

Dynamic 2-D/3-D Rigid Registration Framework Using Point-To-Plane Correspondence Model

Jian Wang, Roman Schaffert, Anja Borsdorf, Benno Heigl, Xiaolin Huang,
Joachim Hornegger, and Andreas Maier

Abstract—In image-guided interventional procedures, live 2-D X-ray images can be augmented with preoperative 3-D computed tomography or MRI images to provide planning landmarks and enhanced spatial perception. An accurate alignment between the 3-D and 2-D images is a prerequisite for fusion applications. This paper presents a dynamic rigid 2-D/3-D registration framework, which measures the local 3-D-to-2-D misalignment and efficiently constrains the update of both planar and non-planar 3-D rigid transformations using a novel point-to-plane correspondence model. In the simulation evaluation, the proposed method achieved a mean 3-D accuracy of 0.07 mm for the head phantom and 0.05 mm for the thorax phantom using single-view X-ray images. In the evaluation on dynamic motion compensation, our method significantly increases the accuracy comparing with the baseline method. The proposed method is also evaluated on a publicly-available clinical angiogram data set with “gold-standard” registrations. The proposed method achieved a mean 3-D accuracy below 0.8 mm and a mean 2-D accuracy below 0.3 mm using single-view X-ray images. It outperformed the state-of-the-art methods in both accuracy and robustness in single-view registration. The proposed method is intuitive, generic, and suitable for both initial and dynamic registration scenarios.

Index Terms—Rigid 2-D/3-D registration, dynamic registration, point-to-plane correspondence model.

I. INTRODUCTION

IN CURRENT clinical practice, interventional radiology becomes a standard routine for image-guided procedures, e.g. endovascular aneurysm coiling and stenting, catheter embolization and needle guidance. Interventional C-arm systems are commonly used to provide live streams of X-ray

images, known as fluoroscopy. In many scenarios, the 2-D fluoroscopic images can be augmented via fusion with 3-D CT or MRI images. The 2-D/3-D image fusion provides additional information from preoperative planning (e.g. needle guidance), pre-contrasted 3-D vascular structures (e.g. 3-D DSA) and enhanced spatial perception [1]. It helps to reduce the radiation dose, to lessen the injected contrast agent, and to shorten the procedure time [2]. An accurate alignment between 2-D and 3-D images is a prerequisite for fusion applications, which leads to the topics of C-arm calibration and image registration.

The C-arm calibration endorses the interventional C-arm system in diverse fields of 3-D imaging, i.e. to acquire 3-D computed tomography (C-arm CT) [3], [4] and to set up 2-D/3-D image fusion [5]. In this work, we assume that accurate calibration is readily available and therefore do not further address this issue.

Thorough reviews on recent 2-D/3-D registration techniques are found in [6] and [7]. Our work is in the scope of rigid 2-D/3-D registration. Although this topic has been widely discussed over the last decades, achieving high accuracy and robustness still remains challenging. Conventional approaches commonly use the digitally reconstructed radiographs (DRR), e.g. [8], where the optimization problem is complicated due to the loss of depth information. Alternatively, back-projection-based and reconstruction-based approaches have been investigated [9], [10], where the similarity is measured in 3-D space using back-projected or reconstructed image gradients. However, these approaches still require multiple-view X-ray images to achieve high 3-D accuracy.

For single-view registration, the accuracy in depth is generally considered as an ill-posed problem [11]. However, interventional fluoroscopic imaging is usually limited to a single view, since mono-planar C-arm systems are more widespread in clinical practice. In recent investigations on single-view 2-D/3-D registration [11]–[13], it has been shown that depth information and image gradients play important roles. Špiclin et al. [11] measured the similarity between 3-D and 2-D gradient covariances using back-projection. They demonstrated a higher success rate and capture range but did not report on the 3-D accuracy. Schmid and Chênes [12] modeled the similarity as the sum of image forces between the projected shape contour and the extracted X-ray silhouettes. Block matching between the DRR and the X-ray image at the projected silhouette points explicitly measures the image force that was modeled to follow the Hooke’s law.

Manuscript received March 15, 2017; revised April 30, 2017; accepted May 3, 2017. Date of publication May 8, 2017; date of current version August 31, 2017. (Corresponding author: Jian Wang.)

J. Wang is with the Pattern Recognition Lab, Friedrich-Alexander University of Erlangen-Nuremberg, 91058 Erlangen, Germany, and also with Siemens Healthcare GmbH, 91301 Forchheim, Germany (e-mail: jian.wang@cs.fau.de).

R. Schaffert, J. Hornegger, and A. Maier are with the Pattern Recognition Lab, Friedrich-Alexander University of Erlangen-Nuremberg, 91058 Erlangen, Germany.

A. Borsdorf and B. Heigl are with Siemens Healthcare GmbH, 91301 Forchheim, Germany.

X. Huang is with the Pattern Recognition Lab, Friedrich-Alexander University of Erlangen-Nuremberg, 91058 Erlangen, Germany, and also with the Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, 200400 Shanghai, China.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TMI.2017.2702100

The model requires additional pivot constraints to achieve accurate registration in 3-D. Berger et al. [14] proposed 2-D/3-D registration-based motion-free reconstruction in weight-bearing cone-beam CT of the knee joint, where gradient-based similarity measures are demonstrated to be promising for 2-D/3-D registration. Otake et al. [15] proposed robust 2-D/3-D registration for spine interventions, where multi-start covariance matrix adaptation evolution strategy (CMA-ES) [16] with local restarts demonstrated a significantly increased success rate in the presence of deformation and content mismatch. Dum  nil et al. [17] proposed a versatile rigid registration method that employs reduced exhaustive search involving a multi-resolution scheme and decomposition of the transformation. However, both [15] and [17] rely on large populations of start transformations, which is suboptimal for real-time motion correction.

In general, the above mentioned approaches are proposed to serve as initial registration at the beginning of the intervention. Patient motion can still introduce misalignment after initial registration. In practice, if certain misalignment caused by patient motion is observed, the clinicians have to pause the intervention and trigger the registration process. Therefore, automatic patient motion compensation is essential for more reliable fusion applications. Rohlfing et al. [18] proposed a markerless real-time 3-D target region tracking for patient motion compensation, where multiple projection images acquired from different viewing directions are required to estimate the 3-D motion from back-projected 2-D motions. Aichert et al. [19] proposed a tracking approach that estimates 3-D rigid patient motion using the Epipolar Consistency Condition (ECC). The ECC-based tracking gives promising results using only a few X-ray images, but the images must fulfill the ECC assumptions [20].

Rigid 2-D/3-D registration using single-view images is closely related to the single-view rigid pose estimation problem in computer vision, e.g. the Perspective- n -Point (PnP) problem. It aims at determining the pose and the orientation of a camera given its intrinsic parameters and a set of n point correspondences between 3-D points and their 2-D projections [21]. Lepetit et al. [21] proposed a closed-form $\mathcal{O}(n)$ solution to the PnP problem, shortened as EPnP, which efficiently estimates both planar and non-planar configurations by expressing the n points ($n \leq 4$) as a weighted sum of four virtual control points. However, direct point correspondences between 3-D volume and 2-D X-ray images are usually not available, because the pixel intensity of an X-ray image is determined by the attenuation of the X-ray through the target object.

In [13], Wang et al. introduced the differential patient motion compensation approach, where the differential 3-D rigid motion is estimated from the 2-D optical flow together with a correspondence model that incorporates both the observable and the unobservable 3-D motion. It demonstrated the potential of the correspondence model to recover 3-D motion out of 2-D observations. However, the approach in [13] relies on a highly accurate initial registration to set up the 2-D/3-D correspondences. Without an good initialization, the 3-D motion cannot be accurately estimated because the

depth information from the 2-D/3-D correspondences is crucial for the accuracy. Moreover, the correspondences are only initialized once at the beginning and later on maintained by 2-D tracking. Significant systematic error is induced by accumulated large motion because the tracked 2-D targets do not match the initial 3-D correspondences anymore. Strictly speaking, the differential motion compensation in [13] is not capable of providing an accurate registration and only maintains the established overlay up to a certain range of motion.

In this paper, we suggest a novel dynamic 2-D/3-D registration framework, where the 2-D/3-D correspondences are iteratively updated by the direct measure of local 3-D to 2-D misalignment, and the 3-D transformation estimation is constrained by the point-to-plane correspondence (PPC) model introduced in [13]. As initialization, a set of 3-D surface points are selected as sparse representation of the salient structures for the registration. In each registration iteration, the 3-D points defining the contour generator [22] are selected. These are the points with their gradient perpendicular to the viewing ray, and their projection defines the apparent contour [22]. The goal of this approach is to minimize the visual misalignment between the projection of the contour generator points from the 3-D volume and the actual apparent contour points from the 2-D X-ray image. The dynamic registration iterates between two steps: 1) given an updated registration, contour generator points are updated and projected onto the image plane, with their depths and 3-D gradients preserved. The depth-aware gradient projections (DGP) are also computed; 2) the local misalignment is explicitly measured between DGPs and the X-ray image, and serves as input for the PPC model for robust estimation of the registration update. Comparing to [13], the iterative update of the 3-D contour generator and the 2-D/3-D correspondences resolve the problem of the accumulating systematic error mentioned in [13]. Instead of incremental 2-D tracking as in [13], the direct 3-D/2-D misalignment measurement enables the use of the PPC model even without an accurate initial registration. Given the strength of the PPC model in estimating both planar and non-planar transformations, high 3-D accuracy can be achieved for single-view images without additional constraints as in [12]. Our dynamic registration framework is capable of initializing and maintaining an accurate 2-D/3-D registration, even for single-view X-ray images. In the evaluation, we demonstrate high 3-D accuracy and robustness using different datasets. The proposed approach is intuitive and generic, thus can be applied in many relevant clinical applications.

II. GEOMETRY SETUP OF 2-D/3-D IMAGE FUSION

An interventional C-arm system can be modeled as a pinhole camera for X-ray imaging [2], [22], where the X-ray source is the camera center and the detector is the imaging plane, see Fig. 1(a). Its projection geometry is described by the projection matrix

$$\mathbf{P}_0 = \mathbf{K}[\mathbf{R}_{\text{view}}|\mathbf{t}_{\text{view}}] \in \mathbb{R}^{3 \times 4},$$

where

$$\mathbf{K} = \begin{pmatrix} f & 0 & a_x \\ 0 & f & a_y \\ 0 & 0 & 1 \end{pmatrix} \quad (1)$$

is the camera matrix containing the intrinsic parameters, i.e. focal length f and the principal point (a_x, a_y) , $\mathbf{R}_{\text{view}} \in \mathbb{R}^{3 \times 3}$ and $\mathbf{t}_{\text{view}} \in \mathbb{R}^3$ are the extrinsic parameters representing respectively the orientation and the position of the X-ray source in the world coordinate system [22]. For a calibrated C-arm system, both intrinsic and extrinsic parameters are available. Without loss of generality (w.l.o.g.), the following derivation is done in the camera coordinate system, where the projection matrix is simplified as

$$\mathbf{P} = \mathbf{K}[\mathbf{I}|\mathbf{0}]. \quad (2)$$

To create a 2-D/3-D image fusion, the volume can be rendered as imaged from the X-ray source, and the resulting image can be blended onto the X-ray image [2]. The task of 2-D/3-D registration is to estimate the 3-D transformation of the volume to the actual patient position during the X-ray scan, such that the rendered image of the volume is well aligned with the X-ray image. In this paper, 3-D rigid transformation is used to represent the general patient movement.

Usually an initial pose estimation $\mathbf{T}_0 \in \mathbb{R}^{4 \times 4}$ is available, e.g. by performing rough manual alignment. It does not necessarily to ensure an accurate image fusion but provides a good starting condition for the automatic registration. The objective of the automatic registration is to estimate the registration matrix $\mathbf{T}_r \in \mathbb{R}^{4 \times 4}$ based on \mathbf{T}_0 . Given the total registration $\mathbf{T} = \mathbf{T}_r \mathbf{T}_0$, the volume V is rendered as

$$I^{\text{proj}} = \mathcal{R}_{\mathbf{PT}}(V)$$

and overlaid onto the X-ray image I^{flr} , where $\mathcal{R}_{\mathbf{PT}}$ denotes the volume rendering under the projection geometry \mathbf{PT} .

III. THE POINT-TO-PLANE CORRESPONDENCE MODEL

The intuition behind 2-D/3-D image registration is to minimize the visual misalignment of the common structures between the rendered image I^{proj} and the X-ray image I^{flr} by optimizing the registration matrix \mathbf{T}_r . In this paper, we measure the visual misalignment from 3-D rendered image to 2-D X-ray image, thus the visual misalignment is referred as 3-D/2-D misalignment. Intuitively, the observable 3-D/2-D misalignment is related to the regions with distinctive image gradient. In both I^{proj} and I^{flr} , the regions with distinctive gradient appear as boundaries or edges of anatomical structures, known as apparent contour [22]. The 2-D apparent contour usually corresponds to the 3-D contour generator [22] of the regions having remarkably different attenuation values compared to their surroundings, e.g. for bone structures surrounded by water-like environment or soft tissue. Therefore, our proposed registration method focuses on the 3-D contour generator and the corresponding 2-D apparent contour.

The 3-D/2-D misalignment is observed in the image space, where the depth information is missing. Thus, directly estimating the 3-D transformation from the 2-D observation is not trivial. The point-to-plane correspondence (PPC) model is initiated by the relation between the incremental movement and 3-D gradient [13]. It tackles the difficulty of estimating 3-D rigid transformation by leveraging the 3-D image gradient and the depth information.

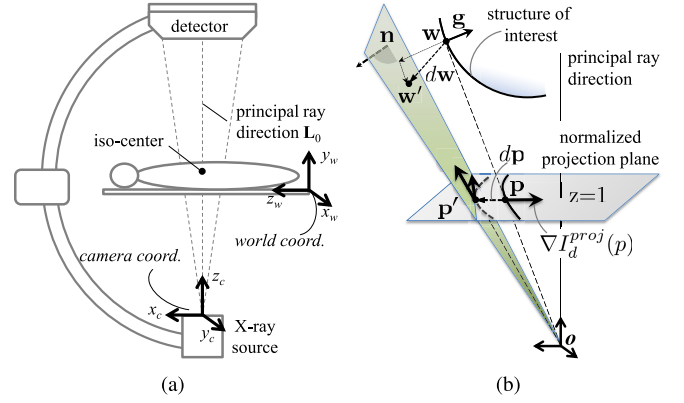


Fig. 1. (a) The projection model of a C-arm system and the coordinate systems; (b) Point-to-plane correspondence (PPC) model, details in Sect. III.

In the 2-D/3-D registration scenario, the 2-D observation and the 3-D transformation differ in spatial dimension but are associated by the projection geometry. To ease the demonstration of the point-to-plane correspondence model, the normalized image coordinate is used to represent the 2-D image content in 3-D space. W.l.o.g., a 2-D image point $\mathbf{x} \in \mathbb{R}^2$ 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 of unit depth $z = 1$ in the camera coordinate system, see Fig. 1(b). Under the convention of Eq. (2) with Eq. (1), given the image coordinate $\mathbf{x} = (u, v)$, the normalized image coordinate is formulated as

$$\mathbf{p} = \mathcal{N}_H(\mathbf{K}^{-1} [u; v; 1]) = \begin{pmatrix} \frac{(u - a_x)}{f} \\ \frac{(v - a_y)}{f} \\ 1 \end{pmatrix},$$

where $\mathcal{N}_H(\cdot)$ denotes the homogeneous normalization with the last element as 1. Later in the derivation, a 2-D point is represented by the normalized image coordinate \mathbf{p} . Accordingly, the 2-D image is scaled with respect to the focal length, which is later denoted as I in the normalized image coordinate system. The image gradient at point \mathbf{p} is denoted as $\nabla I(\mathbf{p})$.

Consider that an incremental movement takes place under continuous X-ray acquisition. For a 2-D apparent contour point \mathbf{p} , only the motion component perpendicular to the edge can be determined, i.e. the motion along the image gradient $\nabla I(\mathbf{p})$. Similarly for the corresponding 3-D contour generator point $\mathbf{w} \in \mathbb{R}^3$, the incremental movement only causes a change of intensity values if the movement has a component in the direction of its 3-D gradient $\mathbf{g} = \nabla f(\mathbf{w}) \in \mathbb{R}^3$. This allows to assume that only the motion along the image gradient is locally observable. Nevertheless, the local observation does not directly determine the real movement. However, with constraints from different gradient directions, the real motion can be determined by considering all observations under a certain motion model.

As shown in Fig. 1(b), consider that a small 3-D movement is applied to the structure of interest. For a contour generator point \mathbf{w} and its apparent contour point \mathbf{p} , the new position is observed as $\mathbf{p}' = \mathbf{p} + d\mathbf{p}$. Based on the above considerations,

the observation \mathbf{p}' is along the image gradient but does not necessarily determine the true movement. In other words, the 3-D position $\mathbf{w}' = \mathbf{w} + d\mathbf{w}$ after the movement does not necessarily lie on the back-projected ray of \mathbf{p}' , because the motion components either in depth or perpendicular to the gradient direction cannot be determined. To sum up, the observable component is determined by the vector \mathbf{p}' and the unobservable component is constrained by the span of the depth direction \mathbf{w} and the gradient direction \mathbf{g} , namely $\mathbf{w} \times \mathbf{g}$. In this way, the real motion is constrained in the span space of the observable and unobservable components, i.e. the span of $(\mathbf{w} \times \mathbf{g})$ and \mathbf{p}' . Locally, we say that \mathbf{w}' lies on a plane through the origin with the normal defined as

$$\mathbf{n} = \mathcal{N}_U((\mathbf{w} \times \mathbf{g}) \times \mathbf{p}'), \quad (3)$$

where $\mathcal{N}_U(\cdot)$ normalizes a vector to unit length.

Following the above derivations, the new 3-D position \mathbf{w}' after motion fulfills the point-on-plane condition

$$\mathbf{n}^\top \mathbf{w}' = 0. \quad (4)$$

The above formulation constrains a contour generator point to a 3-D plane under 3-D movement, which is thus denoted as the point-to-plane correspondence (PPC) model.

Under the rigid motion, the new 3-D position \mathbf{w}' is transformed from \mathbf{w} as

$$\mathbf{w}' = \delta \mathbf{R} \mathbf{w} + d\mathbf{v}, \quad (5)$$

where $\delta \mathbf{R} \in \mathbb{R}^{3 \times 3}$ denotes the rotation matrix and $d\mathbf{v}$ denotes the translation vector. Since only incremental motion is involved in each registration iteration, we assume only small rotation angles ($\omega < 15^\circ$). Around $\omega = 0$, the Taylor series of sine and cosine functions are respectively $\sin(\omega) \simeq \omega + \mathcal{O}(\omega^3)$ and $\cos(\omega) \simeq 1 + \mathcal{O}(\omega^2)$. In the Euler angle matrix entries, $\sin(\omega)$ approximately equals to ω and $\cos(\omega)$ approximately to one. By substituting the above approximations, the Euler angle rotation matrix becomes linear to the vector $d\boldsymbol{\omega} = (\omega_x, \omega_y, \omega_z)$ as

$$\delta \mathbf{R} = \begin{bmatrix} 1 & -\omega_z & \omega_y \\ \omega_z & 1 & -\omega_x \\ -\omega_y & \omega_x & 1 \end{bmatrix} = \mathbf{I}_{3 \times 3} + [d\boldsymbol{\omega}]_\times, \quad (6)$$

where $[d\boldsymbol{\omega}]_\times \in \mathbb{R}^{3 \times 3}$ is a skew symmetric matrix that represents cross products as matrix multiplications. By substituting Eq. (6) in Eq. (5), it yields

$$\mathbf{w}' = \mathbf{w} + [d\boldsymbol{\omega}]_\times \mathbf{w} + d\mathbf{v} = \mathbf{w} + d\boldsymbol{\omega} \times \mathbf{w} + d\mathbf{v}. \quad (7)$$

Thus, the incremental motion vector $d\mathbf{w} = d\boldsymbol{\omega} \times \mathbf{w} + d\mathbf{v}$. By substituting Eq. (7) to Eq. (4), it yields

$$\mathbf{n}^\top (d\boldsymbol{\omega} \times \mathbf{w}) + \mathbf{n}^\top d\mathbf{v} + \mathbf{n}^\top \mathbf{w} = 0.$$

After reformulation, the linear constraint between the differential motion and a contour point can be formulated as

$$((\mathbf{n} \times \mathbf{w})^\top - \mathbf{n}^\top) \delta \mathbf{v} = \mathbf{n}^\top \mathbf{w}, \quad (8)$$

where $\delta \mathbf{v} = (d\boldsymbol{\omega}^\top d\mathbf{v}^\top)^\top \in \mathbb{R}^6$ denotes the 3-D differential transformation vector.

Each contour generator point and its misalignment observation provide one local constraint according to Eq. (8) on the global 3-D rigid motion. Given all misalignment observations of the contour generator points, a system of linear equations can be assembled by combining the set of contour point correspondences $\{\mathbf{w}_i, \mathbf{p}_i, \mathbf{p}'_i\}$ as

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

where $\mathbf{a}_i = ((\mathbf{n}_i \times \mathbf{w}_i)^\top - \mathbf{n}_i^\top) \in \mathbb{R}^6$ is the i -th row of matrix $\mathbf{A} \in \mathbb{R}^{N \times 6}$, $b_i = \mathbf{n}_i^\top \mathbf{w}_i$ is the i -th entry of vector $\mathbf{b} \in \mathbb{R}^N$ and N is the number of point correspondences. At least $N = 6$ points are required for the closed-form solution of the motion vector $\delta \mathbf{v}$ using the pseudo-inverse as

$$\delta \mathbf{v} = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b}. \quad (10)$$

Given the differential motion vector $\delta \mathbf{v} = (d\boldsymbol{\omega}^\top d\mathbf{v}^\top)^\top$, the rotation matrix $\delta \mathbf{R} \in \mathbb{R}^{3 \times 3}$ can be calculated from $d\boldsymbol{\omega}$. Note that the linear approximate in Eq. (6) does not ensure the properties of a rotation matrix, i.e. an orthogonal matrix with determinant as one. Instead, the original Euler angle matrix can be employed. Alternatively but equivalently, the rotation vector $d\boldsymbol{\omega}$ can also be converted into its Euler axis-angle representation as the axis $\mathbf{r} = d\boldsymbol{\omega} / \|d\boldsymbol{\omega}\|$ and the angle $\theta = \|d\boldsymbol{\omega}\|$. The rotation matrix can be calculated as

$$\delta \mathbf{R} = \cos \theta \mathbf{I} + (1 - \cos \theta) \mathbf{r} \mathbf{r}^\top + \sin \theta [\mathbf{r}]_\times.$$

The 3-D rigid transformation $\mathbf{T}' \in \mathbb{R}^{4 \times 4}$ is then updated as

$$\mathbf{T}' = \begin{bmatrix} \delta \mathbf{R} & d\mathbf{v} \\ \mathbf{0} & 1 \end{bmatrix} \mathbf{T}.$$

IV. DYNAMIC 2-D/3-D RIGID REGISTRATION FRAMEWORK USING PPC MODEL

After the mathematical derivation of the PPC model in Sect. III, this section presents the dynamic 2-D/3-D registration framework based on the PPC model.

An overview of our dynamic registration framework is shown in Fig. 2. In the initialization step, surface points of the bone structures are extracted from the 3-D volume. These are considered as the sparse representation of the salient structures, a subset of which counts for the pose-dependent contour generator (Sect. IV-A). The registration is an iterative procedure as follows. The contour generator points are selected and projected onto the image plane, where both depth and gradient of the points are preserved for later computation. In addition, the volume V is projected as depth-aware gradient projections (DGPs) (Sect. IV-B). Given the above depth-aware contour points and DGPs, the 3-D/2-D misalignment can be measured by patch matching using local gradient correlation (Sect. IV-C). Robust estimation schemes are combined with the PPC model to estimate the 3-D transformation, which serves as an update to the next iteration. This procedure is iterated until convergence (Sect. IV-D). The 3-D/2-D misalignment measurement updates the 2-D/3-D correspondences in each iteration, which does not rely on an accurate initial alignment. The framework essentially extends the PPC model from differential motion compensation [13]

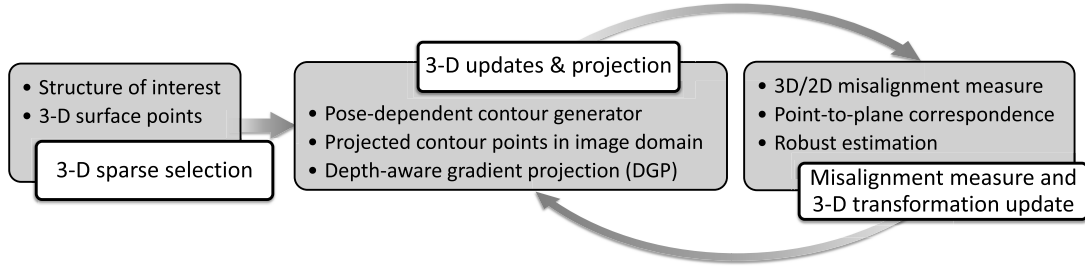


Fig. 2. Overview of the dynamic 2-D/3-D registration pipeline.

to dynamic registration scenarios. Furthermore, involving the intermediate 3-D updates resolves the error accumulation problem raised in differential motion compensation [13].

A. Sparse Surface Representation of the 3-D Volume

In order to reduce the dimensionality of the problem, we begin by identifying the criteria of salient structures in the input 3-D volume, i.e. the structures to be aligned, which have distinctive contrast and are visible in the 2-D X-ray image. Since our objective is the 3-D rigid transformation, we consider only the structures with high rigidity, e.g. bone structures or cerebral vessel trees. W.l.o.g., intensity windowing on the voxels is applied to select the salient structures.

After identifying the salient structures, we further focus on their contours. In computer vision, the set of surface points with their normal perpendicular to the viewing rays defines the contour generator in 3-D space, and their projection defines the apparent contour in 2-D space [22]. As initialization, surface points of the salient structures are firstly extracted from the 3-D volume as sparse representation. A 3-D Canny edge detector [23] is used to extract the surface points $\{w_i\}_{\text{all}}$. For further 3-D updates, the contour generator points are selected from the sparse surface set $\{w_i\}_{\text{all}}$.

The 3-D Canny edge detector involves the 3-D gradient computation. The derivative computation is sensitive to noise and artifacts. To avoid this, the input volume is preprocessed using the 3-D guided image filter [13], [24] (GIF), which effectively enhances the edges and preserves the image gradient [25]. Moreover, the gradient is computed using the error function of third degree (3EF) suggested in [26].

B. Pose-Dependent 3-D Updates and Projection

As shown in Fig. 2, pose-dependent 3-D updates and projection provide the up-to-date input for 3-D/2-D misalignment measurement and motion estimation. In this step, the contour generator points are selected using the perpendicularity condition between gradients and the viewing ray, and projected onto the 2-D image plane. The 3-D gradients and depth are also preserved and later serve as input for the PPC model.

1) Pose-Dependent Contour Generator: Most notably, the contour generator points have their gradients perpendicular to the viewing rays and produce the apparent contours in the 2-D image [1], [26]. The gradients along the viewing ray do not contribute much information that can be used for registration. Thus, 3-D contour generator points are further

updated that fulfill the perpendicular condition [13]

$$\theta_i = \arccos(|(\mathbf{g}_i \cdot \mathbf{w}_i) / (\|\mathbf{g}_i\| \cdot \|\mathbf{w}_i\|)|) \geq \theta_T,$$

where θ_i is the angle between the 3-D gradient \mathbf{g}_i and the corresponding viewing ray \mathbf{w}_i of the 3-D point $w_i \in \{w_i\}_{\text{all}}$. The perpendicular threshold θ_T controls the contour thickness and is empirically set close to 90° , e.g. 87° . A selected contour point w_i is projected to 2-D space as

$$p_i = \mathcal{P}_{\mathbf{PT}}(w_i),$$

where $\mathcal{P}_{\mathbf{PT}}$ denotes the projection operation under the given projection geometry \mathbf{PT} . A subset of points is extracted from the surface points as contour generator points $\{\mathbf{p}_i, \mathbf{w}_i, \mathbf{g}_i\}$, preserving the 3-D position (depth) and gradient.

2) Pose-Dependent Depth-Aware Gradient Projections: Furthermore, the volume V is forward projected generating 2-D images. The projection images show the appearance of the 3-D volume under the given projection geometry and provide an essential basis for the 3-D/2-D misalignment measure. As suggested by [27] and [28], we chose the gradient projection (GP) rendering that samples on 3-D gradient during ray casting and directly computes the gradient projection images.

In the standard GP, 3-D structures are projected and overlap each other in the rendered image. Since there is no distinction to depths of overlapping structures, standard GP does not adequately present the appearance of the contour generator which overlaps with other contours. To better distinguish the overlapping contours from different depth intervals, we refer to the concept of depth layers [29] and render the depth-aware gradient projections (DGP). Those are the GP images from different depth intervals that are uniformly distributed along the principal ray direction, denoted as

$$\nabla I_d^{\text{proj}} = \mathcal{R}_{\mathbf{PT}}^{\text{DGP}}(V, d),$$

where $\mathcal{R}_{\mathbf{PT}}^{\text{DGP}}$ denotes the projection operation and d is the depth index of a DGP image ∇I_d^{proj} . The stack of DGP images, denoted as $\{\nabla I_d^{\text{proj}}\}$, contains the gradient appearance of V in 2-D space with distinction to depths of overlapping structures.

In each iteration k , the contour generator points $\{\mathbf{p}_i, \mathbf{w}_i, \mathbf{g}_i\}^k$ and the DGPs $\{\nabla I_d^{\text{proj}}\}^k$ are computed from 3-D. Each contour generator point \mathbf{w}_i is associated with the corresponding DGP image ∇I_d^{proj} , the depth interval of which contains \mathbf{w}_i . The image patch $\Omega(\nabla I_d^{\text{proj}}, \mathbf{p}_i)$ around \mathbf{p}_i provides the gradient projection appearance of \mathbf{w}_i , where $\Omega(I, \mathbf{p})$ denotes the image patch from an image I centered at \mathbf{p} . On one hand, as the gradient projection appearance is limited within the depth

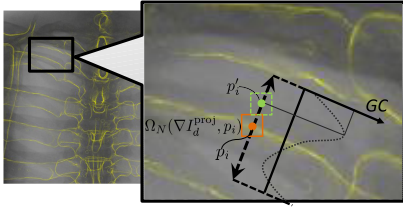


Fig. 3. The 3-D/2-D misalignment measure via patch matching. An example region of an 2-D/3-D overlay is zoomed in, where the contour rendering (yellow contours) of the 3-D volume is overlaid onto an X-ray image. The template patch $\Omega_N(\nabla I_d^{\text{proj}}, p_i)$ (orange square) centered at p_i is taken from the corresponding DGP image. The search range (dotted arrow line) is defined by the gradient projection direction and the range scale. Within the search range, gradient correlation (GC) is computed (dotted curve). The match position p'_i is the one with the maximum GC. The difference vector $dp_i = p'_i - p_i$ is considered as 3-D/2-D misalignment.

interval, it reduces the chance of overlapping structures. On the other hand, the patch $\Omega(\nabla I_d^{\text{proj}}, p_i)$ is more robust compared to the 3-D gradient \mathbf{g}_i (Fig. 1(b)), because the DGP patch considers the 2-D neighborhood and the depth integral.

C. Measure of 3-D/2-D Misalignment

Figure 3 shows the 3-D/2-D misalignment measure for a contour generator point \mathbf{p}_i given its corresponding DGP ∇I_d^{proj} . An image patch of the DGP $\Omega(\nabla I_d^{\text{proj}}, p_i)$ is taken around \mathbf{p}_i . The misalignment is measured by searching along the gradient direction $\nabla I_d^{\text{proj}}(\mathbf{p}_i)$ within a certain range for the best match on the X-ray gradient image ∇I^{fir} . Within the search range, the local gradient correlation (GC) [28] is computed between the template patch $\Omega(\nabla I_d^{\text{proj}}, p_i)$ and the candidate patch $\Omega(\nabla I^{\text{fir}}, \tilde{\mathbf{p}}'_i)$ at the candidate position $\tilde{\mathbf{p}}'_i$, and the one with the maximum GC value is considered as the best match \mathbf{p}'_i , which serves as input to the PPC model. The difference $\mathbf{p}'_i - \mathbf{p}_i$ is the observed 3-D/2-D misalignment.

In practice, a threshold τ_{GC} is used to filter out non-valid matches, where the maximal GC is below τ_{GC} . Non-valid matches are neglected for further transformation estimation. In our implementation, we use a low threshold (e.g. $\tau_{GC}^l = 0.1$) in the initial iterations to capture rough matches, and a higher threshold (e.g. $\tau_{GC}^h = 0.5$) in the last iterations and in dynamic motion compensation to ensure high accuracy.

D. Robust Motion Estimation

In order to enhance the robustness of the framework, we use an alternative formulation of the PPC model based on the DGP gradient $\nabla I_d^{\text{proj}}(p)$ (Sect. IV-B.2). In the original formulation of PPC, the local 3-D gradient \mathbf{g} at \mathbf{w} is used in the span of unobservable component in Eq. (3). As shown in Fig. 1(b), $\mathbf{p} \times \nabla I_d^{\text{proj}}(p)$ provides the identical span as $\mathbf{w} \times \mathbf{g}$ does. As already stated in Sect. IV-B.2, the DGP gradient is the integral within the corresponding depth interval and thus considered more robust against noise compared to the local gradient \mathbf{g} . Additionally, since DGPs are already computed by volume rendering for misalignment measure, this modification does not introduce extra computation expense. Therefore,

we choose the more robust formulation of the plane normal

$$\mathbf{n}' = \mathcal{N}_U \left((\mathbf{p} \times \nabla I_d^{\text{proj}}(p)) \times \mathbf{p}' \right). \quad (11)$$

The registration problem remains as estimating the transformation parameters that fulfill the linear constraints composed of Eq. (8) using all selected contour points and their correspondences, with the plane normal formulated as Eq. (11).

The objective of the transformation estimation is to find $\delta \mathbf{v}$ from observations \mathbf{b} . Ideally, the relationship between $\delta \mathbf{v}$ and \mathbf{b} is modeled linearly as in Eq. (9). However, noise and outliers on the observations are hardly to avoid in practice, thus Eq. (9) can not be exactly satisfied. In this case, estimating $\delta \mathbf{v}$ becomes an optimization problem for losses introduced by the residuals, i.e.

$$\min_{\delta \mathbf{v}} \sum_{i=1}^N L(b_i - \mathbf{a}_i^T \delta \mathbf{v}).$$

When the noise is i.i.d. and follows the Gaussian distribution, minimizing the sum of the squared residuals $L(r) = r^2$ is used to estimate the true transformation. Moreover, its analytical solution is given by Eq. (10). However, when the noise is non-Gaussian, especially when the observations are corrupted by outliers, the least squares estimator could be far from the real motion. This sensitivity comes from the fact that in $L(r) = r^2$ the loss increases quadratically with the residual. Then, its result tends to follow large residuals, which nevertheless are more likely to be outliers.

To reduce the effect of outliers, we need to change the quadratic loss by some robust loss functions. One popular choice is to use the absolute deviation $L(r) = |r|$, resulting in the following *least absolute deviation regression (LADR)*:

$$\min_{\delta \mathbf{v}} \sum_{i=1}^N |b_i - \mathbf{a}_i^T \delta \mathbf{v}|. \quad (12)$$

Statistically, LADR estimates the median as discussed in [30] and it is robust to outliers. However, there is no analytic solution for Eq. (12). To effectively solve it, we apply the iteratively reweighted least squares (IRLS), see [31]. In IRLS, we iteratively solve a sequence of weighted least squares problems, which have analytic solutions. Specifically, the iteration can be written as

$$\delta \mathbf{v}^{(t)} = (\mathbf{A}^T \mathbf{W}^{(t-1)} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W}^{(t-1)} \mathbf{b}, \quad (13)$$

where $\mathbf{W}^{(t-1)}$ is a diagonal matrix in the $(t-1)$ -th iteration. For LADR, the diagonal elements of $\mathbf{W}^{(t-1)}$ are determined by

$$\mathbf{W}_{ii}^{(t-1)} = \frac{1}{|r_i^{(t-1)}|},$$

where $r_i^{(t-1)} = b_i - \mathbf{a}_i^T \delta \mathbf{v}^{(t-1)}$ is the residual of the i -th observation in the $(t-1)$ -th iteration.

In the LADR equation Eq. (12), the loss increases linearly with the residual and hence the effect of outliers is significantly smaller in the least squares regression. To further weaken their effect, some non-convex losses can be considered. In this

paper, we also consider Maximum Correntropy Criterion for Regression (MCCR), of which the loss function is given by

$$L_\sigma(r) = \exp\left(-r^2/2\sigma^2\right),$$

where $\sigma > 0$ is the scale parameter related to the robustness. This loss is introduced by correntropy, which is a generalized similarity measure [32]. Since the proposal of MCCR, it has been successfully applied in signal processing, machine learning, and image processing, see, e.g., [33], [34], [35]. The statistical learning interpolation of MCCR has been analyzed in [36]. Intuitively, on one hand, $L_\sigma(r)$ gives a relatively small penalty for a large residual, which makes it robust to outliers. On the other hand, for a small residual, $L_\sigma(r)$ performs like the squared error, which makes MCCR perform well for Gaussian noise. Moreover, $L_\sigma(r)$ is sufficiently smooth, which benefits the optimization process.

Based on MCCR, the motion estimation is formulated as

$$\min_{\delta \mathbf{v}} \sum_{i=1}^N \exp\left(-\frac{(b_i - \mathbf{a}_i^T \delta \mathbf{v})^2}{2\sigma^2}\right), \quad (14)$$

which could be also solved within the scheme of IRLS Eq. (13). The weight matrix for Eq. (14) is given by

$$\mathbf{W}_{ii}^{(t-1)} = \exp\left(-\frac{(r_i^{(t-1)})^2}{2(\sigma^{(t-1)})^2}\right).$$

Notice that σ is here adaptively tuned during the iterations. Following the suggestion in [35], we set $\sigma^{(t-1)}$ as the standard deviation of $\{r_i^{(t-1)}\}$.

V. IMPLEMENTATION

In this section, several important aspects related to implementation are revealed.

A. Scenarios for Initial Registration and Motion Compensation

The proposed registration framework is capable of both establishing an accurate initial 2-D/3-D overlay and compensating the patient motion during continuous fluoroscopy.

As indicated in Fig. 2, the sparse surface representation is computed once as initialization. The dynamic registration iterates between pose-dependent 3-D updates and transformation update out of 3-D/2-D misalignment measure.

In the initial registration scenario, either single-view or multi-view images can be used for registration. Using a multi-scale scheme (Sect. V-B) will likely increase the capture range. In the initial registration scenario, the iteration involves only static 2-D image(s). Note that the small angle assumption ($< 15^\circ$) holds for each iteration. Therefore, more iterations are needed for the convergence such that the registration can capture relatively large initial misalignment.

In the dynamic registration scenario, X-ray images are continuously acquired and the iteration is always associated to the latest X-ray image, such that the 3-D volume is dynamically registered to the 2-D scene.

For performance considerations, different criteria can be setup for initial registration, continuous motion compensation and refinement. For initial registration, multi-scale and

multi-view schemes with iterations associated to one acquisition can be considered. During continuous fluoroscopy, usually only one iteration for an incoming image is required to compensate the patient motion on the fly. If the fluoroscopy is paused, the refinement can be performed with more iterations associated to the latest 2-D image.

B. Multi-Scale Scheme

In the initial registration scenario, the 3-D/2-D misalignment can be relatively large, which requires a large search range and more iterations. Therefore, the initial registration is performed on different resolution levels. In this case, the search range and patch size remain the same for all resolution levels. The low resolution levels enhance the capture range without increasing the computation effort with large search range and patch size.

C. Convergence Criteria

For initial registration and motion compensation refinement scenarios, the registration iterates over the same 2-D image until convergence (or the maximal iteration) is reached. Unlike the registration methods where the similarity function is directly optimized, our registration employs implicitly the local image similarity (gradient correlation) for misalignment measurement to drive the PPC model for motion estimation. There are several convergence indicators in our registration framework:

- A direct convergence criterion is the residual of the PPC estimation approaching to zero. Since only the contour generator points with corresponding apparent contour points serve as input to the PPC estimation, the residual only reflects the quality on those input points;
- The motion vector (see Eq.(10)) estimated in each iteration approaching to zero also indicates the convergence;
- Alternatively, the local GC (LGC) for all contour generator points is computed for the patch matching in the misalignment measurement step. The mean LGC also indicates the convergence by approaching to a certain upper limit.

Note that, the misalignment measure suggests the correspondence locally at each contour point and the motion update is globally constrained by the PPC model. The estimated motion does not necessarily lead each contour generator point to its found correspondence, especially when outliers exist (low image quality or presence of mismatch structures). In this case, the registration update oscillates about a certain point. We observed the case where the last registration update before the iteration finishes (or gets terminated) is not as accurate as one of the previous registration updates. In our implementation, we define a registration quality factor for the k -th iteration

$$Q_k = \frac{\bar{r}_k}{\bar{s}_k}, \quad (15)$$

where \bar{r}_k denotes the mean residual of the PPC input points and \bar{s}_k denotes the mean local similarity over all contour generator points with current registration update. Here we define the local similarity as

$$s_i = LGC(\Omega(\nabla I^{\text{flr}}, \mathbf{p}_i), \Omega(\nabla I_d^{\text{proj}}, \mathbf{p}_i)) + \eta,$$

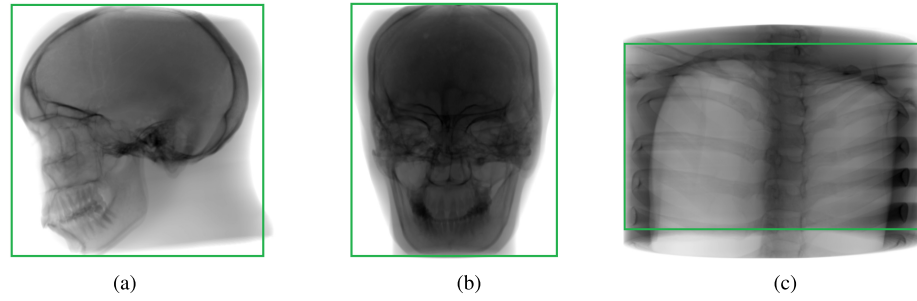


Fig. 4. The simulated X-ray images (digitally reconstructed radiographs) of the head phantom and the thorax phantom, rendered using the C-arm CT volumes. The green boxes show the regions of interest for registration. (a) Head (LAT). (b) Head (AP). (c) Thorax (AP).

where LGC denotes the local gradient correlation at point p_i and η is a constant to avoid zero division. The quality factor Q_k takes both the image similarity and the PPC estimation into account. Along with the iterations $k = 0, \dots, K$, the registration with the smallest quality factor is marked as the best registration. The convergence must fulfill two criteria: 1) the change of the quality factor between two iterations approaches to zero (below a small threshold) and 2) the change of the minimal quality factor approaches to zero.

VI. EXPERIMENTS AND EVALUATION

To thoroughly evaluate the proposed registration framework, experiments are organized as follows:

- 1) In Sect. VI-A, evaluation on simulated X-ray images of phantoms is presented. Static single-view DRRs of phantoms serve as simulated input X-ray images for the initial registration scenario;
- 2) In Sect. VI-B, evaluation on X-ray image sequences of a thorax phantom is presented. The dynamic registration scenario applies to correct the continuous motion of the phantom during the acquisition;
- 3) In Sect. VI-C, quantitative evaluation on clinical datasets with “gold-standard” registrations is presented. Static single-view DSA or X-ray images serve as input as in the initial registration scenario;
- 4) In Sect. VI-D, qualitative evaluation on other clinical datasets is presented for the initial registration scenario.

The experiments focus on the two algorithms proposed in this work, namely

PPC-LADR the PPC-based dynamic registration with the LADR estimation strategy (Eq. (12)) and

PPC-MCCR the PPC-based dynamic registration with the MCCR estimation strategy (Eq. (14)).

Beyond these algorithms, we also include the following algorithms for comparison in different sections:

BGB a back-projection gradient-based (BGB) 2-D/3-D registration method proposed in [9], Sect. VI-A and VI-C;

DGP-EPNP the EPnP solution [21] is used instead of PPC for motion estimation. DGP-EPNP aims to demonstrate the advantage of the PPC model over the Point- n -Point correspondence model, Sect. VI-A, VI-B and VI-C;

DT the differential tracking (DT) approach for motion compensation is previously proposed in [13], Sect. VI-B;

GP-PPC standard gradient projection (GP) is used instead of depth-aware gradient projections (DGP). It aims to

demonstrate the contribution of the depth-layer concept [29], Sect. VI-C.

A. Evaluation on Simulated X-Ray of Phantoms

As already stated, the strength of the proposed method compared to [13] is that our method extends the use of the PPC model from the differential tracking to a self-contained registration framework. In this section, we assess the dynamic registration in the initial registration scenario.

A head phantom and a thorax phantom are used in the evaluation. Note that the head phantom consists of a real human skull. A floor-mounted Artis zeego C-arm system (Siemens Healthcare GmbH, Forchheim) is employed to acquire the 3-D scans of the phantoms. The voxel resolution of the 3-D volumes is $512 \times 512 \times 399$ and the voxel size is 0.485 mm in all three dimensions.

Unlike the approaches that optimize both the rigid-body transformation and the intrinsic calibration parameters [37]–[39], our approach assumes an accurately calibrated C-arm system. In order to assess the 3-D accuracy of our method, the calibration error can be eliminated using the digitally reconstructed radiograph (DRR) as simulated X-ray images. Using DRR as X-ray image is an established way of evaluating 2-D/3-D registration methods [28], [38], [40], because the ground-truth position is accurately given. The DRRs of the phantoms are rendered using the method described in [41], where the position parameters of the C-arm system are taken. Figure 4 shows the simulated X-ray images for the evaluation in this section. The images have the same pixel and spatial resolution as the X-ray images, i.e. 1240×960 with the pixel size of 0.308 mm in both dimensions. The regions of interest for the image registration are defined by rectangular bounding boxes.

Two experiments are performed using the pairs of simulated single-view X-ray images and the C-arm CT volumes, described as follows.

1) *Random Study*: Evaluation using randomly generated start positions is part of the standardized evaluation methodology for 2-D/3-D registration [42]. In our random study, the 3-D surface points of the salient structures are used as target points for the assessment of the registration accuracy. The 3-D mean target registration error (mTRE) is computed as the mean distance between the target points at the ground-truth position and the target points at the registration position [42]. The mean reprojection distance (mRPD) is computed as the mean

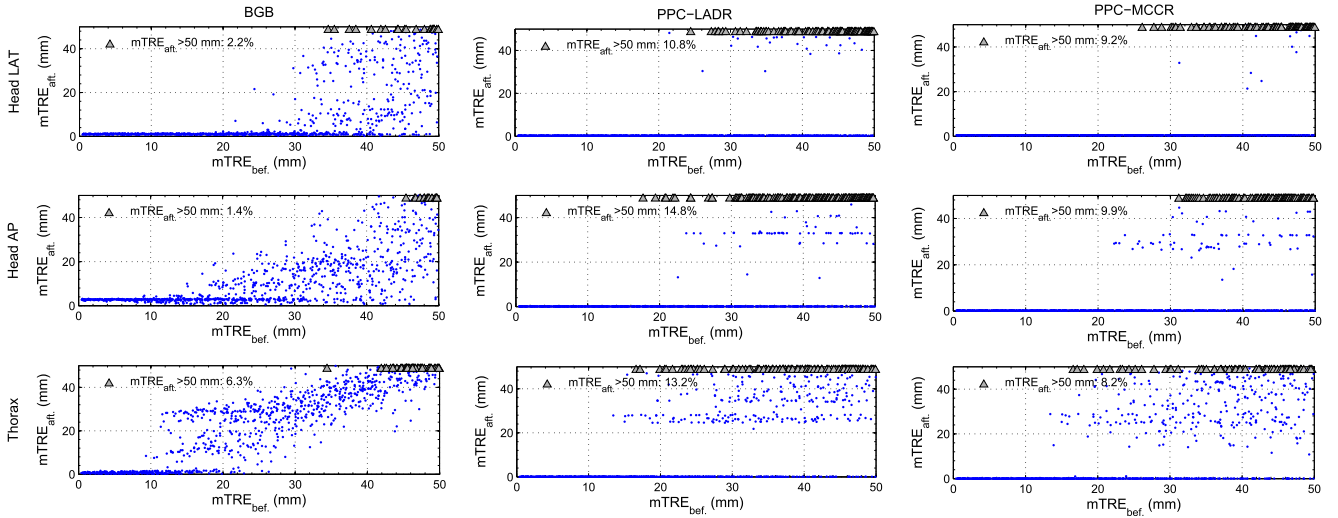


Fig. 5. Registration error plots of mTREs before and after the registration. Single-view registrations with 1000 start positions with initial mTRE up to 50 mm are evaluated using the head and the thorax phantoms. The reference method BGB [9] is compared with the PPC-based methods using LADR and MCCR estimation strategies. To keep a reasonable scale, the mTREs after registration larger than 50 mm are only shown by the gray triangle markers with their percentage over all random test cases.

distance between the 3-D target points at the ground-truth position and the lines from the X-ray source to the projections of the same points at the registration position [42]. The mTRE fundamentally assesses the actual 3-D accuracy, which is crucial when, e.g., the C-arm changes the working angulation, or the depth information is later used for motion compensation as in [13] and [29]. Therefore, we mainly use mTRE to assess the registration accuracy. As in [6], [37], and [42], mRPD is reported for the single-view registration, thus we also report the mRPD for comparison.

To generate the random start positions, we follow the standardized evaluation methodology proposed in [42]. For both the head and the thorax phantom, 1000 random start positions are generated with initial mTRE up to 50 mm and 20 random samples are generated for each 1 mm interval.

The success criterion is defined as $\text{mTRE} < 2$ mm, because misalignment within 2 mm is considered to be clinically plausible [43] and is chosen as success criterion in many works [9]–[11], [27], [37], [42], [44]. The success rate (SR) is computed as the percentage of the successful cases over all test cases. As suggested by [42], the capture range (CR) is defined as the first 1 mm sub-interval with SR less than 95%.

In the evaluation, both PPC-LADR and PPC-MCCR, are evaluated with the three image pairs (Fig. 4). To provide comparative evaluation, we also evaluated the reference method BGB proposed in [9]. In our implementation of BGB, gradient descent is used instead of Powell's method as in [9]. In order to be comparable to the proposed method, it is extended to a multi-resolution scheme with four resolution levels (scale-factor as 2). The maximum number of iterations is decreased from low to high resolution, namely 80 (lowest resolution), 40, 15 and 3 (full resolution).

The results are represented in Table I and Figure 5.

Table I shows the metrics assessing the registration accuracy (mTRE and mRPD) and the robustness (SR and CR). The mTRE and mRPD are only computed for the successful

TABLE I

RESULTS OF RANDOM STUDY USING SIMULATED X-RAY IMAGES OF HEAD AND THORAX PHANTOMS. FOR EACH IMAGE PAIR, 1000 RANDOM START POSITIONS WITH INITIAL mTRE UP TO 50 mm ARE EVALUATED. THE REFERENCE METHOD [9] IS COMPARED WITH THE PPC-BASED REGISTRATION USING LADR AND MCCR ESTIMATION STRATEGIES

Method	BGB [9]	PPC(LADR)	PPC(MCCR)
Head LAT, mTRE (mm)	1.03 ± 0.23	0.29 ± 0.01	0.46 ± 0.00
Head AP, mTRE (mm)	1.36 ± 0.35	0.07 ± 0.00	0.17 ± 0.00
Thorax, mTRE (mm)	0.75 ± 0.39	0.05 ± 0.00	0.04 ± 0.06
Head LAT, mRPD (mm)	0.21 ± 0.11	0.07 ± 0.00	0.05 ± 0.00
Head AP, mRPD (mm)	0.36 ± 0.16	0.05 ± 0.00	0.03 ± 0.00
Thorax, mRPD (mm)	0.16 ± 0.03	0.03 ± 0.00	0.03 ± 0.02
Head LAT, SR (%)	70.3	86.4	89.5
Head AP, SR (%)	7.3	79	82.9
Thorax, SR (%)	29.3	60.8	61.7
Head LAT, CR (mm)	23	27	27
Head AP, CR (mm)	0	20	22
Thorax, CR (mm)	10	15	13

registrations, shown with mean and standard deviations. For the three image pairs (Head LAT, Head AP and Thorax, see Fig. 4), our PPC-based registrations are superior in all metrics. The statistics on mTRE and mRPD show that our PPC-based registrations (PPC-LADR and PPC-MCCR) converge more accurately and consistently comparing to BGB. The t-Tests at 0.1% significance level (with p -value < 0.001) show the statistical significance for PPC-LADR vs. BGB and for PPC-MCCR vs. BGB.

Comparing to PPC-LADR, PPC-MCCR achieves higher SRs, but does not show advantages in other metrics. This is due to the fact that the outliers are limited in the simulated X-ray images. For Head LAT, both BGB and PPC (LADR and MCCR) achieve relatively higher SR ($\geq 70\%$) compared to the other two image pairs. However, for PPC-based registrations, the 3-D accuracy ($\text{mTRE} = 0.29$ mm / 0.46 mm) is lower as the other two image pairs ($\text{mTRE} < 0.1$ mm). The t-Tests prove

the statistical significance of the above comparison. Head LAT has relatively simple apparent contours, which have limited diversity in depth. The main reason of a lower SR and CR for the thorax is that repetitive rib structures are present in the phantom which induce false correspondences.

Figure 5 shows the plots of the relationship between mTREs before registration ($mTRE_{bef.}$) and mTREs after registration ($mTRE_{aft.}$) of BGB [9], PPC-LADR and PPC-MCCR on the three image pairs (Head LAT, Head AP and Thorax, see Fig. 4). To ensure a reasonable scale on both axes, only the registrations with $mTRE_{aft.} \leq 50$ mm are shown in the plots. The percentage of the left-out cases over all test cases is given with the plots. Comparing to BGB [9], the PPC-based registrations show the following distinctive characteristics: 1) it converges significantly more consistently and accurately than BGB if a registration is successful; 2) The margin between failed and successful registrations is much clearer compared to BGB. The consistent convergence and the clear success/failure margin can provide useful input for extensions, e.g., to recognize failed registrations.

2) Convergence Analysis: The strength of the PPC model is that the correspondence planes efficiently constrain both planar and non-planar transformations. In the random study, we have also observed that the PPC-based registration converges already after a few iterations. In order to demonstrate the efficiency of the PPC model, a comparative convergence analysis is performed using the simulated head phantom image.

In each registration iteration, the step of measuring 3-D/2-D misalignment (Sect. IV-C) gives the “correspondences” between the 3-D contour generator points and the 2-D apparent contour points. As an alternative to the PPC model, these point correspondences can also serve as input to the EPnP proposed by Lepetit et al. [21]. In our registration framework, the step of PPC-based motion estimation (Sect. IV-D) can be replaced by the EPnP-based motion estimation. For comparison, the convergence behavior of our PPC-based registration is compared with the EPnP-based registration. The publicly available implementation of EPnP [21] is directly used to replace the PPC-based motion estimation, later noted as DGP-EPnP.

For demonstration purpose, both the PPC-LADR registration and the DGP-EPnP registration are performed with 20 iterations per start position under the full resolution. As test image data, the Head AP image pair (Fig. 4) is used and the iteration results are stored for the first 100 random start positions.

The comparative results are shown in Fig. 6, where the registration errors ($mTRE_{aft.}$ and $mRPD_{aft.}$) are normalized with the errors before the registration ($mTRE_{bef.}$ and $mRPD_{bef.}$) to reflect the error evolution and the convergence of the registration. Each curve shows the error evolution within 20 iterations of the corresponding start position and 100 curves are plotted for all the test start positions. For both the EPnP-based registration and the PPC-based registration, enlarging of the registration error at the first iterations is involved in mTRE evolution while mRPD is reduced along the iterations. It indicates the fact that the out-of-plane transformation is sometimes

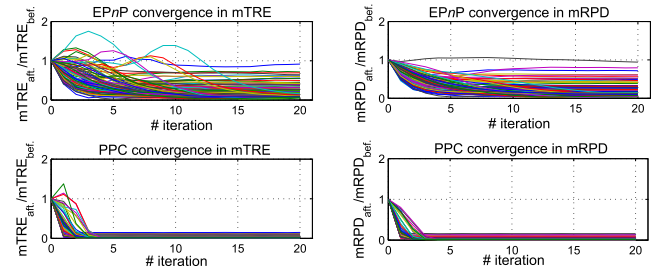


Fig. 6. The convergence behaviors of EPnP and PPC within 20 iterations for 100 start positions. Each curve represents the rate between the registration error after the iteration with the initial registration error. PPC shows an efficient and consistent convergence behavior compared to EPnP.

“over-corrected” in the beginning to reduce the projection error. The mTRE and mRPD of DGP-EPnP registration converge from iteration 3 to approximately iteration 15, and the rates $mTRE_{aft.}/mTRE_{bef.}$ and $mRPD_{aft.}/mRPD_{bef.}$ are spread between zero and one. The mTRE and mRPD of the PPC-LADR registration converge within five iterations for all tested start positions and the rates $mTRE_{aft.}/mTRE_{bef.}$ and $mRPD_{aft.}/mRPD_{bef.}$ are strongly and consistently reduced. It demonstrates the strength of the PCC model in efficiently constraining both planar and non-planar transformations, especially in the X-ray image registration scenario, where direct 2-D/3-D point correspondences are usually not available.

B. Evaluation on X-Ray Sequences With Continuous Motion

In order to demonstrate the benefits of our extension of the PPC model from [13], the frame-wise PPC-based dynamic registration is compared with the differential-tracking-based (DT) motion compensation in [13].

The evaluation is performed using the datasets acquired for DT-based motion compensation [13]. The experiment setup is as follows. The initial 2-D/3-D registration was derived from a calibrated C-arm system, and no position change occurred 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 attached to the phantom during the 2-D acquisition. As shown in Fig. 7, an OptoTrak Certus motion capture system was employed to track the motion of the markers attached on the phantom as “ground truth” for the evaluation. The OptoTrak Certus is a research-grade motion capture system with an accuracy of up to 0.1 mm and resolution of 0.01 mm. The markers were attached on the borders of the phantom, such that no markers were visible in the X-ray image sequences. The motion sequence $\{\mathbf{T}^{opt}\}$ is acquired in the OptoTrak coordinate system, see Fig. 7. The OptoTrak system is registered with the C-arm system using the markers that are visible in both the OptoTrak and the C-arm CT. The OptoTrak motion sequences are synchronized with the X-ray image sequences manually. Ten sequences consisting of 1010 monoplanar fluoroscopic images of the thorax phantom were acquired for the evaluation. Over all sequences, the maximum structure shift in the 2-D projection varies from 17.3 mm to 33.2 mm [13]. Note that experiment setup involves systematic errors including the



Fig. 7. The experiment setup for dynamic registration using the thorax phantom [13]. An interventional C-arm system is used to acquire both 2-D image sequence and 3-D C-arm CT volume. A CT volume is available for thorax phantom. The motion of the thorax phantom was triggered manually. The OptoTrak motion capture system is used to acquire the “ground truth” [13].

C-arm calibration error and the OptoTrak to C-arm registration error. We consider the precision of the “ground truth” as acceptable for comparative analysis since the accumulated motion introduces large misalignment and the differential tracking [13] also causes a non-neglectable registration error.

In this evaluation, the DGP-EPNP is also evaluated by replacing the PPC motion estimation within our dynamic registration framework. For the purpose of motion compensation, the registration result from the last frame is used as initial pose in the next frame and only one registration iteration is performed. As baseline, the results of differential 2-D tracking (DT) approach [13] is repeated in this evaluation. DT-based motion compensation sets up the 3-D/2-D correspondences from the initial frame based on the accurate initial registration. The feature points are tracked using 2-D optical flow along the frames as described in [13].

Figure 8(a) and 8(b) show the motion compensation results of two example image sequences (Sequence 6 and 10). The plots show the motion curves of respectively the rotation ($^\circ$) and translation (mm) components along the frames. R^* and t^* are the “ground-truth” motion curves captured by OptoTrak. R_{PPC} and t_{PPC} are the motion curves estimated by PPC-LADR (1 iteration per frame). R_{EPnP} and t_{EPnP} are the motion curves estimated by DGP-EPNP (1 iteration per frame). R_{DT} and t_{DT} are the motion curves estimated by the baseline approach DT [13]. Since the proposed framework involves 3-D updates and the updates of the correspondences, the 3-D motion is more accurately estimated compared to the baseline method, especially regarding the off-plane rotation ($R(z)$).

Figure 8(c) shows the resulting statistics of all image sequences. The plots show the means and standard deviations of mTRE and mPE (mean projection error) of all frames in all ten sequences. As a test acquisition, sequence 1 contains large error in ground-truth measurement and is not involved in the further discussion. Compared to the baseline algorithm [13], the proposed method generally improves the accuracy among all frames. As already demonstrated by the motion curve plots, the proposed method shows significant improvement in 3-D accuracy. In 2-D registration error plot, both PPC- and

EPnP-based estimation (resp. PM_{PPC} and PM_{EPnP}) delivers comparable results. However, in 3-D registration error plot, the PPC-based estimation outperforms EPnP-based estimation over all sequences except the test acquisition sequence.

Despite of the systematic error in the experiment, the quantitative results still reflect the performance difference of different approaches. Over all 1010 frames, the mean and standard deviation of the mTREs are respectively $mTRE_{DT} = 3.26 \pm 2.76$ mm, $mTRE_{PPC} = 1.48 \pm 0.73$ mm and $mTRE_{EPnP} = 1.68 \pm 0.88$ mm. The t-Tests show the statistical significance (significance level of 0.1%) for DT vs. PPC and for PPC vs. EPnP.

The above evaluation demonstrates the significant benefit of our method against the accumulated systematic error stated in [13]. It also shows that our method can precisely compensate 3-D motion with only one iteration per frame. Moreover, unlike the DT-based motion compensation [13], our dynamic registration does not rely on an accurate initial alignment. As demonstrated in Sect. VI-A.1, it can establish an accurate alignment in the initial registration stage by iteratively updating the transformation and the correspondences. Note that the small angle assumption ($< 15^\circ$) needs to hold for each registration iteration.

C. Quantitative Evaluation on Clinical Datasets With “Gold-Standard” Registration

In order to quantitatively evaluate the accuracy and robustness of the proposed registration framework, experiments on clinical datasets with highly accurate “gold-standard” registrations are performed.

There are several publicly available “gold-standard” datasets for 2-D/3-D image registration, e.g. the lumbar vertebrae dataset proposed by Tomažević et al. [45], the porcine head phantom proposed by Pawiro et al. [46] and the cerebral angiograms proposed by Mitrović et al. [44] and Madan et al. [47]. We used the cerebral angiogram datasets [44], [47] in the evaluation of our registration framework under the following considerations. Firstly, endovascular treatment of cerebrovascular diseases (e.g. cerebral aneurysm and arteriovenous malformation) is an important application of interventional C-arm systems and 2-D/3-D image fusion applications [2]. The cerebral angiogram datasets [44], [47] contain ten datasets acquired during actual intervention, including cerebral aneurysm and arteriovenous malformation cases. Secondly, each angiogram dataset contains a digitally subtracted 3-D rotational angiography (3D-DSA) volume, 2-D digitally subtracted angiography (2D-DSA) as well as native fluoroscopic images (2D-MAX) from anterior-posterior (AP) and lateral (LAT) views. It provides the possibility to evaluate both the accuracy and robustness of an algorithm; Furthermore, the initial registrations and the evaluation framework with standardized evaluation methodology [42] are also publicly available. Several state-of-the-art methods are also evaluated on this dataset [11], [44]. It eases the standardized evaluation and comparison to the state-of-the-art.

Given the angiogram datasets, the following aspects are considered in the experiments:

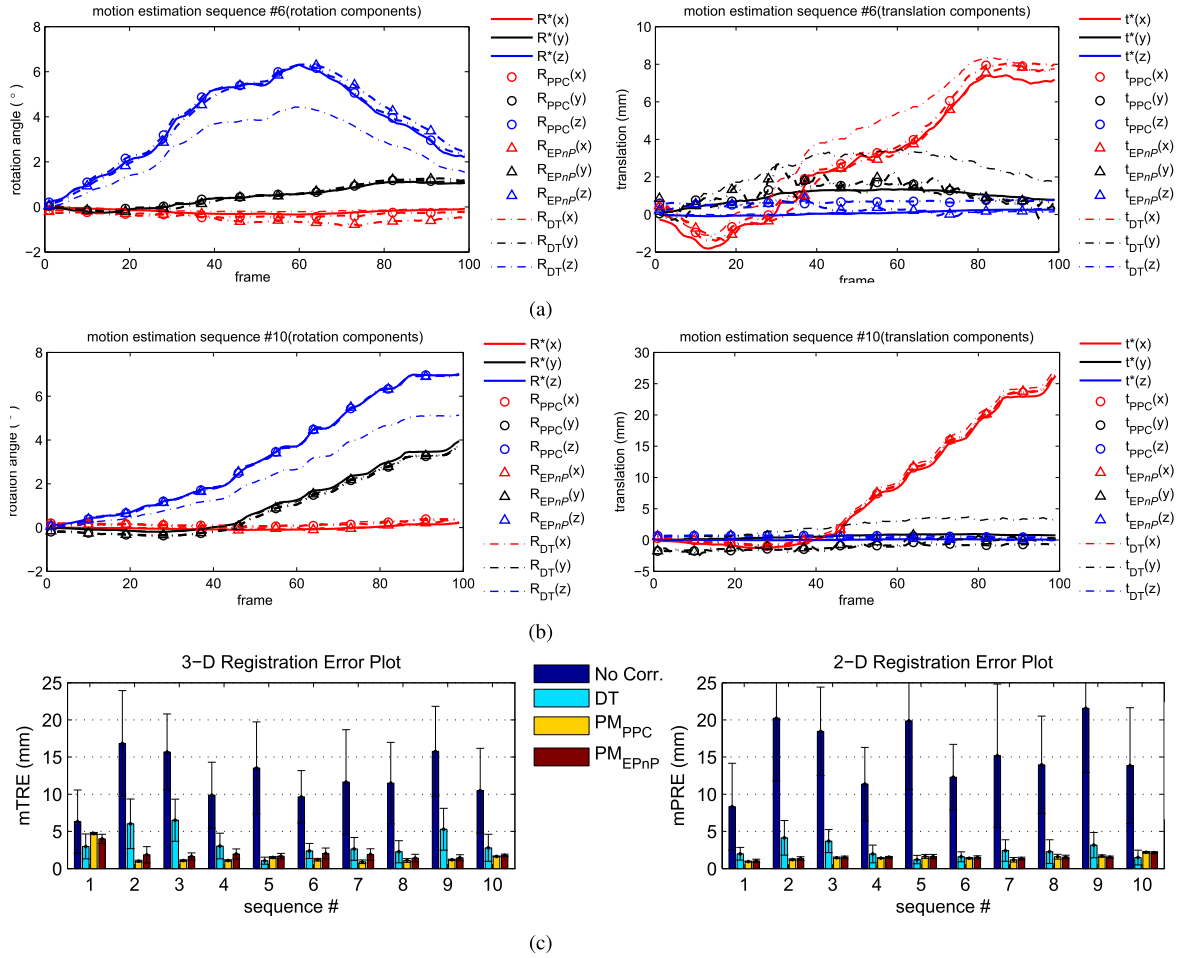


Fig. 8. Evaluation of the dynamic registration for motion tracking and compensation. (a) Sequence 6, comparisons of motion curves estimated by PCC-based dynamic registration, EPnP [21] registration and differential tracking [13] (DT). A sequence of fluoroscopic images is given, where the phantom was in continuous movement. The rotation (left) and translation (right) components of the motion sequence are compared along all frames. R^* and t^* show the ground truth motion curves acquired by external motion sensor (described in [13]). R_{PPC} and t_{PPC} show the motion curves computed by frame-wise registration using the proposed approach. R_{EPnP} and t_{EPnP} show the motion curves computed by frame-wise registration using EPnP [21] instead of PPC. R_{DT} and t_{DT} show the motion curves computed by tracking [13] given the initial alignment. (b) Sequence 10, comparisons of motion curves estimated by our proposed PCC dynamic registration, EPnP [21] registration and differential tracking [13] (DT). (c) Comparison of the registration errors for dynamic registration. The plots show respectively the 3-D and 2-D mean registration errors with standard deviation of the ten image sequences. The errors without correction (No Corr.) are shown in blue bars. The results of the differential 2-D tracking (DT) [13] provides the comparison baseline (cyan). The results of the proposed method using PPC (PM_{PPC}) and EPnP (PM_{EPnP}) are respectively shown in yellow and red. Details refer to Sect. VI-B.

The 2D-DSA and 3D-DSA image pairs contain only the cerebral vessel tree, which is relatively ideal for image registration algorithms. We mainly used the DSA image pairs to assess the registration accuracy.

The 2D-MAX images contain both the contrasted vessel tree and the bone structures. The additional bone structures introduce disturbances that strongly challenge the robustness of the registration algorithms. Thus, the 2D-MAX and 3D-DSA image pairs are used to assess the robustness.

The PPC-LADR and PPC-MCCR are the main methods in the evaluation. Both of the variations are performed on the 2D-DSA and 2D-MAX images, in the single-view scenario. The DGP-EPnP and GP-PPC are variations for comparison purpose. Both of the variations are only performed on the 2D-DSA images.

The evaluation is done as suggested in [44]. For each dataset, 400 random start positions are provided [44], which were generated by following the methodology of van de

Kraats et al. [42]. The translations and rotations were randomly sampled in the range of $[-20, 20]$ mm and $[-10, 10]$ degrees, in order to generate 20 initial displacements per 1 mm mTRE interval up to 20 mm of mTRE [44].

The registration results are shown in Tab. II. In order to give a direct comparison, the results published in citevspiclin2014fast and [44] are repeated in the same table. The results are discussed from the following aspects:

1) **Single-View 2D-DSA:** In this scenario, all our four variants are evaluated, i.e. PPC-LADR, PPC-MCCR, DGP-EPnP and GP-PPC. For the DSA case, the outliers are limited because both 2-D and 3-D images contain only the cerebral vessel tree. In this case, MCCR does not show the advantage in accuracy (mTRE and mRPD), but already shows better SRs and CRs that reflect robustness. For AP view, both PPC-LADR and PPC-MCCR outperform DGP-EPnP and GP-PPC in mTRE, SR and CR. For LAT view, DGP-EPnP shows a better mTRE (0.86 ± 0.57 mm) comparing to PPC-LADR

TABLE II

SUMMARY OF THE REGISTRATION RESULTS USING THE “GOLD STANDARD” DATASETS [44]. THE MEAN AND STANDARD DEVIATION (std) OF mTRE AND mRPD OF SUCCESSFUL REGISTRATIONS, SUCCESS RATE (SR) AND CAPTURE RANGES (CR) OVER ALL TEN DATASETS USING DIFFERENT OPTIMIZATION STRATEGIES/CORRESPONDENCE MODELS ARE LISTED, I.E. PPC WITH LADR OPTIMIZATION (PPC-LADR), PPC WITH MCCR OPTIMIZATION (PPC-MCCR), DGP MISALIGNMENT MEASURE WITH EPnP (DGP-EPNP) AND THE GRADIENT PROJECTION WITH PPC (GP-PPC, NO MULTIPLE DEPTH INTERVALS). THE RESULTS PUBLISHED IN [11] AND [44] ARE ALSO REPEATED HERE FOR COMPARISON. THE BOLD VALUES ARE THE BEST RESULTS OF EACH EVALUATION GROUP. THERE ARE BLANK CELLS MARKED WITH “-” DUE TO THE FOLLOWING SITUATIONS: DGP-EPNP AND GP-PPC ARE NOT EVALUATED ON 2D-MAX IMAGES, THE mTREs ARE NOT REPORTED FOR SINGLE-VIEW REGISTRATION IN [11] AND [44]

View	Method	mTRE (mean±std) [mm]		mRPD (mean±std) [mm]		SR [%]		CR [mm]	
		DSA	MAX	DSA	MAX	DSA	MAX	DSA	MAX
AP	PPC-LADR	0.57 ± 0.38	0.76 ± 0.46	0.15 ± 0.08	0.15 ± 0.08	97.25	82.23	16	0
	PPC-MCCR	0.59 ± 0.27	0.68 ± 0.32	0.16 ± 0.08	0.16 ± 0.08	99.38	94.68	20	13
	DGP-EPNP [21]	0.90 ± 0.49	-	0.16 ± 0.07	-	93.78	-	13	-
	GP-PPC	0.79 ± 0.45	-	0.19 ± 0.13	-	93.73	-	9	-
	MGP+BGB [44]	-	-	0.28 ± 0.19	0.37 ± 0.29	95.45	86.8	12	8
LAT	PPC-LADR	0.91 ± 0.50	1.03 ± 0.50	0.22 ± 0.08	0.26 ± 0.07	95.60	74.78	16	0
	PPC-MCCR	0.64 ± 0.31	0.73 ± 0.43	0.23 ± 0.08	0.24 ± 0.08	98.30	91.75	18	11
	DGP-EPNP [21]	0.86 ± 0.57	-	0.25 ± 0.08	-	85.15	-	0	-
	GP-PPC	1.07 ± 0.47	-	0.25 ± 0.09	-	95	-	11	-
	MGP+BGB [44]	-	-	0.28 ± 0.21	0.31 ± 0.25	79.45	75.23	6	4
	PB-BGC [11]	-	-	0.51 ± 0.29	-	82.2	-	9	-
	QA-BGC [11]	-	-	0.42 ± 0.25	-	78.2	-	8	-

(0.91±0.50 mm), but the SR (85.15%) and CR (0 mm) are the lowest compared to the other three variants. The reason is that if DGP-EPNP converges to the “gold-standard” position, the point-*n*-point correspondences and the point-to-plane correspondences are both true. But the point-*n*-point correspondence model has less tolerance against the wrong correspondences and does not lead to consistent convergence, also reflected by Fig. 6. The results between PPC-LADR and GP-PPC were also compared, where the only difference was that GP-PPC did not separate the depth intervals. As shown in Tab. II, PPC-LADR delivers better registration results as GP-PPC. The t-Tests on mTREs between PPC-LADR and GP-PPC of AP and LAT views pass at 0.1% significance level.

The single-view accuracy with the angiogram datasets in [11] and [44] is assessed with mRPD instead of mTRE. The mRPDs of our evaluation are also computed to compare with the state-of-the-art methods evaluated with the same datasets. The t-Tests (at 0.1% significance level) show significant improvement of our proposed methods compared to the state-of-the-art methods [11], [44]. Both PPC-LADR and PPC-MCCR deliver higher SR and CR compared to the state-of-the-art methods. PPC-MCCR even achieves CR=20 mm for AP view, which is the upper limit of the randomly generated start positions [44]. The high CR proves the contribution of our extension on the PPC model from differential tracking to general 2-D/3-D registration.

The mTRE for single-view 2-D/3-D registration is also reported in [37]. Although a different dataset was used for the evaluation in [37], an indirect comparison is still possible because their test datasets are also cerebral DSA images. By simultaneous registration and calibration, [37] have achieved the mTRE as 0.64 mm with SR of 93.3% over 30 pairs of 3-D/2-D DSA images. Our proposed registration method (PPC-MCCR) achieves (indirectly) comparable performance, i.e. mTRE for AP view as 0.59 mm with SR of 99.38% and mTRE for LAT view as 0.64 mm with SR of 98.30%.

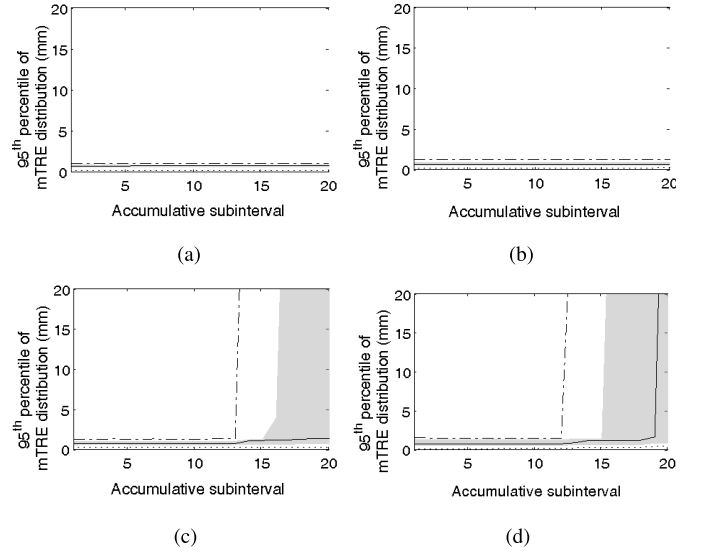


Fig. 9. Registration accuracy plots of PPC-MCCR. The registration accuracy is defined as 95th percentile of mTRE distribution for each of accumulative sub-intervals of initial mTRE displacements [48]. Similar plots for the same datasets of the MGP method are available in [44]. The minimal, median and maximal accuracies in each sub-interval were calculated across all ten clinical datasets and is represented by dash-dotted, solid and dotted lines, respectively and the grey area denotes the range of registration accuracies between first and third quartile. [44]. Note that only mTREs are used here to assess the registration accuracy. (a) DSA AP view. (b) DSA LAT view. (c) MAX AP view (d) MAX LAT view.

Note that the PPC-MCCR does not consider the calibration correction.

2) *Single-View 2D-MAX*: The evaluation on single-view 2D-MAX images is done between PPC-LADR and PPC-MCCR. As already stated, both the vessel tree and the bone structures (skull) are present in the 2D-MAX images. Compared to DSA images, the presence of bone structures introduces strong disturbances and challenges the robustness of the registration algorithms. For our PPC-based registration,

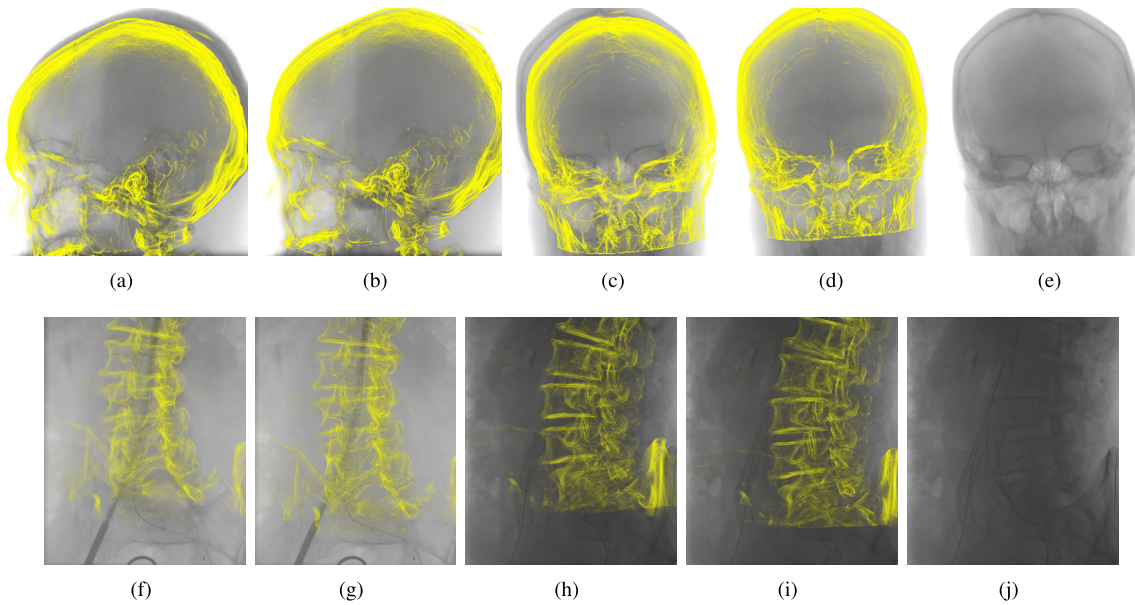


Fig. 10. Clinical evaluation on a cerebral aneurysm case (a-e) and an abdominal aortic aneurysm case (f-j). Registrations are based on single-view X-ray images (a,b) and (f,g). Quasi-perpendicular views (c,d,e) and (h,i,j) show the registration accuracy.

the wrong correspondences found during misalignment measure introduce outliers to the motion estimation. The t-Tests at 0.1% significance level show statistically significant improvement of PPC-MCCR compared to PPC-LADR using single-view 2D-MAX images. Note that, although PPC-LADR on single-view 2D-MAX images delivers SRs of 82.23% for AP view and 74.78% for LAT view, but it failed in CRs (0 mm) for both views over the ten datasets. Meanwhile, PPC-MCCR still achieves to maintain the SRs for single-view 2D-MAX images (94.68% for AP and 91.75% for LAT). The CRs are also over 10 mm for both views (13 mm for AP and 11 mm for LAT). Regarding to SR and CR, the single-view PPC-MCCR even outperforms all evaluated state-of-the-art methods evaluated in [44] using two X-ray images (AP-LAT). The results show the significant robustness of the MCCR estimation strategy comparing to LADR.

To provide an alternative comparison to the state of the art, Fig. 9 shows the accuracy plots of the PPC-MCCR method in accumulative sub-intervals, for both single-view DSA and MAX images. The accuracy plot is suggested in the standardized evaluation methodology by Markelj et al. [48]. To show the registration, the range of mTREs before registration is divided into accumulated sub-intervals, i.e. 0-1 mm, 0-2 mm,..., 0-20 mm for the “gold-standard” angiogram evaluation. For each sub-interval, the accuracy is represented by the 95th percentile of mTREs after the registration [44]. Similar plots are found for MGP method in [44]. Note that only mTREs are used in our plots. For evaluation on DSA images, the 95th percentile curves of PPC-MCCR method consistently below 2 mm over all sub-intervals. For the evaluation on MAX images, the curves stay under 2 mm up to certain ranges (about to 12 mm sub-interval) and go steeply up in the next sub-interval. It again confirms the observation from Fig. 5 that the PPC-based registration has a clear margin between failed and successful registrations.

The quantitative evaluation on “gold-standard” angiogram datasets confirms the strength of the proposed registration framework from the following perspectives:

The high potential in single-view registration by extending the PPC model from differential tracking [13] to dynamic registration: although the PPC model is derived under the small motion assumption, but PPC shows the strength in single-view registration with the iterative 3-D update provided by the direct 3-D/2-D misalignment measure;

PPC model better handles bad correspondences or outliers than PnP in X-ray image registration scenario: As already demonstrated in Fig. 6, the evaluation confirms that the PPC model is better suited in X-ray image registration compared to EPnP;

Improved accuracy and robustness by using depth-aware gradient projections: Derived from the depth-layer concept [29], DGP reduces the overlapping of structures in depth and outperforms the standard gradient projection (GP) in our registration framework;

MCCR as the more robust strategy compared to LADR with presence of disturbances: The clear improvements in accuracy and robustness provided by PPC-MCCR comparing to PPC-LADR on 2D-MAX images confirms the statement in Sect. IV-D, showing MCCR as a more robust estimation method compared to LADR.

D. Qualitative Evaluation on Other Clinical Datasets

In the further evaluation using clinical datasets, we used the datasets from two types of endovascular treatments, i.e. cerebral aneurysm (CA) and abdominal aortic aneurysm (AAA). Evaluations are carried out in terms of accuracy and robustness. Since no ground truth is available for the clinical data, the accuracy is assessed qualitatively using the X-ray images taken from other views during the procedures. The

robustness is assessed by the registration consistency along image sequences, where no patient movement but external disturbances are present, e.g. moving external devices and injected contrast medium.

Figure 10 shows two examples respectively of the CA and AAA cases. Registrations are performed using single-view images, i.e. Fig. 10(a) for the CA dataset and 10(f) for the AAA dataset. Images acquired from other directions are used to assess the out-of-plane registration error, i.e. Fig. 10(c) for CA and 10(h) for AAA. In this example, the initial registration contains both in-plane and out-of-plane errors. As view pairs (Fig. 10(b), 10(d)) and (Fig. 10(g), 10(i)) demonstrate, the registration estimates accurately the 3-D transformation (especially in depth), even though only single-view images were used. The fluoroscopic images of the reference views are explicitly shown in Fig. 10(e) and 10(j).

To assess the robustness, random tests are performed independently through all frames of the AAA sequences that are acquired during the injection of contrast medium. In those AAA datasets, no patient movement was present during the acquisition, but external disturbances were present in the images, e.g. flow of injected contrast medium and moving interventional devices. For each image sequence, the registration method is supposed to produce consistent results despite of those external disturbances. Therefore, we compute for each sequence the variance of the registration difference regarding to the result of the first frame. For all the three tested AAA datasets, the variances of the registered 3-D targets (structures of interest) are respectively 0.59 ± 0.31 mm, 0.15 ± 0.05 mm and 0.19 ± 0.11 mm comparing to the first frames. The consistency of the registration result indicates the high robustness against external disturbances.

VII. CONCLUSION

In this paper, we proposed a dynamic 2-D/3-D rigid registration framework that sets up the initial registration and dynamically compensates patient motion. Using the novel point-to-plane correspondence (PPC) model, the dynamic registration framework is capable of (but not limited to) recovering 3-D motion out of single-view 2-D images. Structures of interest are extracted in 3-D as sparse input. The registration iterates between two steps until convergence: 1) Given the updated registration, contour generator points are selected and projected with their 3-D gradients and depths preserved. DGPs of the volume are also computed; 2) the 3-D/2-D misalignment is measured for robust estimation of the registration update using the PPC model. The results of phantom studies show the remarkable strength of our approach in 3-D accuracy using single-view X-ray images. The convergence analysis indicates the high efficiency of the motion constraint provided by the PPC model. Comparing to the 2-D-tracking-based motion compensation, our dynamic registration shows significant improvement in accuracy. In addition, unlike the differential-tracking-based motion compensation, our dynamic registration does not rely on an accurate initial alignment as the DT motion compensation [13]. As demonstrated in Sect. VI-A.1, it can establish an accurate alignment in the initial registration

stage by iteratively updating the transformation and the correspondences with holding the small angle ($< 15^\circ$) assumption within each iteration.

The quantitative and qualitative clinical evaluations confirm the 3-D accuracy and robustness, showing the clinical value of our method. Especially, the quantitative evaluation using the clinical angiogram datasets with “gold-standard” registrations demonstrate the strength of the method compared to the state-of-the-art methods in single-view registration. Our approach is intuitive and generic, and can potentially be applied in many relevant clinical applications. Moreover, its design allows for straight forward parallelization. A preliminary GPU implementation reduces the registration time to 80 ms per iteration [49]. Considering that the frame rate of fluoroscopy is usually 5 to 10 Hz and only one iteration is required in dynamic motion compensation, our dynamic registration can compensate the patient movement in real time.

VIII. FUTURE WORK

The PPC model that constrains the 3-D motion updates relies on the differential approximation of the motion model. Strategies as global multi-start [15] and constrained exhaustive search [17] can be considered to increase the convergence range and the success rate with large initial misalignment. More realistic piece-wise rigid motion model can be considered within the PPC-based registration framework. The PPC model can be also extended to the multiple-view scenario in a way that the observations are jointly optimized. To further enhance the robustness, the selection of contour generator points can be improved by considering neighborhood condition (highly curved or flat region) and distribution condition (whether points are evenly distributed in 3-D space). Similar as [37] and [38], simultaneous optimization of the intrinsic calibration parameters can be embedded into the dynamic registration framework to ensure the accuracy in case that an accurate calibration is not available, e.g. for a mobile C-arm or a system based on an image-intensifier.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their constructive comments that improved significantly the quality of this paper.

DISCLAIMER

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

REFERENCES

- [1] J. Wang, M. Kreiser, L. Wang, N. Navab, and P. Fallavollita, “Augmented depth perception visualization in 2D/3D image fusion,” *Comput. Med. Imag. Graph.*, vol. 38, no. 8, pp. 744–752, 2014.
- [2] S. Rossitti and M. Pfister, “3D road-mapping in the endovascular treatment of cerebral aneurysms and arteriovenous malformations,” *Intervent. Neuroradiol.*, vol. 15, no. 3, pp. 283–290, 2009.
- [3] N. Navab, “C-arm calibration method for 3D reconstruction,” U.S. Patent 6049582 A, Apr. 11, 2000.
- [4] K. Wiesent *et al.*, “Enhanced 3-D-reconstruction algorithm for C-arm systems suitable for interventional procedures,” *IEEE Trans. Med. Imag.*, vol. 19, no. 5, pp. 391–403, May 2000.

- [5] H. Livyatan, Z. Yaniv, and L. Joskowicz, "Robust automatic C-arm calibration for fluoroscopy-based navigation: A practical approach," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2002, pp. 60–68.
- [6] 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, 2012.
- [7] R. Liao, L. Zhang, Y. Sun, S. Miao, and C. Chéd'Hotel, "A review of recent advances in registration techniques applied to minimally invasive therapy," *IEEE Trans. Multimedia*, vol. 15, no. 5, pp. 983–1000, Aug. 2013.
- [8] A. Kubias, F. Deinzer, T. Feldmann, and D. Paulus, "Extended global optimization strategy for rigid 2D/3D image registration," in *Computer Analysis of Images and Patterns* (Lecture Notes in Computer Science), W. G. Kropatsch, M. Kampel, and A. Hanbury, Eds., vol. 4673. Berlin, Germany: Springer, 2007, pp. 759–767.
- [9] D. Tomaževič, B. Likar, T. Slivnik, and F. Pernuš, "3-D/2-D registration of CT and MR to X-ray images," *IEEE Trans. Med. Imag.*, vol. 22, no. 11, pp. 1407–1416, Nov. 2003.
- [10] P. Markelj, D. Tomaževič, F. Pernuš, and B. Likar, "Robust gradient-based 3-D/2-D registration of CT and MR to X-ray images," *IEEE Trans. Med. Imag.*, vol. 27, no. 12, pp. 1704–1714, Dec. 2008.
- [11] Ž. Špiclin, B. Likar, and F. Pernuš, "Fast and robust 3D to 2D image registration by backprojection of gradient covariances," in *Proc. 6th Int. Workshop Biomed. Image Registration*, 2014, pp. 124–133.
- [12] J. Schmid and C. Chênes, "Segmentation of X-ray images by 3D-2D registration based on multibody physics," in *Proc. 12th Asian Conf. Comput. Vis.*, Nov. 2014, pp. 674–687.
- [13] J. Wang, A. Borsdorf, B. Heigl, T. Köhler, and J. Hornegger, "Gradient-based differential approach for 3-D motion compensation in interventional 2-D/3-D image fusion," in *Proc. 2nd Int. Conf. 3D Vis. (3DV)*, vol. 1, Dec. 2014, pp. 293–300.
- [14] M. Berger *et al.*, "Marker-free motion correction in weight-bearing cone-beam CT of the knee joint," *Med. Phys.*, vol. 43, no. 3, pp. 1235–1248, 2016.
- [15] Y. Otake *et al.*, "Robust 3D–2D image registration: Application to spine interventions and vertebral labeling in the presence of anatomical deformation," *Phys. Med. Biol.*, vol. 58, no. 23, p. 8535, 2013.
- [16] N. Hansen, "The CMA evolution strategy: A comparing review," in *Towards a New Evolutionary Computation*. Berlin, Germany: Springer, 2006, pp. 75–102.
- [17] A. Duménil, A. Kaladji, M. Castro, C. Göksu, A. Lucas, and P. Haigron, "A versatile intensity-based 3D/2D rigid registration compatible with mobile C-arm for endovascular treatment of abdominal aortic aneurysm," *Int. J. Comput. Assist. Radiol. Surgery*, vol. 11, no. 9, pp. 1713–1729, 2016.
- [18] T. Rohlfing, J. Denzler, C. Grassl, D. B. Russakoff, and C. R. Maurer, Jr., "Markerless real-time 3-D target region tracking by motion backprojection from projection images," *IEEE Trans. Med. Imag.*, vol. 24, no. 11, pp. 1455–1468, Nov. 2005.
- [19] A. Aichert, J. Wang, R. Schaffert, A. Dörfler, J. Hornegger, and A. Maier, "Epipolar consistency in fluoroscopy for image-based tracking," in *Proc. Brit. Mach. Vis. Conf.*, Swansea, U.K., 2015, pp. 82.1–82.10.
- [20] A. Aichert *et al.*, "Epipolar consistency in transmission imaging," *IEEE Trans. Med. Imag.*, vol. 34, no. 11, pp. 2205–2219, Nov. 2015.
- [21] V. Lepetit, F. Moreno-Noguer, and P. Fua, "EPnP: An accurate $O(n)$ solution to the PnP problem," *Int. J. Comput. Vis.*, vol. 81, no. 2, p. 155, 2009.
- [22] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2003.
- [23] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 6, pp. 679–698, Nov. 1986.
- [24] J. Wasza, S. Bauer, and J. Hornegger, "High performance GPU-based preprocessing for time-of-flight imaging in medical applications," in *Bildverarbeitung für die Medizin*. Berlin, Germany: Springer, 2011, pp. 324–328.
- [25] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013.
- [26] P. Rheingans and D. Ebert, "Volume illustration: Nonphotorealistic rendering of volume models," *IEEE Trans. Vis. Comput. Graphics*, vol. 7, no. 3, pp. 253–264, Jul./Sep. 2001.
- [27] H. Livyatan, Z. Yaniv, and L. Joskowicz, "Gradient-based 2-D/3-D rigid registration of fluoroscopic X-ray to CT," *IEEE Trans. Med. Imag.*, vol. 22, no. 11, pp. 1395–1406, Nov. 2003.
- [28] W. Wein, B. Röper, and N. Navab, "2D/3D registration based on volume gradients," *Proc. SPIE*, vol. 5747, pp. 144–150, May 2005, doi: 10.1117/12.595466.
- [29] J. Wang, A. Borsdorf, and J. Hornegger, "Depth-layer-based patient motion compensation for the overlay of 3D volumes onto X-ray sequences," in *Bildverarbeitung für die Medizin*. Berlin, Germany: Springer, 2013, pp. 128–133.
- [30] R. Koenker, *Quantile Regression*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
- [31] P. W. Holland and R. E. Welsch, "Robust regression using iteratively reweighted least-squares," *Commun. Statist.-Theory Methods*, vol. 6, no. 9, pp. 813–827, 1977.
- [32] I. Santamaría, P. P. Pokharel, and J. C. Principe, "Generalized correlation function: Definition, properties, and application to blind equalization," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2187–2197, Jun. 2006.
- [33] W. Liu, P. P. Pokharel, and J. C. Principe, "Correntropy: Properties and applications in non-Gaussian signal processing," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5286–5298, Nov. 2007.
- [34] R. He, W.-S. Zheng, and B.-G. Hu, "Maximum correntropy criterion for robust face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1561–1576, Aug. 2011.
- [35] R. He, W.-S. Zheng, T. Tan, and Z. Sun, "Half-quadratic-based iterative minimization for robust sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 2, pp. 261–275, Feb. 2014.
- [36] Y. Feng, X. Huang, L. Shi, Y. Yang, and J. A. K. Suykens, "Learning with the maximum correntropy criterion induced losses for regression," *J. Mach. Learn. Res.*, vol. 16, pp. 993–1034, 2015.
- [37] U. Mitrović, F. Pernuš, B. Likar, and Ž. Špiclin, "Simultaneous 3D–2D image registration and C-arm calibration: Application to endovascular image-guided interventions," *Med. Phys.*, vol. 42, no. 11, pp. 6433–6447, 2015.
- [38] Y. Otake *et al.*, "3D–2D registration in mobile radiographs: Algorithm development and preliminary clinical evaluation," *Phys. Med. Biol.*, vol. 60, no. 5, p. 2075, 2015.
- [39] C. Syben *et al.*, "Joint calibration and motion estimation in weight-bearing cone-beam CT of the knee joint using fiducial markers," in *Proc. IEEE Int. Symp. Biomed. Imag.*, Apr. 2017, pp. 494–497.
- [40] D. Knaan and L. Joskowicz, "Effective intensity-based 2D/3D rigid registration between fluoroscopic X-ray and CT," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2003, pp. 351–358.
- [41] L. Lemieux, R. Jagoe, D. R. Fish, N. D. Kitchen, and D. G. T. Thomas, "A patient-to-computed-tomography image registration method based on digitally reconstructed radiographs," *Med. Phys.*, vol. 21, no. 11, pp. 1749–1760, 1994.
- [42] E. B. van de Kraats, G. P. Penney, D. Tomaževič, T. Van Walsum, and W. J. Niessen, "Standardized evaluation methodology for 2-D-3-D registration," *IEEE Trans. Med. Imag.*, vol. 24, no. 9, pp. 1177–1189, Sep. 2005.
- [43] A. Brost, R. Liao, N. Strobel, and J. Hornegger, "Respiratory motion compensation by model-based catheter tracking during EP procedures," *Med. Image Anal.*, vol. 14, no. 5, pp. 695–706, 2010.
- [44] U. Mitrović, Ž. Špiclin, B. Likar, and F. Pernuš, "3D–2D registration of cerebral angiograms: A method and evaluation on clinical images," *IEEE Trans. Med. Imag.*, vol. 32, no. 8, pp. 1550–1563, Aug. 2013.
- [45] D. Tomaževič, B. Likar, and F. Pernuš, "'Gold standard' data for evaluation and comparison of 3D/2D registration methods," *Comput. Aided Surgery*, vol. 9, no. 4, pp. 137–144, 2004.
- [46] S. A. Pawiro *et al.*, "Validation for 2D/3D registration I: A new gold standard data set," *Med. Phys.*, vol. 38, no. 3, pp. 1481–1490, 2011.
- [47] H. Madan, F. Pernuš, B. Likar, and Ž. Špiclin, "A framework for automatic creation of gold-standard rigid 3D–2D registration datasets," *Int. J. Comput. Assist. Radiol. Surgery*, vol. 12, no. 2, pp. 263–275, 2017.
- [48] P. Markelj, B. Likar, and F. Pernuš, "Standardized evaluation methodology for 3D/2D registration based on the Visible Human data set," *Med. Phys.*, vol. 37, no. 9, pp. 4643–4647, 2010.
- [49] M. Utzschneider, J. Wang, R. Schaffert, A. Borsdorf, and A. Maier, "Real-time-capable GPU-framework for depth-aware rigid 2D/3D registration," in *Bildverarbeitung für die Medizin*, G. Maier-Hein, K. Fritzsche, G. Deserno, T. Lehmann, H. Handels, and T. Tolxdorff, Eds. Berlin, Germany: Springer-Verlag, 2017, pp. 185–190.