ground-truth motion was acquired by using an OptoTrak motion capture system. The results show that our approach is capable of estimating 3-D rigid motion by 2-D tracking. Our approach achieves a mean recovery of above 95% for all major in-plane motions and a mean recovery of 79.9% for the longitudinal rotation. Over all sequences, the maximum projection shift due to the 3-D motion is from 17.3 mm to 33.2 mm. Our approach reduces the 2-D structure shift to the range from 1.93 mm to 6.52 mm. The mean projection error remains under 5 mm for 8 out of 10 sequences. According to [19], our approach manages to keep 80% of the cases below the failure threshold of 5 mm.

In the current implementation, the initialization of the contour points is done only once and used for all the frames. As future work, a re-initialization strategy of the contour points will be developed for higher accuracy and robustness.

Acknowledgment

This work is supported by Siemens Healthcare, Forchheim, Germany. The authors gratefully acknowledge the funding of Siemens Healthcare and the Erlangen Graduate School in Advanced Optical Technologies (SAOT).

Disclaimer

The concepts and information presented in this paper are based on research and are not commercially available.

References

[1] P. Markelj, D. Tomaževič, B. Likar, and F. Pernuš, "A review of 3D/2D registration methods for image-guided interventions," *Med Image Anal*, vol. 16, no. 3, pp. 642 – 661, April 2012.

[2] A. Brost, R. Liao, N. Strobel, and J. Hornegger, "Respiratory motion compensation by model-based catheter tracking during EP procedures," *Medical Image Analysis*, vol. 14, no. 5, pp. 695–706, 2010.

[3] J. Hadida, C. Desrosiers, and L. Duong, "Stochastic 3D motion compensation of coronary arteries from monoplane angiograms," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*. Springer, 2012, pp. 651–658.

[4] H. Livyatan, Z. Yaniv, and L. Joskowicz, "Gradient-based 2-D/3-D rigid registration of fluoroscopic X-ray to CT," *Medical Imaging, IEEE Transactions on*, vol. 22, no. 11, pp. 1395–1406, 2003.

[5] W. Wein, B. Röper, and N. Navab, "2D/3D registration based on volume gradients," in *Medical Imaging*. International Society for Optics and Photonics, 2005, pp. 144–150.

[6] P. Markelj, D. Tomazevic, F. Pernus, and B. Likar, "Robust gradient-based 3-D/2-D registration of CT and MR to X-ray images," *Medical Imaging, IEEE Transactions on*, vol. 27, no. 12, pp. 1704–1714, 2008.

[7] Ž. Špiclin, B. Likar, and F. Pernuš, "Fast and robust 3D to 2D image registration by backprojection of gradient covariances," in *Biomedical Image Registration*. Springer, 2014, pp. 124–133.

[8] P. Wang, P. Marcus, T. Chen, and D. Comaniciu, "Using needle detection and tracking for motion compensation in abdominal interventions," in *Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium on*. IEEE, 2010, pp. 612–615.

[9] Y. Cao and P. Wang, "An adaptive method of tracking anatomical curves in X-ray sequences," in *Proceedings of the 15th International Conference on Medical Image Computing and Computer-Assisted Intervention-Volume Part I*. Springer-Verlag, 2012, pp. 173–180.

[10] J. Wang, A. Borsdorf, and J. Hornegger, "Depth-layer based patient motion compensation for the overlay of 3D volumes onto X-ray sequences," in *Proceedings Bildverarbeitung für die Medizin 2013*, 2013, pp. 128–133.

[11] S. Rossitti and M. Pfister, "3D road-mapping in the endovascular treatment of cerebral aneurysms and arteriovenous malformations," *Interventional Neuroradiology*, vol. 15, no. 3, p. 283, 2009.

[12] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge Univsersity Press, 2003.

[13] K. He, J. Sun, and X. Tang, "Guided image filtering," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 6, pp. 1397–1409, 2013.

[14] P. Rheingans and D. Ebert, "Volume illustration: Nonphotorealistic rendering of volume models," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 7, no. 3, pp. 253–264, 2001.

[15] J.-Y. Bouguet, "Pyramidal implementation of the affine Lucas Kanade feature tracker description of the algorithm," *Intel Corporation*, vol. 2, p. 3, 2001.

[16] J. A. Scales and A. Gersztenkorn, "Robust methods in inverse theory," *Inverse problems*, vol. 4, no. 4, p. 1071, 1988.

[17] T. Köhler, S. Haase, S. Bauer, J. Wasza, T. Kilgus, L. Maier-Hein, H. Feußner, and J. Hornegger, "Outlier detection for multi-sensor super-resolution in hybrid 3d endoscopy," in *Bildverarbeitung für die Medizin 2014*. Springer, 2014, pp. 84–89.

[18] B. K. Horn, "Closed-form solution of absolute orientation using unit quaternions," *JOSA A*, vol. 4, no. 4, pp. 629–642, 1987.

[19] C. Gendrin, P. Markelj, S. A. Pawiro, J. Spoerk, C. Bloch, C. Weber, M. Figl, H. Bergmann, W. Birkfellner, B. Likar *et al.*, "Validation for 2D/3D registration II: The comparison of intensity-and gradient-based merit functions using a new gold standard data set," *Medical physics*, vol. 38, no. 3, pp. 1491–1502, 2011.