



Title	Robustness and Accuracy of Feature-Based Single Image 2-D-3-D Registration Without Correspondences for Image-Guided Intervention
Author(s)	Kang, X; Yoshito, O; Yau, WP; Cheung, PYS; Hu, Y; Taylor, R
Citation	IEEE Transactions on Biomedical Engineering, 2014, v. 60 n. 1, p. 149-161
Issued Date	2014
URL	http://hdl.handle.net/10722/189097
Rights	IEEE Transactions on Biomedical Engineering. Copyright © Institute of Electrical and Electronics Engineers.

Robustness and Accuracy of Feature-Based Single Image 2-D–3-D Registration Without Correspondences for Image-Guided Intervention

Xin Kang*, *Member, IEEE*, Mehran Armand, Yoshito Otake, Wai-Pan Yau, Paul Y. S. Cheung, *Senior Member, IEEE*, Yong Hu, *Senior Member, IEEE*, and Russell H. Taylor, *Fellow, IEEE*

Abstract—2-D-to-3-D registration is critical and fundamental in image-guided interventions. It could be achieved from single image using paired point correspondences between the object and the image. The common assumption that such correspondences can readily be established does not necessarily hold for image guided interventions. Intraoperative image clutter and an imperfect feature extraction method may introduce false detection and, due to the physics of X-ray imaging, the 2-D image point features may be indistinguishable from each other and/or obscured by anatomy causing false detection of the point features. These create difficulties in establishing correspondences between image features and 3-D data points. In this paper, we propose an accurate, robust, and fast method to accomplish 2-D–3-D registration using a single image without the need for establishing paired correspondences in the presence of false detection. We formulate 2-D–3-D registration as a maximum likelihood estimation problem, which is then solved by coupling expectation maximization with particle swarm optimization. The proposed method was evaluated in a phantom and a cadaver study. In the phantom study, it achieved subdegree rotation errors and submillimeter in-plane (X – Y plane) translation errors. In both studies, it outperformed the state-of-the-art methods that do not use paired correspondences and achieved the same accuracy as a state-of-the-art global optimal method that uses correct paired correspondences.

Manuscript received December 17, 2012; revised July 7, 2013; accepted August 11, 2013. Date of publication August 15, 2013; date of current version December 16, 2013. This work was supported by NIH/NIBIB Grant R21 EB007747, Research Grant and RPg Exchange Funding of the University of Hong Kong, research fellowship from JSPS Postdoctoral Fellowships for Research Abroad, and Johns Hopkins University internal funds. *Asterisk indicates corresponding author.*

*X. Kang was with the Department of Orthopaedics & Traumatology, University of Hong Kong, Hong Kong, and the Engineering Research Center for Computer-Integrated Surgical Systems and Technology (CISST ERC), Johns Hopkins University, Baltimore, MD 21218 USA (e-mail: h0974152@connect.hku.hk).

M. Armand is with the Department of Mechanical Engineering and Orthopaedic Surgery, Johns Hopkins University, Baltimore, MD 21218 USA and also with Applied Physics Laboratory, Laurel, MD 20723 USA (e-mail: Mehran.Armand@jhuapl.edu).

Y. Otake is with the Department of Computer Science, Johns Hopkins University, Baltimore, MD 21218 USA (e-mail: otake@jhu.edu).

W.-P. Yau and Y. Hu are with the Department of Orthopaedics and Traumatology, University of Hong Kong, Hong Kong (e-mail: peterwpy@hkucc.hku.hk; yhud@hkucc.hku.hk).

P. Y. S. Cheung is with the Department of Electrical and Electronic Engineering, University of Hong Kong, Hong Kong (e-mail: paul.cheung@hku.hk).

R. H. Taylor is with the Department of Computer Science, with joint appointments in Mechanical Engineering, Radiology, and Surgery, and Engineering Research Center for Computer-Integrated Surgical Systems and Technology (CISST ERC), Johns Hopkins University, Baltimore, MD 21218 USA (e-mail: rht@jhu.edu).

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/TBME.2013.2278619

Index Terms—2-D–3-D registration, feature-based registration, image-guided interventions (IGIs), particle swarm optimization (PSO).

I. INTRODUCTION

REGISTERING 2-D intraoperative images to 3-D pre- and intraoperative data is an essential component and a crucial step for all image-guided interventions (IGIs) [1], which is called 2-D–3-D registration.¹ 2-D–3-D registration methods are briefly categorized into three classes: feature-based, intensity-based, and gradient-based [2]. In intensity- and gradient-based methods, patient anatomy is generally used, and thus, a same-patient preoperative CT is required. To reduce radiation exposures to patient, we used custom-designed tracking fiducials for intraoperative guidance. A tracking fiducial is much smaller than the patient anatomy and, hence, it occupies a relatively small space and overlaps with the patient anatomy in intraoperative images. The tracking fiducial may hinder the use of intensity-based methods (e.g., mutual information) because the intensity distribution of the tracking fiducial is easy to be overwhelmed by that of the anatomy. Moreover, the tracking fiducial is generally designed using simple geometric entities such as points and lines. This poses challenges to gradient-based methods in terms of accurate estimation of image gradient in noisy image for small structures, especially for small point entities (fiducial beads). Therefore, feature-based methods are comparatively more suitable when using tracking fiducials. In feature-based methods, point features have been used in a wide range of methods [3]–[10]. Practically, point features come from three sources: anatomical points, fiducial markers, and image contours (represented as a set of points).

2-D–3-D registration is a well-studied problem in cases where paired correspondences between a subset of a model and feature points are known, and many solutions have been proposed [11]–[14]. However, in clinical practice point correspondences cannot be readily established because feature points may be obscured and individually undistinguishable, and because cluttered image background and imperfect feature extraction may cause false detection in the image.

¹Note that 2-D–3-D is different from 2-D/3-D. The former means 2-D-to-3-D registration that registers two sets of data in different dimensions, while the latter means 2-D-to-2-D or 3-D-to-3-D registration that registers two sets of data in the same dimension.

Without known paired point correspondences, most existing feature-based registration methods follow a common two-step scheme, where paired correspondences are established in the first step and the registration parameter is computed afterwards using these pairs correspondences in the second step. The two steps are iterated until some stop criterion is reached.

This paper presents a novel scheme that bypasses the paired correspondence establishment step and goes directly to the registration parameter estimation, by formulating the 2-D–3-D registration as a maximum likelihood estimation (MLE) of the transformation parameter. The estimation is solved efficiently using an optimization method that couples the particle swarm optimization (PSO) and the expectation maximization (EM) algorithm. Instead of using PSO in the M-step of EM and resulting in a local optimizer, we embed EM into PSO and make our method potentially a global optimizer. We demonstrate in the experiments that the proposed method, without establishing paired correspondences, has the same accuracy as a state-of-the-art *global optimal method that uses correct paired correspondences*, from only one image.

This paper is a significant extension of our previous work [15], including extended literature review, detailed rationale of method design, a novel optimization method that embeds EM into PSO, additional evaluations in an animal study, and thorough comparisons with state-of-the-art methods.

This paper is organized as follows. Section II reviews related work in the literature. Section III describes the proposed algorithm in detail. In Section IV, the accuracy and robustness of the method is evaluated and the comparison is performed using 100 X-ray images. Conclusions are drawn in Section VII with some brief discussions of our future work.

II. RELATED WORK AND CONTRIBUTIONS

A. Problem Formulation

Given 3-D model points and 2-D image feature points with paired (one-to-one) correspondences known between a subset of these model and feature points, the 2-D–3-D registration is commonly formulated as a least squares using these paired correspondences to solve

$$\min_{\theta \in SE(3)} \sum_{n=1}^N \|\mathbf{u}_n - \mathcal{T}(\mathbf{X}_{\eta(n)}; \theta)\|^2 \quad (1)$$

where $\mathbf{u}_n, n = 1, \dots, N$, represent 2-D image feature points, $\mathbf{X}_{\eta(n)}$ represent 3-D model points that correspond to \mathbf{u}_n , $\mathcal{T}: \mathbb{R}^3 \mapsto \mathbb{R}^2$ denotes a 3-D-to-2-D projection following a rigid transformation with parameters $\theta = \{\mathbf{R}, \mathbf{t}\}$ including a rotation $\mathbf{R} \in SO(3)$ and a translation $\mathbf{t} \in \mathcal{R}^3$, and $\eta: \{1, \dots, N\} \mapsto \{1, \dots, M\}$ is a function that assigns an image feature point to its model counterpart.

It is also known that (1) can be equivalently written as

$$\min_{\theta \in SE(3)} \sum_{n=1}^N \sum_{m=1}^M \eta_{mn} \|\mathbf{u}_n - \mathcal{T}(\mathbf{X}_m; \theta)\|^2 \quad (2)$$

where $\mathbf{X}_m, m = 1 \dots M$, represent the 3-D model points. The η_{mn} equals one if \mathbf{X}_m corresponds to \mathbf{u}_n and zero otherwise.

A 2-D–3-D registration method also needs to deal with outliers. Here “outliers” are defined as the false positive or irrelevant points detected in the image. This is different from the commonly used outliers in pose estimation in computer vision, which are defined as incorrect paired correspondences.

B. Related Work

The most critical part in the aforementioned formulation is how to establish paired correspondences in \mathbf{u}_n and \mathbf{X}_m , i.e., how to obtain η_{mn} . According to the essential idea on establishing paired correspondences, the related methods in the literature can be categorized into ICP-like, hypothesis-test-based, softassign-based, and correspondence-less methods.

1) *ICP-Like Methods*: A majority of methods [3]–[10] adapt the well-known iterative closest point (ICP) algorithm [16], by assigning paired correspondences to η_{mn} based on the closest distance criterion. The basic ICP performs poorly if the data include spurious data points (points without correspondences, also called outliers). To deal with outliers, robust methods such as M-estimators [3] and extended Kalman filters [5], are employed. Moreover, ICP can easily be trapped in local minima, making it sensitive to initializations. Therefore, these methods generally need an initialization in the neighborhood of the true transformation.

Zheng *et al.* [6]–[10] adopted 2-D–3-D registration to address the construction of patient-specific surface models from a few X-ray images. Before estimating the global deformation using point distribution models (PDMs) [17] and the local deformation using TPSs, an affine transformation between the mean shape of the PDM and the X-ray was determined using an adapted ICP. To determine paired correspondences, the cross-matching elimination [9] was used.

2) *Hypothesis-Testing-Based Methods*: Pose estimation in computer vision (also called Perspective-n-Point, PnP) is a problem very close to the 2-D–3-D registration in this paper. To solve PnP, many methods utilize the random sample consensus (RANSAC) algorithm [18]. Their significant difference from the 2-D–3-D registration is that a set of candidate paired correspondences is first generated using feature descriptors. Then, a “hypothesis-testing” procedure is performed to find the most reliable subset of paired correspondences. Finally, the optimal pose is estimated using these paired correspondences.

Enqvist *et al.* [19], [20] proposed a method to deal with incorrect matchings, and hence, obtaining optimal correspondences. Applied to pose estimation [19], the method is used to find the most consistent subset from a set of candidate paired correspondences [20], while the candidate paired correspondences were generated using scale-invariant feature transform (SIFT) [21].

In the 2-D–3-D registration in this paper, the local feature descriptors used in computer vision for generating candidate paired correspondences are not applicable. In addition, generating hypotheses using all the combinations of model points and image feature points is very time consuming, while a fast algorithm is need in clinical practice.

3) *Softassign-Based Method*: Chui and Rangarajan [22] proposed to use softassign [23], [24] and deterministic annealing

for nonrigid matching of two point sets, both in 2-D or in 3-D. The nonrigid spatial mapping was formulated using the TPS. For each point set, the outlier was modeled using a Gaussian distribution with a very large variance and the mean placed at the center of mass. This is not optimal to reliably set the robustness parameter, and setting the weight of the smoothness term can be difficult [22]. Moreover, it implemented a simpler form than (2) since including outliers in its formulation is very cumbersome [22]. Last but not least, this method cannot be employed for 2-D-3-D registration since the perspective projection is a nonlinear mapping from 3-D to 2-D, while TPS models a nonrigid warp in 2-D or in 3-D.

For 2-D-3-D registration (pose estimation), David *et al.* [25] combined the softassign and deterministic annealing to determine paired correspondences and the POSIT [13] to estimate a 3-D transformation using these paired correspondences. This method, called SoftPOSIT, is arguably the most computationally effective method due to its accuracy and efficiency [26]. It iteratively estimates paired correspondences and transformation by minimizing the objective function

$$\min_{\theta \in SE(3)} \sum_{n=1}^N \sum_{m=1}^M \eta_{mn} (\|\mathbf{u}_n - \mathcal{T}(\mathbf{X}_m; \theta)\|^2 - \alpha) \quad (3)$$

where α is a threshold that defines two points as unmatchable if their distance is greater than α , and $\eta_{mn} \in \{0, 1\}$ are binary weights [13]. This method is sensitive to initializations, and is not robust in the presence of a large number of outliers.

4) *Correspondenceless Methods*: Several correspondenceless methods have been proposed. These methods either used simplified camera models [27] or have special requirements [28], [29]. In [28], it is assumed that the object consists of planar surfaces with closed curves drawn on them. In [29], a closed-form solution was proposed. However, the method has a very strong constraint: the camera can only have in-plane translations and rotations (i.e., moving in the X - Y plane and rotating around the optical axis parallel to the Z -axis). In [5], bitangent lines and bitangent planes were employed to find an initial pose under projective projection. However, this method is not suitable for the registration of point sets that do not form curves or surfaces.

5) *Related Methods in Point Set Registration*: Our method adopts the Gaussian mixture model (GMM) [30] and EM [31]. These methods have been employed for 2-D-2-D and 3-D-3-D point set registration [32]–[34], but to our knowledge not been used for 2-D-3-D registration.

For 2-D-2-D/3-D-3-D point set registration, Myronenko *et al.* [32] proposed the *coherent point drift* (CPD) method. They considered the problem as a probability density estimation, where one point set represents the GMM centroids and the other represents the data points. An EM algorithm was derived to solve the problem. However, CPD cannot be extended to 2-D-3-D registration since, unlike in point set registration where the rotational and translational parameters can be decomposed, the perspective projection is nonlinear and the two parameters cannot be decomposed.

For 3-D-3-D surface registration, Granger *et al.* [33] proposed the EM-ICP method. The scene points were modeled as noisy measurements of the model points. It models the localization error using a Gaussian. The variance of the Gaussian (called “scale”) is monotonically decreased by dividing it with a predefined value. However, doing so can sometimes lead to an increase of the criterion value in consecutive iterations [33]. This method anneals the variance rather than considering it as a parameter to be estimated. The initial value of the scale depends heavily on the initialization of registration. To handle outliers, a distance threshold was used rather than modeling outliers explicitly.

For 3-D-3-D rigid and articulated point set registration, Horaud *et al.* [34] have proposed the *expectation conditional maximization for point registration* (ECMPR) method using GMM and expectation conditional maximization [35]. Different from the CPD method [32], ECMPR uses general covariance matrices for the GMM components and it estimates the rotational and translational parameters using a method based on semidefinite positive relaxation. However, the ECMPR method cannot be extended to 2-D-3-D registration due to the nonlinear perspective projection involved. The nonlinearity of perspective projection complicates the estimation of the rotational parameter.

In summary, the existing literature shows that GMM and EM methods are well suited for 2-D-2D/3-D-3-D point set registration. However, these methods cannot be extended directly to 2-D-3-D registration. We propose to combine GMM and EM methods with the global optimizer PSO for solving the 2-D-3-D registration robustly and effectively.

C. Contributions

To the best of our knowledge, this study is the first time that these methodologies are used for 2-D-3-D registration without using paired correspondences. This paper has the following original contributions.

1) Given a set of 3-D and 2-D *unmatched points*, the proposed method bypasses the paired correspondence establishing step and goes directly to the registration parameter estimation; it neither requires nor establishes paired correspondences. The method estimates the registration parameter solely based on geometry without considering the appearance of the feature points, which can become an unreliable measure when the scene contains many similar and easily confused fiducials. In addition, the method is able to achieve accurate 2-D-3-D registration from one image, although it may also be adapted readily to cases where more than one image is available.

2) The method couples PSO with EM to solve the optimal estimation of the registration parameter. The intuitive manner of combining PSO and EM is to use POS as a numerical solver in the M-step since there is no closed-form solution. However, we integrate the E-step into the PSO.

First, PSO is a global optimizer, while EM is a local one. If we use PSO in the M-step, EM will dominate the optimization procedure. This leads to a local optimizer, making the method sensitive to initializations. On the contrary, PSO ensures that

the optimization is over the original parameter space and not over some subspace. Most importantly, PSO has advantages on solving the problems that are partially irregular and noisy, and change over time, as the 2-D–3-D registration formulated in this paper, for which other commonly used methods are not suitable. Benefiting from making PSO govern the optimization procedure, the proposed method is robust to initializations and outliers, and is as accurate as a state-of-the-art global optimal method that *uses correct paired correspondences*, as illustrated in the experiments.

Second, the direct use of PSO is hampered by the unknown paired correspondences and the variance of the GMM components, called *nuisance parameters*. A nuisance parameter is a parameter which is not of immediate interest but which must be accounted for in the analysis of the parameters of interest. When only PSO is used, these nuisance parameters have to be put into the parameters to be optimized. So doing increases the dimension of the problem, make the optimization more difficult and more computational intensive. Most importantly, these nuisance parameters are not free parameters; they are related to the transformation parameters. In the proposed method, the nuisance parameters are estimated by their closed-form MLEs derived using the EM, ensuring that the estimates are optimal in statistical meaning. The closed-form solutions also accelerate the optimization. Therefore, EM deals with the nuisance parameters, allowing the transformation parameters be optimized via the global optimizer PSO.

3) Extensive experiments were performed with the proposed 2-D–3-D registration method using real images. We thoroughly study the behavior of the method with respect to the initial parameter values and the presence of outliers. The proposed method is also compared with several state-of-the-art methods.

III. REGISTRATION WITHOUT CORRESPONDENCES

A. 2-D–3-D Registration as MLE

We trace back the 2-D–3-D registration problem to the statistical inference solution of the transformation parameters, without establishing paired correspondences. We briefly derive the 2-D–3-D problem in the EM framework to make exposition of our method clearer and to introduce consistent notation. Similar derivation can be found in [32] and [34].

Without knowing the “true” paired correspondences, we assume that the n th 2-D point, \mathbf{u}_n , has a certain probability, p_{mn} , to correspond to the m th 3-D point, \mathbf{X}_m . We consider the 2-D–3-D registration as an MLE of the transformation parameters θ , where the projected 3-D points are represented by a mixture model, whereas the 2-D points are observations drawn from this model.

The EM algorithm is a generic mathematical framework for solving incomplete-data problems by formulating a complete-data problem through augmenting the missing data. To form complete data, an unknown paired correspondence between a 2-D point, \mathbf{u}_n , and a 3-D point, \mathbf{X}_m , is modeled using a normal distribution $p(\mathbf{u}_n|m) = \mathcal{N}_m(\mathcal{T}(\mathbf{X}_m; \theta), \sigma^2)$. This models the *a priori* probability of the m th 3-D point being the correspondence of the n th 2-D point. To account for outliers a dummy

point, \mathbf{X}_{M+1} , is introduced, which allows multiple points to be matched to its projection. Further, the probability of an image point being an outlier is modeled using a uniform distribution, $p(\mathbf{u}_n|M+1) = \mathcal{U}(0, N)$, since without *a priori* knowledge of the outlier it could be anywhere in the image. Consequently, the observed-data log-likelihood is

$$\ell_{\text{obs}}(\theta, \sigma^2|\mathbf{u}_n) = - \sum_{n=1}^N \log \left(\sum_{m=1}^{M+1} p(\mathbf{u}_n|m)P(m) \right) \quad (4)$$

where $P(m)$ is the prior probability that the projection of \mathbf{X}_m is one of the 2-D feature points detected in image. Generally, *a priori* knowledge on $P(m)$ is not available. Thus, without loss generality, we define $p(M+1) = w$ and $P(m) = \mathcal{U}(0, 1-w)$, $m = 1, \dots, M$, where $0 < w < 1$ is constant.

Directly solving the MLE of θ and σ^2 from (4) is extremely difficult since 1) the correspondence probabilities $p(\mathbf{u}_n|m)$ are constrained by the registration parameters; 2) the nonlinearity of perspective projection complicates the estimation of rotation matrix; 3) the estimate of θ further depends on σ^2 ; and 4) the summation inside the logarithms makes the computation of (4) intractable. By augmenting the unknown paired correspondences the complete-data log-likelihood is simplified as

$$\ell_{\text{com}}(\theta, \sigma^2|\mathbf{u}_n, m) = \sum_{n=1}^N \log (p(\mathbf{u}_n|m)P(m)). \quad (5)$$

Accordingly, the target is to minimize the objective function

$$Q = \frac{1}{2\sigma^2} \sum_{n=1}^N \sum_{m=1}^M p_{mn} \|\mathbf{u}_n - \mathcal{T}(\mathbf{X}_m; \theta)\|^2 + C \log \sigma^2 \quad (6)$$

where $C = \sum_{m=1}^M \sum_{n=1}^N p_{mn}$ and $p_{mn} = P(m|\mathbf{u}_n)$ denotes the posterior of a correspondence, which is calculated via Bayes' formula ($p_{mn} = p(\mathbf{u}_n|m)P(m)/p(\mathbf{u}_n)$) as

$$p_{mn} = \frac{\exp\left(-\frac{\|\mathbf{u}_n - \mathcal{T}(\mathbf{X}_m; \theta)\|^2}{2\sigma^2}\right)}{\sum_{m=1}^M \exp\left(-\frac{\|\mathbf{u}_n - \mathcal{T}(\mathbf{X}_m; \theta)\|^2}{2\sigma^2}\right) + \frac{2\pi\sigma^2 w M}{(1-w)N}}. \quad (7)$$

Furthermore, one can easily obtain the MLE of σ^2 as

$$\hat{\sigma}^2 = \frac{1}{2C} \sum_{m=1}^M \sum_{n=1}^N p_{mn} \|\mathbf{u}_n - \mathcal{T}(\mathbf{X}_m; \theta)\|^2. \quad (8)$$

B. Optimal Transformation Estimation Using PSO

Obviously minimizing (6) over θ does not lead to a closed-form solution because \mathcal{T} is a nonlinear perspective projection. To deal with this, one can estimate θ using some numerical optimization approach. However, the EM algorithm is a local optimization method. To a local method, good initial parameter estimates are essential. In our method, PSO [36] is adopted to estimate the global optimum of the objective function (6) because it combines a broader region search and a locally oriented search to obtain closer to an global optimum [37].

In PSO, the optimal solution is achieved by having a swarm of particles (candidate solutions) moving around in the search space \mathcal{S} . The movements of the particles are guided by their

own best known position $\theta_{i,\text{best}}$ and the entire swarm's best known position θ_{best} . At beginning, each particle's position and velocity are initialized as uniformly distributed random vectors $\theta_i \sim \mathcal{U}(\theta_0 - \mathbf{r}/2, \theta_0 + \mathbf{r}/2)$ and $\mathbf{v}_i \sim \mathcal{U}(-\mathbf{r}, \mathbf{r})$, respectively, where θ_0 and \mathbf{r} are the *center* and the *range* of \mathcal{S} . And each particle's best known position is set to its initial position (i.e., $\theta_{i,\text{best}} = \theta_i$). Then, each particle updates its position and velocity according to the following equations:

$$\mathbf{v}_i = \omega \mathbf{v}_i + c_p r_p (\theta_{i,\text{best}} - \theta_i) + c_g r_g (\theta_{\text{best}} - \theta_i) \quad (9)$$

$$\theta_i = \theta_i + \mathbf{v}_i \quad (10)$$

where $r_p, r_g \sim \mathcal{U}(0, 1)$, ω is the inertia weight, and c_p and c_g are the ‘‘cognitive’’ and ‘‘social’’ parameters. This process is iterated until a minimum error criterion or a predefined maximum iteration is attained.

The coupling of the PSO and the EM allows us to solve 2-D-3-D registration problems robustly, accurately, and efficiently, as we will demonstrate in our experiments.

C. Implementation

The proposed method is as follows.

- 1) Set $\sigma^{(0)} = \sqrt{(\text{image size})/2}$, the search space center θ_0 , the search space range \mathbf{r} and the outlier prior w
- 2) Initialize K random particles by generating $\theta_i \sim \mathcal{U}(\theta_0 - \mathbf{r}/2, \theta_0 + \mathbf{r}/2)$ and $\mathbf{v}_i \sim \mathcal{U}(-\mathbf{r}, \mathbf{r})$, $i = 1, \dots, K$.
- 3) Compute the posterior modes $p_{mn,i}^{(t)}$ using (7) with θ_i .
- 4) Obtain $\theta_{i,\text{best}}$ and θ_{best} by evaluating the objective function (6), and set $\theta^{(t)} = \theta_{\text{best}}$ and $p_{mn}^{(t)} = p_{mn,\text{best}}^{(t)}$.
- 5) Compute $(\sigma^2)^{(t+1)}$ by (8) using $\theta^{(t)}$ and $p_{mn}^{(t)}$.
- 6) Update each particle's position and velocity using (9) and (10)
- 7) Iterate Step 3 to 6 until meeting the stop criteria
- 8) (Optional) Compute the maximum *a posteriori* (MAP) paired correspondence of \mathbf{u}_n as

$$\hat{\mathbf{X}}_n \equiv \mathbf{X}_{\hat{m}}, \quad \hat{m} = \arg \max_{m \in \{1, \dots, M+1\}} p_{mn}. \quad (11)$$

and the set of valid MAP paired correspondences

$$L = \{\hat{\mathbf{X}}_n \equiv \mathbf{X}_{\hat{m}} | \hat{m} \neq (M+1)\}. \quad (12)$$

If $|L| > c_L$ (a predefined constant), reinitialize $K = 2K$ particles and go to Step 10.

- 9) (Optional) Compute the root mean square (RMS) error

$$e = \sqrt{\frac{1}{|L|} \sum_{\hat{\mathbf{X}}_n \in L} \|\mathbf{u}_n - \mathcal{T}(\hat{\mathbf{X}}_n; \theta_{\text{best}})\|^2} \quad (13)$$

If $e > \epsilon$ (a predefined threshold), reinitialize $K = 2K$.

- 10) (Optional) Check whether the maximum re-initialization is attained, otherwise go to Step 2.

There are three common ways to set a stop criterion in Step 7: 1) defining a maximum number of iterations, 2) checking the change of the objective function values and/or the transformation parameters between two consecutive iterations, and 3) checking the number of inactive particles [38]. In both phantom and cadaver studies, we used the threshold on the change

of objective function values of 10^{-6} , the maximum iteration of 250, the initial number of particles $K = 200$, the maximum reinitialization of 3 and the outlier priori $w = 0.01$.

The Steps 8–10 intend to accelerate the method. Using a larger amount of particles could benefit PSO as more function values can be evaluated in one iteration. However, this leads to a higher computational load. More importantly, a sufficient number is application dependent. We design to start from a small number of particles and increase the number if the result is not desirable, avoiding using a large number of particles and handling the possibility of obtaining poor result using a small number of particles.

The Steps 8 and 9 are used to measure the plausibility of a result. A minimal set of three noncollinear point pairs can give a 2-D-3-D registration result (with ambiguities), but at least five noncollinear, noncoplanar pairs are preferred in practice for a stable and accurate result. In our application, we set $c_L = 5$. This is also a tradeoff for allowing occlusions and false negatives in feature detection. In our case, nine 3-D model points were used for registration, but their counterparts in the image were occluded by other strong features or other model points, or too weak to detect. Thus, some model points may not have their MAP paired correspondence. In other applications, this number can be defined as desired.

In our experiments, the threshold of the RMS error was empirically set to $\epsilon = \sqrt{2}$ pixels as the projected model points are expected to be within the eight neighborhood of the detected image points. If the maximum reinitialization is attained, the result with the minimal RMS error e is given as the final result.

Steps 8–10 are optional for the proposed method and the relevant thresholds are application dependent, but the scheme is effective in our application as illustrated in the experiments.

IV. EXPERIMENTS

The experiments aim to evaluate the performance of the proposed algorithm. For this purpose, we have chosen femoroplasty (injection of bone cement into the proximal femur) as a specific example application. Femoroplasty is an effective countermeasure to reduce the risk of fracture in an osteoporotic hip. An image-guided, robotic-assisted system [39] has been proposed to facilitate the femoroplasty surgery of hip fractures in at-risk patients. The workflow of the system is described in detail in [39]. This paper focuses on the critical step of estimating the pose of the C-arm via 2-D-3-D registration from a single X-ray image.

A hybrid fiducial called ‘‘FTRAC’’ [40] is used in the system for C-arm pose estimation. The FTRAC is a mathematically optimized fluoroscope tracking fiducial initially developed for prostate brachytherapy [40] and extended to a hybrid fiducial for femoroplasty [39]. It comprises nine stainless steel beads and four straight lines and two ellipses made of stainless steel wires. The size of the FTRAC used in this experiment is $18 \times 18 \times 72$ mm. It was designed for estimating the 6-DOF pose of the C-arm from its projection in the X-ray image.

Our method was evaluated in a phantom and a cadaver study. In both studies, only the nine beads were used for the 2-D-3-D

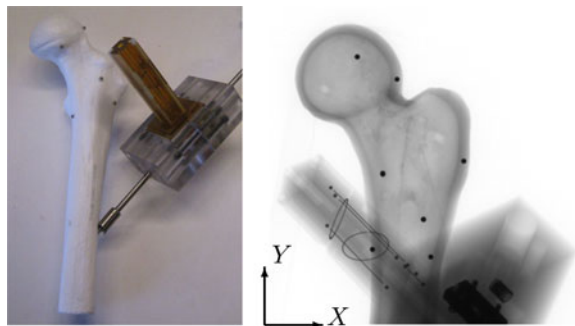


Fig. 1. (Left) Plastic femur phantom with the fiducial affixed on it and (right) a typical X-ray image of them taken by the bench system. The Z -axis of the image points out of the paper. The nine beads on the fiducial were used in the experiments, whereas the larger beads affixed on the femur phantom were not used and they were outliers in our experiments.

registration. In all experiments, the image points were automatically detected from the X-ray images, and exactly the same points detected were used in all the methods. Our detection method is based on a multiscale voting scheme utilizing both image intensity and gradient. A detailed description of the method will be the subject of another publication. But it is worth noting that the detection method is generic and does not favor any particular registration method.

In both studies, the proposed method were compared with SoftPOSIT, which is arguably the most computationally effective method due to its accuracy and efficiency [26], and a global optimal PnP solver [41] (referred to as “gOp”). In the cadaver study, we further compare the proposed method with a newly proposed multiview intensity-based method [39].

A. Phantom Study

The experiments were performed using X-ray images of the FTRAC attached on a plastic bone phantom (Sawbone, Pacific Research Laboratories, WA, USA) (see Fig. 1) acquired by a cone-beam CT (CBCT) imaging bench [42]. The flat-panel detector (FPD) was a PaxScan 4030CB (Varian Imaging Products, Palo Alto, CA, USA) that provide distortionless images. Geometric calibration was performed using the method proposed by Cho *et al.* [43] with submillimeter and subdegree accuracy. In total 100 images were acquired from different viewpoints. The nine radio-opaque metal beads with known configuration on the FTRAC were used to evaluate our method.

In the X-ray images, the fiducial occupies a relatively small portion and overlaps with the sawbone, due to its structure and size. In addition, there were some metal beads affixed on the sawbone. Furthermore, the metal beads on the fiducial (and the phantom) are not distinguishable from each other in the X-ray images. These conditions pose challenges in intensity-based registration methods. Our method achieved the 2-D-3-D registration using the nine small beads on the fiducial without knowing the correspondences between the beads on the model and the points detected in the X-ray images. The larger beads affixed on the phantom (see Fig. 1) were *not* used for registration.

1) *Ground-Truth Transformations:* The ground-truth transformations ($\mathbf{r}_{gt}, \mathbf{t}_{gt}$) was obtained from high-resolution CBCT

images. Each bead in the CBCT data was manually segmented and fit with a sphere so that a surface model of each bead was created. Then, a point-to-point rigid registration between the surface model and the CAD model was carried out, using the centers of the fitted spheres. The fiducial registration error of this procedure for all the beads is 0.14 ± 0.06 mm.

2) *Robustness to Initializations:* The proposed method was evaluated using 100 X-ray images taken from different viewpoints. For each image, 50 trials were conducted using 50 different initializations generated using the ground-truth rotation \mathbf{r}_{gt} and the C-arm source-to-detector distance d_{SD} .

Each initialization (\mathbf{r}, \mathbf{t}) represents a 6-DOF transformation using a 3-vector for translation and Euler angles for rotation. The initial translations were independent from \mathbf{t}_{gt} ; they were set to $\mathbf{t} = (0, 0, d_{SD}/2)$ since practically the imaged object is oftentimes around the middle of C-arm. In the phantom study, $d_{SD} = 1184$ mm. The rotations were initialized as uniformly distributed random vectors $\mathbf{r} \sim \mathcal{U}(\mathbf{r}_{gt} - 20^\circ, \mathbf{r}_{gt} + 20^\circ)$.

For each image, 50 random independent initializations were generated without duplicate. Using these initializations, the projected FTRAC beads were outside the image in some cases.

The search range of the PSO, \mathbf{r} , was set to 40° in rotations and 200 mm in translations. These ranges are much larger than the ranges used in literature.

3) *Robustness to a Large Amount of Outliers:* In the proposed method, the fiducial beads were automatically localized as feature points in the X-ray images and directly used for 2-D-3-D registration. In practice, automatic feature point detection is preferable, requiring that the method is robust to outliers. As mentioned before, there were also seven beads affixed on the femur phantom for calculating ground-truth transformations. These beads were automatically detected and they were outliers.

To evaluate the robustness of our method to a large amount of outliers, we added 80 uniformly generated random false detections in the image. Thus, together with the seven beads on the phantom, there were in total 87 outliers in each testing image. Still, only the nine beads on FTRAC were used for registration.

The proposed method was carried out on these 100 images without providing any paired point correspondences. The same initializations and search range in Section IV-A2 were used.

4) *Comparison With a Global Optimal PnP Method:* To further demonstrate the accuracy of our method, it is compared with a global optimal PnP solver [41] (referred to as “gOp”).

Note that, as a PnP solver the gOp requires *correct paired correspondences*. In this comparison, the ground-truth paired correspondences were used in gOp, whereas no paired correspondences were used in our method. The results of the gOp were compared with the results obtained in Section IV-A2.

B. Cadaver Study

We also evaluate our method in a cadaver study containing 27 images taken from different viewpoints. The experimental setup and an example image is shown in Fig. 2. The images were acquired using a prototype mobile C-arm [44] from a cadaver hip with the FTRAC affixed on the femur shaft. Images

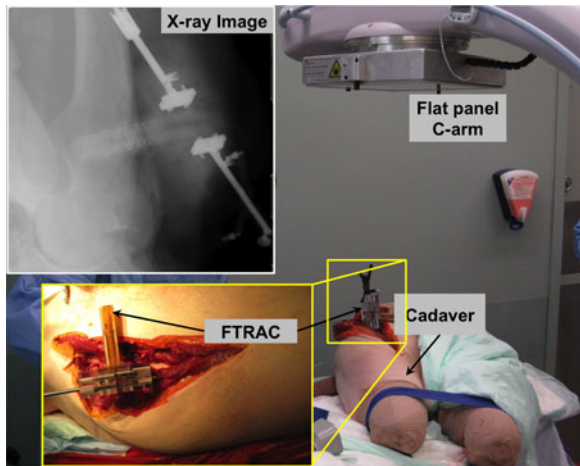


Fig. 2. Setup of the cadaver study in a well-calibrated environment, and an example X-ray image. The FTRAC fiducial was affixed firmly on the cadaver hip on the right femur shaft. The flat-panel C-arm was used to acquire X-ray images by rotating approximately around the long axis of the right femur.

without distortion were acquired by rotating the scanner about the long axis of the femur.

1) *Comparison With a Multiview Intensity-Based Method:* We compared our method with a multiview mutual information-based 2-D-3-D registration method [39] (referred to as MV-MI) for C-arm pose estimation.

In MV-MI, an initial pose is obtained by performing POSIT [13] using manually established paired correspondences. Starting from this initial pose, the C-arm pose is iteratively updated by maximizing the mutual information [45] between a DRR of the CAD model generated at the estimated pose and the original X-ray image. The Downhill Simplex algorithm [46] was used to maximize the objective function. Finally, a multiview refinement is carried out to improve the accuracy. For details, readers are referred to [39].

In the comparison, six images were used in MV-MI for pose estimation, whereas only one image was used in our method.

For initialization, MV-MI used four paired correspondences obtained by manually selecting four nonplanar fiducial beads in the image and their counterparts on FTRAC, whereas our method took the same initialization procedure and search range used in Section IV-A2. The source-to-detector distance is $d_{SD} = 1330$ mm. Since there is no ground truth, the initial rotations were generated using the results of MV-MI. However, to ensure a “poor” initialization, the estimated rotation of the image three apart from the current one in order was used. That is, when performing registration on the #1 image, the estimated rotation of the MV-MI on the #4 image was used.

2) *Comparison With SoftPOSIT and gOp:* Our method was also compared with SoftPOSIT [25] and gOp [41]. SoftPOSIT used the same initializations in Section IV-B1 and ran until convergence without setting a maximum iteration. For gOp, the ground-truth paired correspondences were used.

Note that the comparison between the MV-MI and the proposed method is a comparison between a multiview intensity-based method and a single-view feature-based method, while

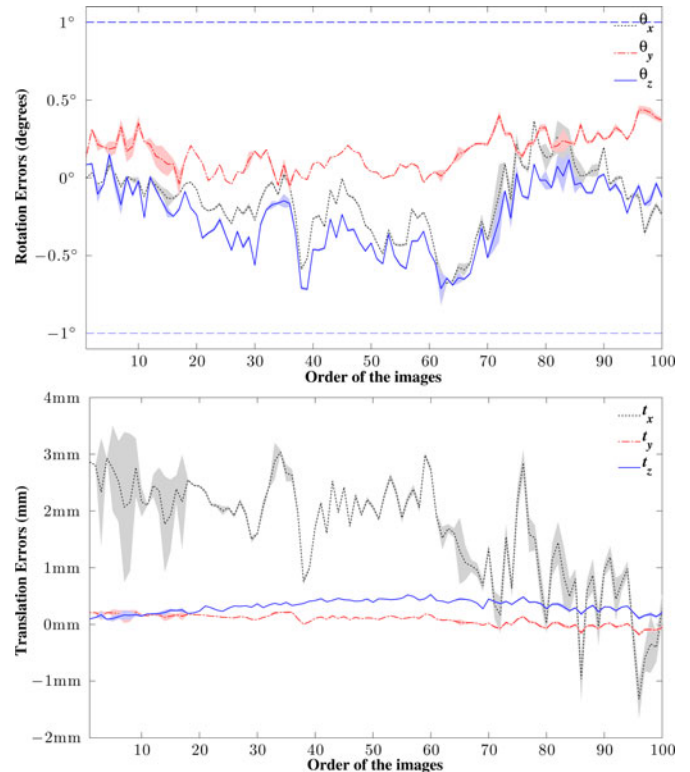


Fig. 3. Means and standard deviations of the registration errors in (top) rotations in degrees and in (bottom) translations in millimeter, respectively, on each of the 100 X-ray images over 50 trials. The solid, dash-dot and dotted lines are the mean values, and the shadow regions show the standard deviations.

the comparison between the proposed method and SoftPOSIT is a comparison between two feature-based methods.

V. EXPERIMENTAL RESULTS

A. Phantom Study

1) *Robustness to Initializations:* For each image, the mean and standard deviation of the errors between the estimated and ground-truth transformation parameters are shown in Fig. 3. The means are plotted using the solid, dash-dot and dotted lines, and the standard deviations are shown using the regions in light colors accordingly. The overall mean errors and standard deviations of each individual parameters over the 5000 experiments were $(-0.26^\circ \pm 0.22^\circ, 0.16^\circ \pm 0.12^\circ, -0.15^\circ \pm 0.22^\circ)$ in rotations around X-, Y-, and Z-axes, and $(0.32 \pm 0.11, 0.08 \pm 0.10, 1.65 \pm 1.00)$ mm in translations along X-, Y-, and Z-axes, respectively.

The aforementioned errors for each individual images over 50 trials are also shown using box-and-whisker plots in Fig. 4. Compared with the plots using means and standard deviations, a box-and-whisker plot give a less biased visualization of the data spread as it shows the registration errors of each individual trials. The area between the upper and lower boundaries of the box, called the interquartile range, shows the spread of the middle 50% of the registration errors. It is a more robust range for interpretation because the middle 50% is not affected by outliers or extreme values.

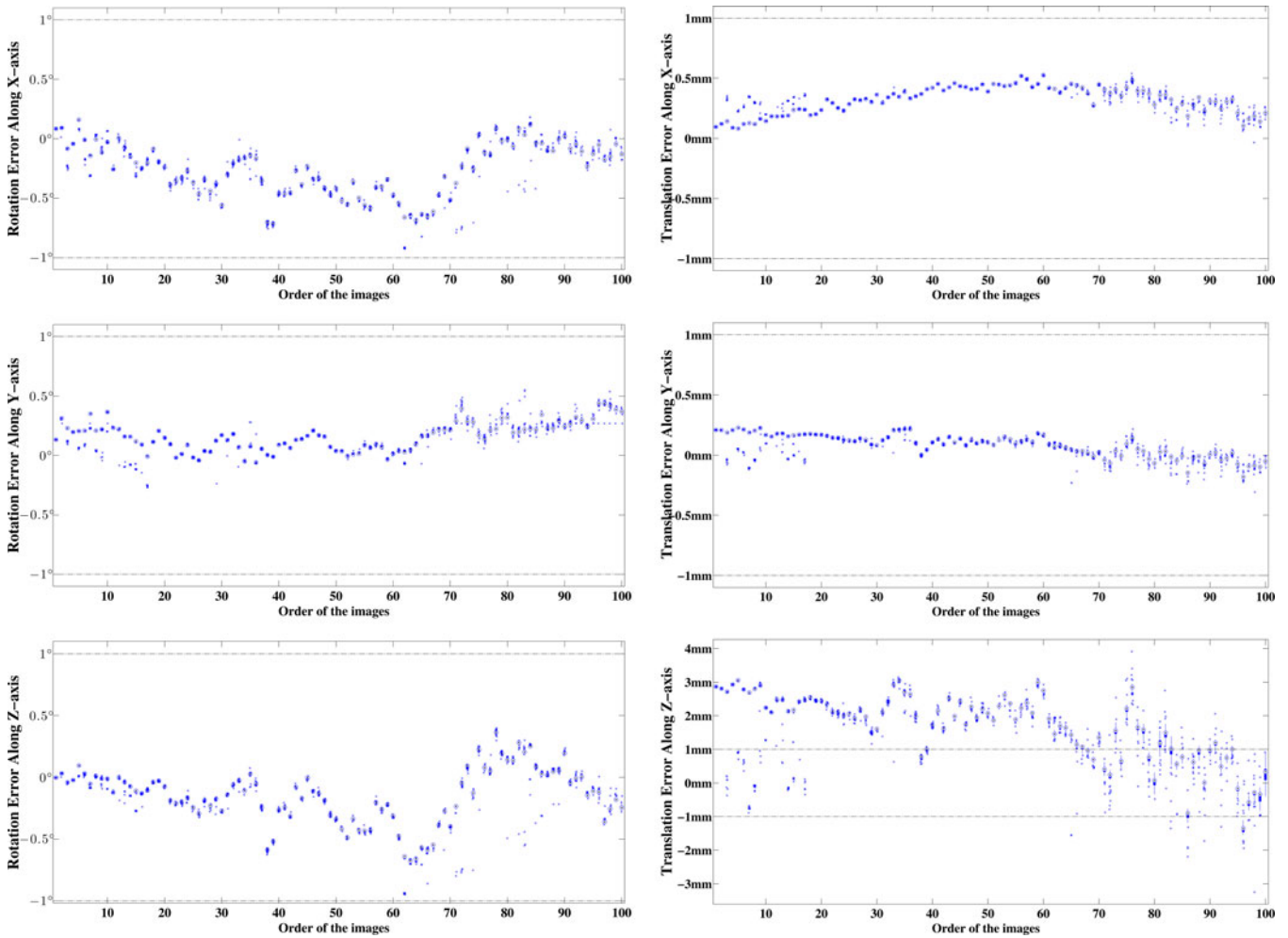


Fig. 4. Registration errors of our method in rotations and translations. In the left column, from top to bottom, the plots show the registration errors (in degrees) around the X -, Y -, and Z -axes over 50 trials on 100 images. In the right column, from top to bottom the plots show the registration errors (in millimeter) along the X -, Y -, and Z -axis. All rotation errors around X -, Y -, and Z -axis were less than 1° (indicated by the horizontal dashed lines), and all translation errors along X - and Y -axes were less than 1 mm, whereas the translation errors along Z -axis were relative larger.

The proposed method estimated rotations and in-plane translations (along the X - and Y -axes) accurately with subdegrees and submillimeter accuracy (see Fig. 4). All rotation errors were less than 1° , and all in-plane translation errors were less than 1 mm. The errors in the projection direction Z -axis (i.e., the depth) were relatively large since the depth can be difficult to accurately estimate from only single image.

2) *Dynamic Behavior of the Proposed Method*: To depict the dynamic procedure of our method, Fig. 5 shows one typical trial of the experiments. It also displays correspondence maps of iterations #0, #5, #10 and #50. A correspondence map is a visualization of the correspondence probabilities p_{mn} , with the horizontal axis the order of image points and the vertical axis the order of model points. In a correspondence map, a block in the bottom row represents the probability of an image point being an outlier, while a block in the n th column and the m th row above the bottom row represents the probability between the n th image point and the m th model point. A higher block value indicates a higher correspondence probability.

At initialization (iteration #0), there was no dominant correspondence as the bottom line shows the highest values for all points. In iteration #5, multiple correspondences with low probabilities were established for several points, indicated by light blue color. In iteration #10, correspondences with high probabilities were found. In iteration #50, the correspondences with high certainty were established. A side product of the proposed method is that, paired correspondences can be obtained using the maximum *a posteriori* (MAP) of the correspondence probabilities. Though our method does not intend to establish them, after the 2-D–3-D registration, an image point that has the largest probability to a model point could be considered as the correspondence of that model point.

3) *Robustness to Outliers*: The registration errors for each individual images are shown in Fig. 6. Although there were a large amount of outliers and the initializations were large, the proposed method produced subdegree rotation errors and submillimeter in-plane translation errors. Comparing the results with that in Fig. 3, it can be seen that the registration

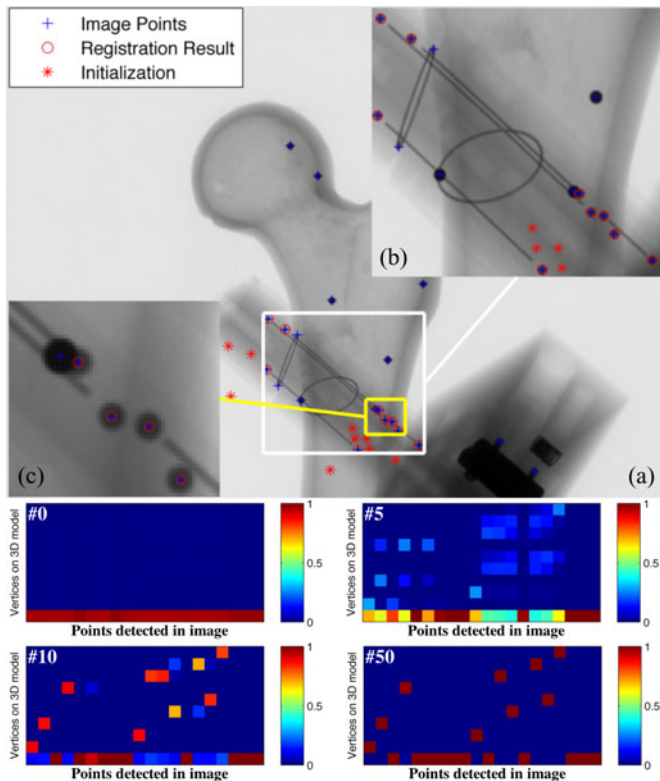


Fig. 5. Behavior of the proposed method. The detected beads, initialization, and final estimate are shown using blue plus signs (“+”), red asterisks (“*”), and red circles (“o”), respectively. Four correspondence maps below the image illustrate the behavior of our method at iterations #0 (initialization), #5, #10, and #50. The correspondence map is a visualization of the correspondence probabilities p_{mn} with the horizontal axis the order of image points and the vertical axis the order of model points. The higher the value of a block in the map, the higher the correspondence probability is. Two enlarged regions give close-up displays of the final estimate (b) and an outlier (a large bead on the femur phantom) that is very close to a FTRAC bead (c).

errors were at the same level as that obtained with much less outliers.

Fig. 5(c) also illustrates the robustness of our method to outliers. In the image, the 3-D structure of FTRAC is degraded and one outlier (a bead on the femur phantom) is very close to the projection of a FTRAC bead. In this case, our proposed method still achieved the registration with errors of $(-0.47^\circ, 0.03^\circ, -0.25^\circ, 0.33 \text{ mm}, -0.12 \text{ mm}, 1.84 \text{ mm})$. As can be seen in Fig. 5(b), the projected registered FTRAC beads well fit the detected beads.

4) *Comparison With gOp and SoftPOSIT*: Using the ground-truth paired correspondences, gOp was performed on the 100 images. The registration errors are shown in Table I and Fig. 7, along with the mean errors of our method for comparison. The average errors of the proposed method were the same as that of gOp for most images, and they were smaller than that of gOp for some images. Also seen from Table I is that the depth is hard to estimate accurately from single image.

The registration errors of SoftPOSIT were larger than 1° and 1 mm in almost all cases, and hence, were not plotted in Fig. 7. To show the accuracy of SoftPOSIT, we performed another experiment using “good” initializations that were very close to

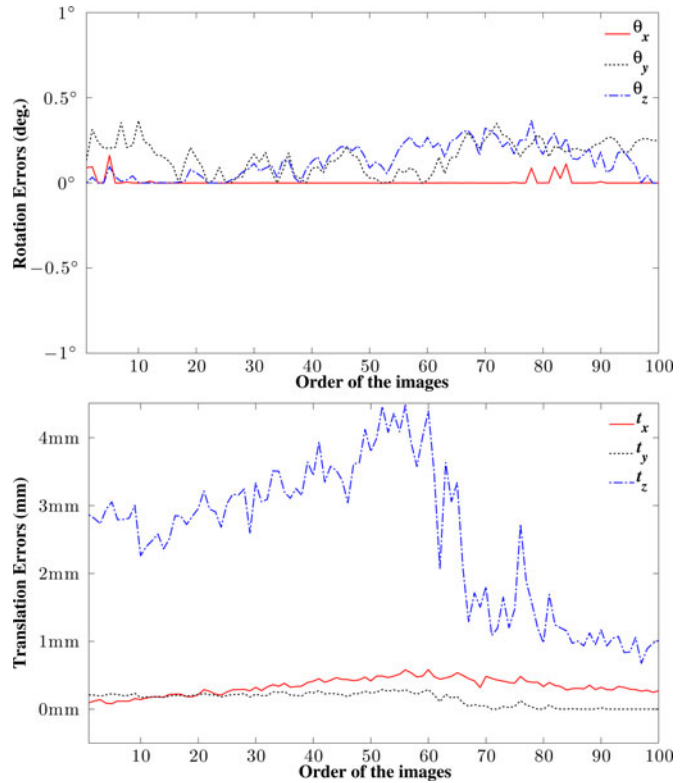


Fig. 6. Robustness of the proposed method to a large amount of outliers (87 outliers versus 9 correct feature points). The rotation errors for all images were within $\pm 0.5^\circ$. The translation errors along X- and Y-axes were less than 1 mm, and the errors along Z-axis were relative large. Notably, the registration errors were at the same levels as that when there were much less outliers, by comparing with Fig. 3.

TABLE I
REGISTRATION ERRORS OF OUR METHOD, gOp [41] AND SOFTPOSIT [25]
IN THE PHANTOM STUDY

Method	Rotation Errors (in degrees)		
	X	Y	Z
Our method	-0.26 ± 0.22	0.16 ± 0.12	-0.15 ± 0.22
gOp	-0.25 ± 0.22	0.17 ± 0.11	-0.13 ± 0.22
SoftPOSIT (best)	1.28 ± 9.40	5.98 ± 3.22	0.78 ± 5.28
Method	Translation Errors (in mm)		
	X	Y	Z
Our method	0.32 ± 0.11	0.08 ± 0.10	1.65 ± 1.00
gOp	0.32 ± 0.11	0.09 ± 0.09	1.76 ± 0.90
SoftPOSIT (best)	0.59 ± 0.62	-0.81 ± 1.51	-7.77 ± 17.05

The errors are the difference between the estimated and ground-truth transformation.

the ground-truth (within 1° and 1 mm). The best registration results in the 50 trials per image, named as SoftPOSIT (best), were used to calculate the statistics for comparison in Table I.

B. Cadaver Study

To measure the accuracy of the proposed method, the RMS projection error was used because there is no accurate ground truth for the cadaver data. The RMS projection errors of the proposed method, MV-MI, SoftPOSIT, and gOp are shown in

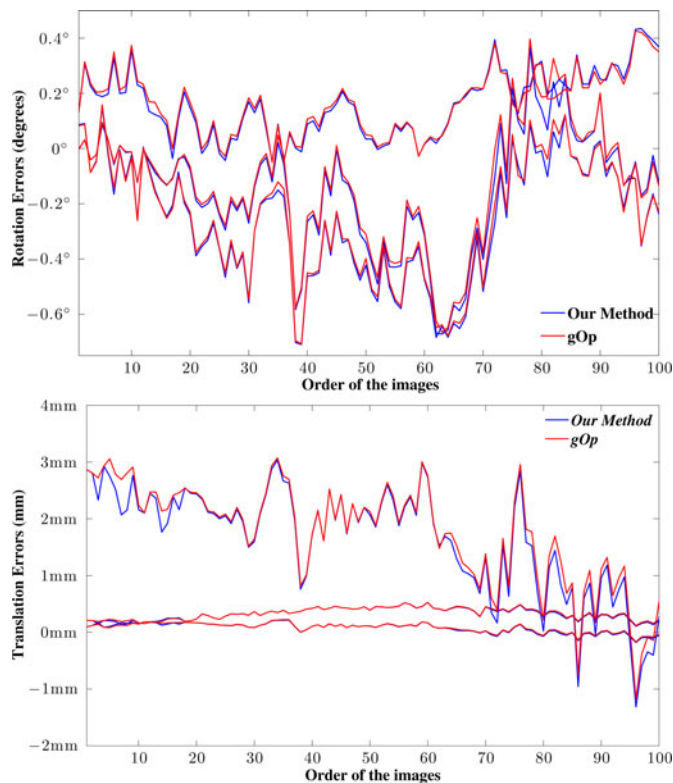


Fig. 7. Comparison of the registration errors between the proposed method with gOp [41]. For most images, the average errors of the proposed method were the same as that of gOp, and they were smaller than that of gOp for some images. Note that gOp requires *correct paired point correspondences*, while no paired correspondences were used in the proposed method.

TABLE II
RMS ERRORS (IN PIXELS) OF DIFFERENT METHODS IN THE CADAVER STUDY

Our method	gOp [41]	MV-MI [39]	SoftPOSIT [25]
0.49 ± 0.08	0.48 ± 0.09	1.29 ± 0.21	26.16 ± 45.46

Table II and Fig. 8, associated with a visual comparison using #14 image. Only the errors no more than 5 pixels are shown in the figure for a better comparison since the errors of SoftPOSIT in some failed cases were more than 100 pixels.

As shown in Table II, our method had the smallest RMS projection error of approximately 0.5 pixels (0.49 ± 0.08). The gOp produced the same accuracy (0.48 ± 0.09) as the proposed method. MV-MI had intermediate accuracy of approximately 1.3 pixels (1.29 ± 0.21). While SoftPOSIT produced the largest errors of 26.16 ± 45.46 pixels.

As a typical example, the registration results of the four methods for the #14 image is also shown in Fig. 8. Using the estimate of the proposed method, the projections of the beads on the FTRAC model (yellow asterisks) are very close to the centers of the fiducial beads (green points). The projections obtained using the estimate of gOp coincide with those of the proposed method. While using the estimate of MV-MI, the projections are a little bit further from the centers of the beads, and using the estimate of SoftPOSIT, some projections are quite far. As for

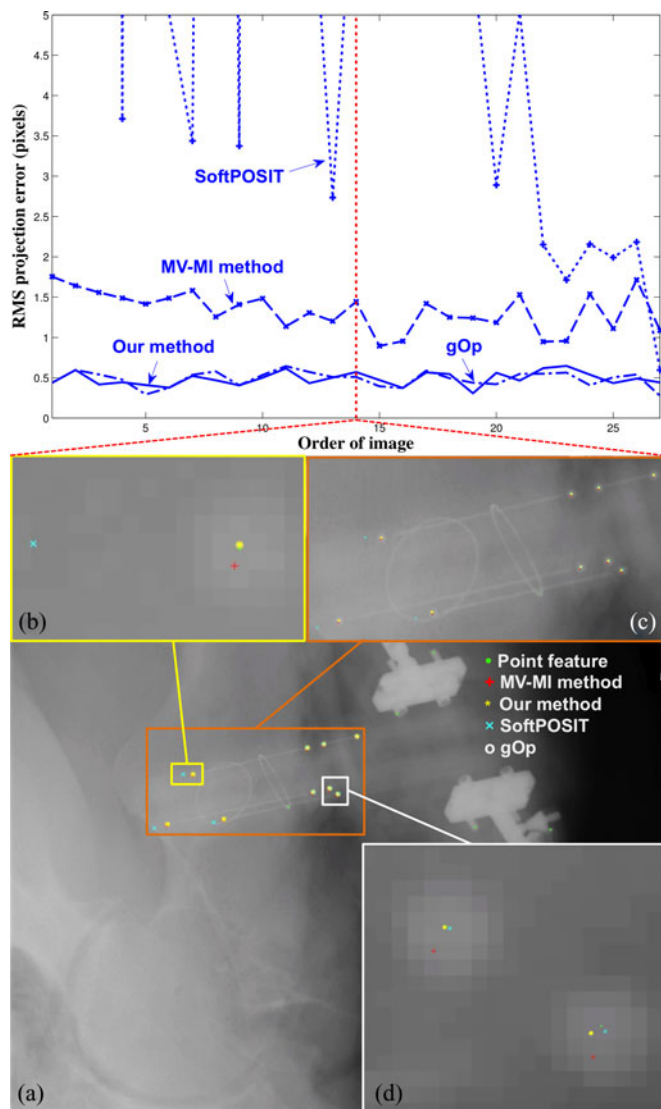


Fig. 8. RMS projection errors and an example registration results of different methods. In the image, the green points are extracted feature points, and the cyan crosses, the red plus signs, the white circles and the yellow asterisks are the projections using the estimate of SoftPOSIT [25], MV-MI [39], gOp [41], and the proposed method, respectively. The yellow asterisks overlapped the white circles, indicating the proposed method had the same accuracy as gOp.

the automatically detected centers of the fiducial beads, they are very close to the centers of the beads in the image [see Fig. 8(b) and (d)].

C. Computational Time

We used a laptop with 2.53 GHz Intel Core 2 Duo CPU and 4 GB memory for this project. The approximate computation time of SoftPOSIT and gOp was 1.8 and 1.3 s, respectively. The approximate computation time for the MATLAB implementation of our method was 2 s. This computation time included the automatic extraction of the fiducial beads, and can be improved using an C++ implementation.

VI. DISCUSSION

A. Comparison With SoftPOSIT

The proposed method has three essential differences from SoftPOSIT. First, $p_{mn} \in [0, 1]$ is the *posterior probability* of paired correspondences. Whereas, $\eta_{mn} \in \{0, 1\}$ in SoftPOSIT is a *binary assignment*. Although softassign is used, it ends up with a “zero-one assignment matrix” that specifies the paired correspondences between image and model points.

Second, a predefined distance threshold α in (3) is used to penalize mismatching when $\|\mathbf{u}_n - \mathcal{T}(\mathbf{X}_m; \boldsymbol{\theta})\|^2 > \alpha$. In our method, the correspondence is modeled using a Gaussian distribution. Although σ^2 can be interpreted as a “distance penalization” parameter, the use of Gaussian makes the penalization adaptive. More importantly, the penalization is driven by the data and optimized. In our method, σ^2 is initially set to a relaxed state (the half of image size), allowing the search of optimal solutions with a large tolerance, and subsequently tightened up, forcing the final solution to approach a minimum projection error. When σ^2 is large, our method loses precise local alignment and focuses on a more global geometry. When σ^2 decreases, the local geometry contributes more and drags the projections toward a better alignment.

Third, in the annealing scheme, the temperature β is simply increased monotonically. This has three shortcomings: 1) the optimal update ratio can be difficult to determine beforehand; 2) it is image- and initialization-related, i.e., a particular value of the update ratio (e.g., 1.05 in [25]) may not be optimal for different images or initializations; and 3) the optimal initial and stop values of β (0.0004 and 0.5 in [25]) are also difficult to determine in advance, which may potentially cause large registration errors. In our experiments, when the initialization was not good, SoftPOSIT produced large errors (see Table II). This is because that the search of SoftPOSIT is local, and there is no guarantee of finding the global optimum given a single initial guess [25]. For our method, σ^2 is optimized during each iteration, making our method robust to the initialization. In all experiments, we simply set σ^2 to the half of image size without influencing the registration accuracy. Since our method is a global optimizer, it produced accurate results.

B. Comparison With gOp

In both phantom and cadaver studies, our method and gOp produced essentially identical accuracy. But gOp requires *correct paired correspondences*, which is a common requirement for any PnP solver. Our method does not have this requirement. In the scenarios where the C-arm pose needs to be continuously estimated, manually establishing paired correspondences is time consuming and interrupts the procedure. In this case, our method is preferable because of its fully automatic fashion.

Our method can also be used when rough or partial paired correspondences are available. In this case, it can start with initial correspondence probabilities or a σ^2 that reflects the certainty of the available paired correspondences.

The gOp can always produce a global solution given correct paired correspondences due to its use of convex optimization.

Our method cannot theoretically guarantee to always produce a global solution, but PSO has demonstrated its capability of finding global solutions in many different applications, which is also shown in the intensive evaluations in this paper.

C. Comparison With MV-MI

Compared with MV-MI, our method has three advantages. First, MV-MI uses multiple images, requiring accurate spatial interrelationship of the images. This means the C-arm has to be tracked, and MV-MI is used to refine the tracking data. Our method uses only one image, leading to a pure image-guided scheme. Second, MV-MI needs to manually identify the beads in images and select their corresponding points on the 3-D model, while our method runs in an automatic manner. Third, our method has better accuracy than MV-MI.

D. Estimation Along Projection Direction

As shown in the experiments, the estimates had relatively large errors in the projection direction along the Z -axis (i.e., the depth). However, the depth can be difficult to accurately estimate for a 2-D-3-D registration method using only one image because the image does not contain distance information in the projection direction, and depth typically remains ambiguous given only local image features [47]. The movement of an object along the projection direction results in a scale change only in its 2-D projection, and it is difficult to accurately solve the ambiguity between the scale and depth using only one image. The limited accuracy in depth is an issue of the proposed method that need to be improved. One solution is to use two or more images if their relative poses are known.

E. Extension to Multiple Views

In general, the use of two or more images taken from different viewpoints will decrease the registration error along the projection direction. In this study, we did not implement multiview 2-D-3-D registration because the C-arm is not tracked in our system due to the large movement range of the C-arm compared with the high-precision tracking volume [39] and the interference of the line-of-sight. Similar setups were also used in other image-guided systems [48]–[50]. In such systems, the C-arm pose has to be accurately and robustly estimated from one image.

The propose method can be easily extended to multiview scenario if accurate spatial interrelationship between images are available. Having I views associated with the transformation $\boldsymbol{\theta}_1^i$ from the i th view to the first view, the multiview 2-D-3-D registration is formulated as

$$\frac{1}{2\sigma^2} \sum_{i=1}^I \sum_{n=1}^N \sum_{M=1}^M p_{imn} \|\mathbf{u}_n - \mathcal{T}(\mathbf{X}_m; \boldsymbol{\theta}, \boldsymbol{\theta}_1^i)\|^2 + C \log \sigma^2$$

where $\boldsymbol{\theta}_1^1 = \mathbf{I}$ is the identity matrix. Theoretically, any view can be chosen as the first view. After convergence, we can, as most multiview methods do in computer vision, further refine the registration using bundle adjustment [51]. This is one of our ongoing work.

F. Extension to Contour-Based 3-D–2-D Registration

The proposed method is not limited to fiducial-based 2-D–3-D registration. It can be readily adapted for contour-based 2-D–3-D registration. Representing image contours using oriented points and a 3-D object using a triangulated surface mesh, the proposed method can be extended to a contour-based 2-D–3-D registration method [52].

The contour-based method can also incorporate image gradient information in our probabilistic representations of image features. Finally, the 2-D–3-D registration is achieved by obtaining the optimal registration parameters using the same method that couples PSO and EM. Some preliminary studies along this direction have been reported in [53] and [52].

VII. CONCLUSION AND FUTURE WORK

We propose a 2-D–3-D registration method without the need for known paired correspondences. Unlike the methods that find the “best” correspondences from a set of candidate matchings, and then, estimate the registration from them, directly or iteratively, in the proposed method, every 3-D point has certain correspondence probabilities to all the 2-D points. Though this formulation in its nature is a combinatory explosion problem, it is solved in the proposed method by making use of a global optimization method coupled with statistical inference effectively. The correspondence probabilities are modeled using a mixture model, and then, the 2-D–3-D registration is solved using our method that couples PSO with EM, without the need for paired correspondences.

Compared with a global optimal method that uses correct paired correspondence, both the phantom and cadaver studies have showed that the proposed method is accurate, in terms of producing estimates with the same accuracy as a state-of-the-art global optimal method.

The experiments also illustrated the robustness of the proposed method to initializations and the presence of a large amount of outliers (false detections in the image). Presenting a large amount outliers (87 false detections versus 9 correct ones), and starting from a poor initialization, the proposed method also produced accurate estimates. Furthermore, the estimation accuracy was at the same level as the global optimal method that uses correct paired correspondences.

Although the evaluations were performed on one phantom and one cadaver, we believe the promising results on 27 images of the cadaver and 5000 trials on 100 images of the phantom demonstrate the efficacy of our method. This method will be further validated and verified in more cadaver and clinical studies in future.

In addition to extending the propose method to multiview and 2-D–3-D rigid registration, another important future work is to extend it to a contour-based 2-D–3-D deformable registration.

ACKNOWLEDGMENT

The authors would like to thank Dr. J. H. Siewerdsen and Dr. W. Zbijewski for their great help in acquiring CBCT images.

REFERENCES

- [1] K. Cleary and T. M. Peters, “Image-guided interventions: Technology review and clinical applications,” *Annu. Rev. Biomed. Eng.*, vol. 12, no. 1, pp. 119–142, 2010.
- [2] P. Markelj, D. Tomazević, B. Likar, and F. Pernuš. (2012.). A review of 3D/2D registration methods for image-guided interventions. *Med. Image Anal.* [Online]. 16(3), pp. 642–661. Available: <http://www.sciencedirect.com/science/article/pii/S1361841510000368>
- [3] A. Gueziec, P. Kazanzides, B. Williamson, and R. Taylor, “Anatomy-based registration of CT-scan and intraoperative X-ray images for guiding a surgical robot,” *IEEE Trans. Med. Imag.*, vol. 17, no. 5, pp. 715–728, Oct. 1998.
- [4] M. Fleute and S. Lavallée, “Nonrigid 3-D/2-D registration of images using statistical models,” in *Medical Image Computing and Computer-Assisted Intervention—MICCAI’99*, vol. LNCS 1679, A. C. C. Taylor, Ed. Berlin, Germany: Springer-Verlag, 1999, pp. 138–147.
- [5] J. Feldmar, N. Ayache, and F. Betting, “3D-2D projective registration of free-form curves and surfaces,” *Comput. Vis. Image Understand.*, vol. 65, no. 3, pp. 403–424, 1997.
- [6] G. Zheng, M. Ballester, M. Styner, and L.-P. Nolte, “Reconstruction of patient-specific 3D bone surface from 2D calibrated fluoroscopic images and point distribution models,” in *Medical Image Computing and Computer-Assisted Intervention—MICCAI’09*, ser. Lecture Notes in Computer Science, vol. 4190, R. Larsen, M. Nielsen, and J. Spurring, Eds. Berlin, Germany: Springer-Verlag, 2006, pp. 25–32.
- [7] G. Zheng, X. Dong, K. Rajamani, X. Zhang, M. Styner, R. Thoranaghatte, L.-P. Nolte, and M. Ballester, “Accurate and robust reconstruction of a surface model of the proximal femur from sparse-point data and a dense-point distribution model for surgical navigation,” *IEEE Trans. Biomed. Eng.*, vol. 54, no. 12, pp. 2109–2122, Dec. 2007.
- [8] G. Zheng, “Statistically deformable 2d/3d registration for accurate determination of post-operative cup orientation from single standard x-ray radiograph,” in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2009*, ser. Lecture Notes in Computer Science, vol. 5761, G.-Z. Yang, D. Hawkes, D. Rueckert, A. Noble, and C. Taylor, Eds. Berlin, Germany: Springer-Verlag, 2009, pp. 820–827.
- [9] G. Zheng, S. Gollmer, S. Schumann, X. Dong, T. Feilkas, and M. A. G. Ballester, “A 2D/3D correspondence building method for reconstruction of a patient-specific 3d bone surface model using point distribution models and calibrated x-ray images,” *Med. Image Anal.*, vol. 13, no. 6, pp. 883–899, 2009.
- [10] G. Zheng, “Statistical shape model-based reconstruction of a scaled, patient-specific surface model of the pelvis from a single standard ap x-ray radiograph,” *Med. Phys.*, vol. 37, no. 4, pp. 1424–1439, 2010.
- [11] B. K. P. Horn, H. M. Hilden, and S. Negahdaripour, “Closed-form solution of absolute orientation using orthonormal matrices,” *J. Opt. Soc. Amer. A*, vol. 5, no. 7, pp. 1127–1135, Jul 1988.
- [12] K. S. Arun, T. S. Huang, and S. D. Blostein, “Least-squares fitting of two 3-D point sets,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-9, no. 5, pp. 698–700, Sep. 1987.
- [13] D. F. Dementhon and L. S. Davis, “Model-based object pose in 25 lines of code,” *Int. J. Comput. Vis.*, vol. 15, pp. 123–141, 1995.
- [14] M. J. Murphy, “Fiducial-based targeting accuracy for external-beam radiotherapy,” *Med. Phys.*, vol. 29, no. 3, pp. 334–344, 2002.
- [15] X. Kang, R. H. Taylor, M. Armand, Y. Otake, W.-P. Yau, P. Y. S. Cheung, and Y. Hu, “Correspondenceless 3D-2D registration based on expectation conditional maximization,” in *Medical Imaging 2011: Image-Guided Procedures, Robotic Interventions, and Modeling*, vol. 7964, no. 1, K. H. Wong and D. R. H. III, Eds. Lake Buena Vista, FL, USA: SPIE, Feb. 12–17, 2011, pp. 79 642Z-1–79 642Z-8.
- [16] P. Besl and N. McKay, “A method for registration of 3D shapes,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 239–256, Feb. 1992.
- [17] T. Cootes, C. Taylor, D. Cooper, and J. Graham, “Active shape models—their training and application,” *Comput. Vis. Image Understand.*, vol. 61, no. 1, pp. 38–59, 1995.
- [18] M. A. Fischler and R. C. Bolles, “Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography,” *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [19] O. Enqvist and F. Kahl, “Robust optimal pose estimation,” in *Computer Vision—ECCV 2008*, ser. Lecture Notes in Computer Science, vol. 5302, D. Forsyth, P. Torr, and A. Zisserman, Eds. Berlin, Germany: Springer-Verlag, 2008, pp. 141–153.

- [20] O. Enqvist, K. Josephson, and F. Kahl, "Optimal correspondences from pairwise constraints," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 29, 2009–Oct. 2, 2009, pp. 1295–1302.
- [21] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, pp. 91–110, 2004.
- [22] H. Chui and A. Rangarajan, "A new point matching algorithm for non-rigid registration," *Comput. Vis. Image Understand.*, vol. 89, no. 2–3, pp. 114–141, 2003.
- [23] S. Gold and A. Rangarajan, "A graduated assignment algorithm for graph matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 4, pp. 377–388, Apr. 1996.
- [24] S. Gold, A. Rangarajan, C.-P. Lu, S. Pappu, and E. Mjølness, "New algorithms for 2D and 3D point matching: Pose estimation and correspondence," *Pattern Recognit.*, vol. 31, no. 8, pp. 1019–1031, 1998.
- [25] P. David, D. DeMenthon, R. Duraiswami, and H. Samet, "SoftPOSIT: Simultaneous pose and correspondence determination," *Int. J. Comput. Vis.*, vol. 59, no. 3, pp. 259–284, 2004.
- [26] F. Moreno-Noguer, V. Lepetit, and P. Fua, "Pose priors for simultaneously solving alignment and correspondence," in *Computer Vision—ECCV 2008*, ser. Lecture Notes in Computer Science, vol. 5303, D. Forsyth, P. Torr, and A. Zisserman, Eds. Berlin, Germany: Springer-Verlag, 2008, pp. 405–418.
- [27] C.-T. D. Lin, D. B. Goldgof, and W.-C. Huang, "Motion estimation from scaled orthographic projections without correspondences," *Image Vis. Comput.*, vol. 12, no. 2, pp. 95–108, 1994.
- [28] K.-I. Kanatani, "Tracing planar surface motion from a projection without knowing the correspondence," *Comput. Vis., Graph., Image Process.*, vol. 29, no. 1, pp. 1–12, 1985.
- [29] Y. Liu and M. A. Rodrigues, "Statistical image analysis for pose estimation without point correspondences," *Pattern Recognit. Lett.*, vol. 22, no. 11, pp. 1191–1206, 2001.
- [30] G. McLachlan and D. Peel. (2005). *Finite Mixture Models*, ser. Wiley Series in Probability and Statistics. New York, NY, USA: Wiley. [Online]. Available: <http://dx.doi.org/10.1002/0471721182.fmatter>
- [31] M. A. Tanner. (1996). *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*, ser. Springer series in statistics. New York, NY, USA: Springer-Verlag. [Online]. Available: <http://books.google.com/books?id=tXocinu3iOoC>
- [32] A. Myronenko and X. Song, "Point set registration: Coherent point drift," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 12, pp. 2262–2275, Dec. 2010.
- [33] S. Granger and X. Pennec, "Multi-scale EM-ICP: A fast and robust approach for surface registration," in *ECCV '02: Proceedings of the 7th European Conference on Computer Vision-Part IV*. London, U.K.: Springer-Verlag, 2002, pp. 418–432.
- [34] R. Horaud, F. Forbes, M. Yguel, G. Dewaele, and J. Zhang, "Rigid and articulated point registration with expectation conditional maximization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 587–602, Mar. 2011.
- [35] X.-L. Meng and D. B. Rubin, "Maximum likelihood estimation via the ECM algorithm: A general framework," *Biometrika*, vol. 80, pp. 267–278, 1993.
- [36] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. IEEE Swarm Intell. Symp. (SIS)*, Indianapolis, IN, USA, 1995, pp. 80–87.
- [37] M. Clerc and J. Kennedy, "The particle swarm—Explosion, stability, and convergence in a multidimensional complex space," *IEEE Trans. Evol. Comput.*, vol. 6, no. 1, pp. 58–73, Feb. 2002.
- [38] X.-F. Xie, W.-J. Zhang, and Z.-L. Yang, "Adaptive particle swarm optimization on individual level," in *Proc. 6th Int. Conf. Signal Process.*, Aug. 2002, vol. 2, pp. 1215–1218.
- [39] Y. Otake, M. Armand, R. Armiger, M. Kutzer, E. Basafa, P. Kazanzides, and R. Taylor, "Intraoperative image-based multi-view 2D/3D registration for image-guided orthopaedic surgery: Incorporation of fiducial-based C-arm tracking and GPU-acceleration," *IEEE Trans. Med. Imag.*, vol. 31, no. 4, pp. 948–962, Apr. 2012.
- [40] A. K. Jain, T. Mustafa, Y. Zhou, C. Burdette, G. S. Chirikjian, and G. Fichtinger, "FTRAC—A robust fluoroscope tracking fiducial," *Med. Phys.*, vol. 32, no. 10, pp. 3185–3198, 2005.
- [41] G. Schweighofer and A. Pinz, "Globally optimal o(n) solution to the pnp problem for general camera models," presented at the Brit. Machine Vision Conf., Leeds, U.K., 2008.
- [42] W. Zbijewski, P. D. Jean, P. Prakash, Y. Ding, J. W. Stayman, N. Packard, R. Senn, D. Yang, J. Yorkston, A. Machado, J. A. Carrino, and J. H. Siewerdsen. (2011). A dedicated cone-beam ct system for musculoskeletal extremities imaging: Design, optimization, and initial performance characterization. *Med. Phys.* [Online]. 38(8), pp. 4700–4713. [Online]. Available: <http://link.aip.org/link/?MPH/38/4700/1>
- [43] Y. Cho, D. J. Moseley, J. H. Siewerdsen, and D. A. Jaffray. (2005). Accurate technique for complete geometric calibration of cone-beam computed tomography systems. *Med. Phys.* [Online]. 32(4), pp. 968–983. Available: <http://link.aip.org/link/?MPH/32/968/1>
- [44] J. H. Siewerdsen, D. J. Moseley, S. Burch, S. K. Bisland, A. Bogaards, B. C. Wilson, and D. A. Jaffray, "Volume ct with a flat-panel detector on a mobile, isocentric c-arm: Pre-clinical investigation in guidance of minimally invasive surgery," *Med. Phys.*, vol. 32, no. 1, pp. 241–254, 2005.
- [45] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans. Med. Imag.*, vol. 16, no. 2, pp. 187–198, Apr. 1997.
- [46] J. A. Nelder and R. Mead, "A simplex method for function minimization," *Comput. J.*, vol. 7, no. 4, pp. 308–313, 1965.
- [47] A. Saxena, S. Chung, and A. Ng, "3-D depth reconstruction from a single still images," *Int. J. Comput. Vis.*, vol. 76, pp. 53–69, 2008.
- [48] A. Jain and G. Fichtinger, "C-arm tracking and reconstruction without an external tracker," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2006*, ser. Lecture Notes in Computer Science, vol. 4190, R. Larsen, M. Nielsen, and J. Sporring, Eds. Berlin, Germany: Springer-Verlag, 2006, pp. 494–502.
- [49] B. Kainz, M. Grabner, and M. Ruther, "Fast marker based C-arm pose estimation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2008*, ser. Lecture Notes in Computer Science, vol. 5242, D. Metaxas, L. Axel, G. Fichtinger, and G. Szekely, Eds. Berlin, Germany: Springer-Verlag, 2008, pp. 652–659.
- [50] E. Moutl, E. C. Burdette, D. Y. Song, P. Abolmaesumi, G. Fichtinger, and P. Fallavollita, "Automatic C-arm pose estimation via 2D/3D hybrid registration of a radiographic fiducial," in *SPIE Medical Imaging*, vol. 7964, no. 1, K. H. Wong and D. R. H. III, Eds. Lake Buena Vista (Orlando), FL, USA: SPIE, Mar. 2011, p. 79642S.
- [51] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon, "Bundle Adjustment—A modern synthesis," in *Vision Algorithms: Theory and Practice*, ser. Lecture Notes in Computer Science, vol. 1883, B. Triggs, A. Zisserman, and R. Szeliski, Eds. Berlin, Germany: Springer-Verlag, 2000, pp. 153–177.
- [52] X. Kang, W. P. Yau, Y. Otake, P. Y. S. Cheung, Y. Hu, and R. H. Taylor, "Assessing 3D tunnel position in ACL reconstruction using a novel single image 3D-2D registration," in *Medical Imaging 2012: Image-Guided Procedures, Robotic Interventions, and Modeling*, vol. 8316, no. 1, D. R. H. III and K. H. Wong, Eds. San Diego, CA, USA: SPIE, Feb. 4–9 2012, pp. 831 628-1–831 628-6. [Online]. Available: <http://link.aip.org/link/?PSI/8316/831628/1>
- [53] X. Kang, W.-P. Yau, R. H. Taylor, and Y. Otake, "A novel method in assessing tunnel positions in ACL reconstruction," presented at the 11th Annu. Meet. CAOS-Int., London, U.K., Jun. 15–19 2011

Authors' photographs and biographies not available at the time of publication.