Correspondenceless 3D-2D Registration Based on Expectation Conditional Maximization

X Kang[a], RH Taylor[b], M Armand[c,d], Y Otake[b], WP Yau[a], PYS Cheung[e], Y Hu[a]

[a]Dept. of Orthopaedics and Traumatology, The University of Hong Kong
[b]Dept. of Computer Science, The Johns Hopkins University
[c]Applied Physics Laboratory, The Johns Hopkins University
[d]Dept. of Orthopaedic Surgery, Johns Hopkins Bayview Medical Center
[e]Dept. of Electrical and Electronic Engineering, The University of Hong Kong

ABSTRACT

3D-2D registration is a fundamental task in image guided interventions. Due to the physics of the X-ray imaging, however, traditional point based methods meet new challenges, where the local point features are indistinguishable, creating difficulties in establishing correspondence between 2D image feature points and 3D model points. In this paper, we propose a novel method to accomplish 3D-2D registration without known correspondences. Given a set of 3D and 2D unmatched points, this is achieved by introducing correspondence probabilities that we model as a mixture model. By casting it into the *expectation conditional maximization* framework, without establishing one-to-one point correspondences, we can iteratively refine the registration parameters. The method has been tested on 100 real X-ray images. The experiments showed that the proposed method accurately estimated the rotations ($< 1°$) and in-plane (X-Y plane) translations (< 1 mm).

Keywords: 3D-2D registration, expectation conditional maximization, mixture of Gaussian.

1. INTRODUCTION

Estimating the registration parameters given a set of 3D object and 2D image feature points is a well understood problem when the correspondences are given. In the registration methods based on point-to-point correspondences, the projection error is commonly used in the objective function that is minimized to estimate the 3D rotation $\boldsymbol{R}$ and translation $\boldsymbol{t}$. This error measure is generally defined as the (Euclidean) distance between the feature points detected in the image (landmarks, edge points, contour points, and/or high dimensional feature points) and the projections of their counterparts on the 3D model. The procedure of getting 2D projection from 3D model, $\mathcal{O} \subset \mathbb{R}^3$, is a nonlinear mapping, $\mathcal{T} : \mathcal{O} \mapsto \Omega$, where $\Omega \subset \mathbb{R}^2$ is the image plane. Specifically, it is a 3D rigid motion $g = [\boldsymbol{R}, \boldsymbol{t}] : \mathbb{R}^3 \mapsto \mathbb{R}^3, \boldsymbol{R} \in SE(3), \boldsymbol{t} \in \mathbb{R}^3$, followed by a perspective projection $\pi : \mathbb{R}^3 \mapsto \Omega$, i.e. $\mathcal{T} = \pi \circ g$. Therefore, the estimation becomes the optimization problem

$$\{\hat{\boldsymbol{R}}, \hat{\boldsymbol{t}}\} = \arg \min_{\boldsymbol{R}, \boldsymbol{t}} \sum_{n=1}^{N} \|\boldsymbol{x}_n - \mathcal{T}(\boldsymbol{X}_n; \boldsymbol{R}, \boldsymbol{t})\|^2 . \tag{1}$$

where $\boldsymbol{x}_n \in \Omega$ and $\boldsymbol{X}_n \in \mathcal{O}$ are corresponding 2D and 3D points.

Vast amount of methods were proposed to solve this problem in computer vision. The registration is generally achieved in two steps: 1) finding point-to-point correspondences and 2) estimating the transformation parameters by solving the optimization problem (1). One essential requirement

of this kind of methods is the establishment of one-to-one point correspondences between 2D image points and 3D model points. Certain strategies for this purpose have been proposed by using a local feature descriptor, such as Harris, SIFT, MSER, or a more complex application specific feature, and a mismatching elimination technique, such as RANSAC (*RANdom SAmple Consensus*), extended Kalman filter, or particle filter. The registration can also be achieved in an iterative manner. The classic method in this class is the *Iterative Closest Point* (ICP) algorithm, where in each iteration the point-to-point correspondence is established through a function that associates to a image point its closest point on the 3D model.

However, the problem is still difficult in practical X-ray image based interventions. Anatomical landmarks are often difficult to locate in X-ray images and the corresponding 3D features may be ambiguously defined. Problems present even where fiducials are used: Anatomy may obscure even strong fiducials; several fiducials may have the same appearance; fiducials may be occluded; and intraoperative clutter may introduce false fiducials in the images. Moreover, there are cases where it is difficult to establish the required correspondences, for example because the 3D points are simply features on a CAD model without associated texture. Therefore, when the correspondences cannot be established *a priori*, a method that can estimate the transformation without the need of known correspondences is desired.

Several correspondenceless registration methods were proposed. However, they either use simplified camera models[1] or have special requirements,[2,3] and cannot be employed in our scenario. The transformation parameters were estimated[2] by assuming the object consists of planar surfaces with closed curves drawn on them. A closed-form solution for 3D-2D registration without point-to-point correspondences[3] was proposed. The method, however, has a very strong constraint: the camera can only has in-plane translation and rotation (i.e., moving in the XY-plane and rotating around the optical axis parallel to Z-axis). Another class of methods adapts well-known ICP algorithm based on point-to-line distance.[4] However, ICP can be easily trapped in local minima due to the binary assignment of one-to-one correspondence, making it sensitive to initialization. To simultaneously estimate the transformation and correspondences, SoftPOSIT is arguably the most efficient and accurate algorithm, which combines an iterative pose estimation method and an soft-assignment technique. However, it tends to fail in the presence of large amounts of cluster and occlusions.

In this paper we propose a method to achieve the 3D-2D registration in a probability framework without the requirement to impose one-to-one point correspondences. Taking fiducial-based registration as an example, we report the result of our experiments demonstrating that our method can obtain accurate registration under the conditions where undistinguishable fiducials were occluded or obscured by others and/or influenced by falsely detected fiducials, even thought the one-to-one correspondence probabilities are not high for all points.

2. CORRESPONDENCELESS 3D-2D REGISTRATION

2.1 The Basic Idea

Given a set of 2D points, $\{\boldsymbol{x}_n\}_1^N$, on the image and a set of 3D points, $\{\boldsymbol{X}_m\}_1^M$, on a 3D model. These two sets of points are in different dimensions and generally have different numbers of points. Without knowing the correspondences between them, it is common to assume that the n-th 2D point, $\boldsymbol{x}_n$, has certain probability p_{mn} to correspond to the m-th 3D point, $\boldsymbol{X}_m$, given the current

transformation parameter $\boldsymbol{\theta} = \{\boldsymbol{R}, \boldsymbol{t}\}$. There, by introducing this probability the optimization problem (1) becomes

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \sum_{n=1}^{N} \sum_{m=1}^{M} p_{mn} \left\| \boldsymbol{x}_n - \mathcal{T}(\boldsymbol{X}_m; \boldsymbol{\theta}) \right\|^2. \tag{2}$$

In this work, the correspondence probability is modeled as a 2D isotropic Gaussian function with variance σ^2 and taking the projected 3D point as its mean. The variance is also a parameter that will be optimized. To account for outliers, we adopt a uniform distribution since there is no prior information for them and they could be anywhere in the image. Several work[5,6] used the similar methodology, but none of them is for 3D-2D registration. Ideally, two point sets become overlapped and σ^2 approaches to zero. And once the optimal transformation has been obtained, the point-to-point correspondence could be obtained by maximizing the a posterior probability. But there is no one-to-one point correspondence used in our registration algorithm.

2.2 The Objective Function

Using the intuitive idea to formulate the object function and then solve it is not so straightforward. In this section, we are going to show that, by casting it into a probability framework, the target object function can be formulated and solved in a tractable way.

We model the correspondence probability $p\left(\boldsymbol{x}_n | m, \boldsymbol{\theta}, \sigma^2\right) = \mathrm{N}\left(\mathcal{T}(\boldsymbol{X}_m; \boldsymbol{\theta}), \sigma^2\right)$ a Gaussian distribution, and $p(\boldsymbol{x}_n | M+1) = 1/N$ a uniform distribution. Consequently, with the i.i.d. assumption, the *observed-data* likelihood function reads

$$L_{obs}(\boldsymbol{\theta}, \sigma^2 | \boldsymbol{x}_n) = \prod_{n=1}^{N} \left(\sum_{m=1}^{M+1} p(\boldsymbol{x}_n | m, \boldsymbol{\theta}, \sigma^2) P(m) \right), \tag{3}$$

where $P(m)$ is the correspondence prior, which equals to $1/M$ since there no prior information for it. Directly solving the maximum likelihood estimation (MLE) of $\boldsymbol{\theta}$ and σ^2 from (3) is extremely difficult. Instead, we treat the correspondence as latent data as if we knew that $\boldsymbol{x}_n$ corresponds to $\boldsymbol{X}_m$, i.e., the probability of that we observed data $\boldsymbol{x}_n$ and knew that it was drawn from the m-th distribution. Thus, the *complete-data* log-likelihood simplifies to

$$\ell_{com}(\boldsymbol{\theta}, \sigma^2 | \boldsymbol{x}_n, m) = \log \left(\prod_{n=1}^{N} p(\boldsymbol{x}_n | m, \boldsymbol{\theta}, \sigma^2) P(m) \right), \tag{4}$$

which is much more tractable. Taking the expectation of ℓ_{com} with respect to the prior correspondence probability given the observations and latent data, and ignoring the constants independent of $\boldsymbol{\theta}$ and σ^2, we obtained our objective function

$$Q(\boldsymbol{\theta}, \sigma^2) = \frac{1}{2\sigma^2} \sum_{n=1}^{N} \sum_{M=1}^{M} p_{mn} \left\| \boldsymbol{x}_n - \mathcal{T}(\mathbf{X}_m; \boldsymbol{\theta}) \right\|^2 + C \log \sigma^2, \tag{5}$$

where $C = \sum_{m,n=1}^{M,N} p_{mn}$ and p_{mn} is the posterior probability of the correspondence, which can be calculated using Bayesian formula. This result coincides with the procedure that we use to register a 3D object to its 2D image: We generally could not register the 3D object with its 2D projection in one trial, but progressively. Each time after (or during) rotating and translating the 3D object, we try to find the correspondences and adjust our correspondence criteria (reflected by σ^2). This is the reason why σ^2 is also optimized during the registration in our method.

2.3 The Registration Algorithm

To find the MLE of $\boldsymbol{\theta}$ and σ^2, we adopt the *expectation maximization* (EM) framework and define (5) as the surrogate function. Minimizing the surrogate function (5) necessarily decreases (4). The EM algorithm is an iterative method for locating posterior modes by minimizing the expectation of the complete-data log-likelihood with respect to the "current" parameter values. It proceeds by alternating between the estimation (E-step) and the maximization (M-step), until convergence.

In the E-step of our method, p_{mn} is estimated via the conventional way. However, there are two major difficulties which prevent us from using the conventional M-step. First, the means of each Gaussian component are constrained by the transformation parameters $\boldsymbol{\theta}$. Second, the estimate of $\boldsymbol{\theta}$ further conditions on the distribution parameter σ^2. To tackle these difficulties, we employ the *expectation conditional maximization* (ECM) algorithm,[7] in which the M-step is replaced by a series of conditional maximization (CM) steps. Theoretically, any ECM is a generalized EM algorithm and it preserves the monotone convergence property. Thus, given the "current" transformation parameter values, one can easily get the conditional MLE of σ^2 as

$$\hat{\sigma}^2 = \frac{1}{2C} \sum_{m=1}^{M} \sum_{n=1}^{N} p_{mn} \|\boldsymbol{x}_n - \mathcal{T}(\boldsymbol{X}_m; \boldsymbol{\theta})\|^2. \tag{6}$$

However, when maximizing (5) over $\boldsymbol{\theta}$ conditioned on σ^2, the CM-step still cannot be accomplished analytically. Fortunately, one could carry out this CM-step using numerical optimization approach, and ECM may still be computationally simpler and more stable because it involves lower-dimensional maximizations than EM. In addition, since the computation of p_{mn} is light, we insert the E-step between each pair of CM-steps, thereby updating (5) at every stage of the CM cycle. To this end, we have the following method:

1. Initialize $\boldsymbol{\theta}^{(0)}$ and $\left(\sigma^2\right)^{(0)}$;
2. *E-step:* Compute the posterior modes $p_{mn}^{(t)}$;
3. *CM-step:* Compute the following CM-cycle:
    - Compute $\left(\sigma^2\right)^{(t+1)}$ by (6) using $\boldsymbol{\theta}^{(t)}$ and $p_{mn}^{(t)}$;
    - Compute $\boldsymbol{\theta}^{(t+1)}$ using an numerical optimization approach, conditioned on $(\sigma^2)^{(t+1)}$;
4. Check convergence.

Practically, in order for ECM to be convergent, one needs to ensure that each CM cycle effectively maximizes over the original parameter space and not over some subspace. In this work, the *pattern search* algorithm[8] was used. However, our method is not limited to this specific adaptation; other numerical optimization methods could be substituted. Convergence is guaranteed by the principle of the generalized EM.

3. EXPERIMENTS AND RESULTS

The experiments were performed using X-ray images acquired from a fiducial attached on a sawbone by a well calibrated C-arm bench system (see Fig. 1). In X-ray images, the fiducial occupies a relatively small portion and overlaps with the sawbone. In addition, there are also some metal beads affixed on the sawbone. Most importantly, the metal beads on the fiducial (as well as on

the sawbone) are not distinguishable from each other. These conditions pose a hurdle in intensity-based registration method. Our method achieved the 3D-2D registration using the nine beads on the fiducial, without knowing the correspondences between the beads on the model and the points detected in the X-ray images. Consequently, the six DOF transformation between the C-arm detector and the fiducial was estimated.
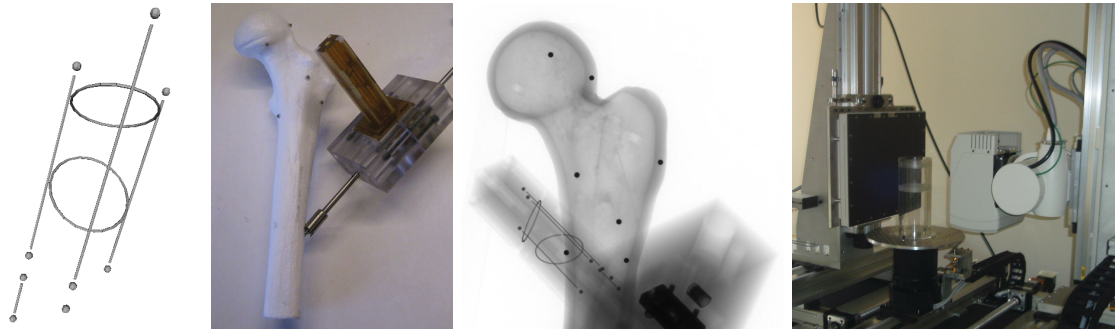

Figure 1: The CAD model of the fiducial (left), the sawbone with the fiducial (middle left) and one typical X-ray image of it (middle right) taken by the bench system (right).

The ground truth of the transformations was obtained from high resolution CT images. Each bead in the CT data was manually segmented and fitted by a sphere, so that 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 beads. The FRE of this procedure for all the nine beads is 0.1428 ± 0.0589 mm.

The experiments were carried out on 100 X-ray images taken from different view points. For each image, fifty trials were conducted using fifty different initial guesses. The initial guesses were randomly and independently selected from transformations uniformly distributed within $\pm 10°$ and ± 15 mm of the ground truth. The mean errors and their standard deviations of each parameter over 50 trials are shown in Fig. 4. The rotation errors were represented in angles for easy reading. For fair comparison, fixed number of iterations was used. The mean errors and standard deviations of each parameter over 50 trials were $(0.18 \pm 3.52, 0.19 \pm 2.41, 0.51 \pm 4.00)$ degrees and $(0.32 \pm 0.75, -0.11 \pm 0.75, 1.17 \pm 4.95)$ mm in rotations and translations, respectively. As can be seen in Fig. 4, our proposed method estimated the rotations and in-plane (X-Y plane) translations with high accuracy. Basically, the differences in rotations between the estimate and the ground truth were less than 1 degree, and the the differences in in-plane translations between the estimate and the ground truth were less than 1 mm. The estimates had relatively large error in the off-plane translation (i.e. the depth along Z-direction). However, it is known that the depth can be difficult to be accurately estimated for a 3D-2D registration algorithm using one single image. This could be significantly improved when using two or more images whose relative positions are known. Fig. 4 also shows large deviations in some cases. This is due to the fact that, when the outliers and inliers formed similar geometric structure as the fiducial beads, and the random guesses were just near that structure, the registration errors became relatively large, leading to large deviations.

To further illustrating the robust of our method to outliers, one selected trial is shown in Fig. 2 (a-b). The errors of the initial guess were $(-3.15°, -3.15°, -3.15°)$ and $(-9.44, -6.29, -12.59)$ mm in rotation and translation, respectively. Although the beads were self-clustered in the image and

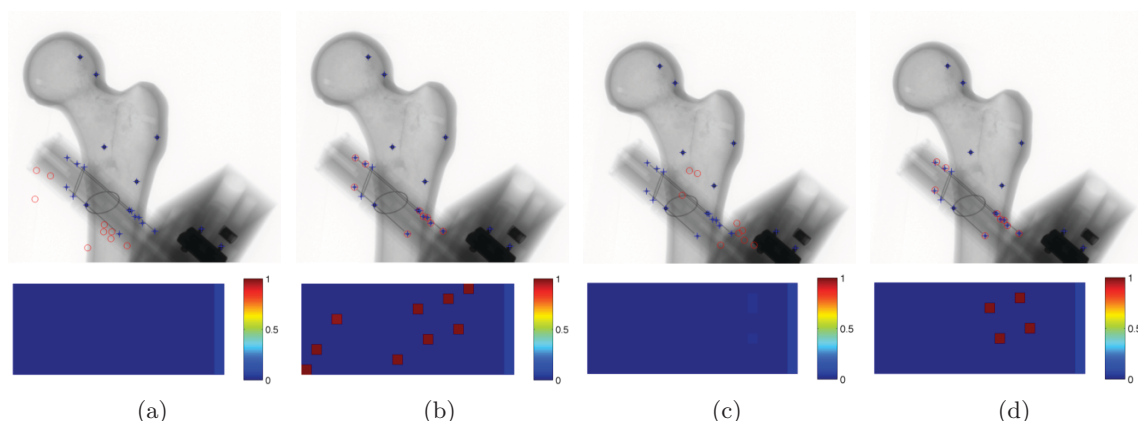<div align="center">(a)   (b)   (c)   (d)</div>

Figure 2: Two trials with different initial guesses (a), (c) and their registration results (b), (d) on the same X-ray image. The top row shows the detected beads and the projections of the beads on CAD model of the fiducial. The bottom row displays the corresponding map.

were with some outliers (the false detections and other beads on the sawbone), our method achieved the desired registration result, with errors of $(-0.11°, 0.16°, -0.02°)$ and $(0.14, -0.05, 2.74)$ mm. Further, the final one-to-one correspondence was not crucial to the accuracy of our method. Fig. 2 (c-d) dipict another trial on the same image. The errors of the initial guess were $(9.01°, -8.05°, -8.90°)$ and $(13.51, -12.08, -8.90)$ mm while the registration errors read $(-0.16°, 0.08°, -1.14°)$ and $(0.30, -0.09, 1.04)$ mm. Fig. 3 illustrates the robustness of our method to the missing points in addition to outliers. In the X-ray image, two beads on the fiducial were obscured by the frame of the fiducial. In addition, the initial guess introduced falsely matched correspondences, as shown in the correspondence map. In this case, our method still achieved a desirable registration with errors of $(-0.170°, 0.052°, 0.045°)$ and $(0.484, -0.023, 4.251)$ mm, respectively, compared with the initial errors of $(-4.430°, 0.938°, 9.150°)$ and $(-6.645, 1.406, 13.725)$ mm.

4. CONCLUSION

We introduce a novel 3D-2D registration algorithm without known correspondences. Unlike the methods that iteratively find the most likely correspondences from a set of candidates and then estimate the registration, our method assumes that every 3D point has certain correspondence probabilities with all the 2D points. Modeling the correspondence probabilities using a mixture model integrating Gaussian distributions and a uniform distribution, the correspondenceless 3D-2D registration is achieved using a ECM algorithm. The experiments using real X-ray images has shown that, our method is robust to outliers and the low probabilities for final one-to-one correspondence does not affect the registration result much.

REFERENCES

[1] Lin, C.-T. D., Goldgof, D. B., and Huang, W.-C., "Motion estimation from scaled orthographic projections without correspondences," *Image and Vision Computing* **12**(2), 95 – 108 (1994).
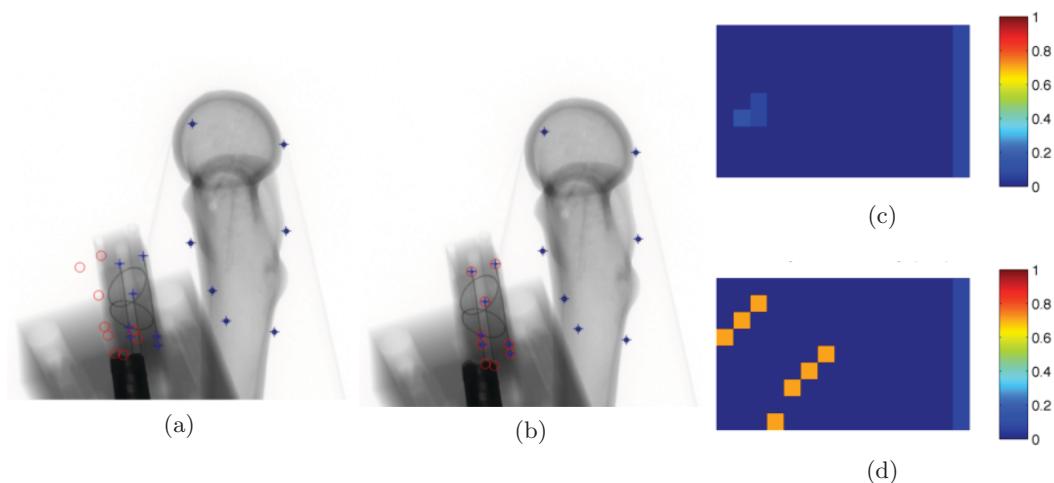
Figure 3: The registration result (b) with a random initial guess (a) for an image in which two beads were obscured by the frame of the fiducial. Although the initial guess introduced additional two falsely matched correspondences, our method still obtained desired registration result.

[2] Kanatani, K.-I., "Tracing planar surface motion from a projection without knowing the correspondence," *Computer Vision, Graphics, and Image Processing* **29**(1), 1 – 12 (1985).

[3] Liu, Y. and Rodrigues, M. A., "Statistical image analysis for pose estimation without point correspondences," *Pattern Recognition Letters* **22**(11), 1191 – 1206 (2001).

[4] Fleute, M. and Lavallée, S., "Nonrigid 3-D/2-D registration of images using statistical models," in [*MICCAI'99*], C. Taylor, A. C., ed., **LNCS 1679**, 138–147, Springer-Verlag, Berlin Heidelberg (1999).

[5] Breitenreicher, D. and Schnörr, C., "Robust 3d object registration without explicit correspondence using geometric integration," *Machine Vision and Applications* **21**, 601–611 (08 2010).

[6] Myronenko, A. and Song, X., "Point set registration: Coherent Point Drift," **Early Access**(99) (2010).

[7] Meng, X.-L. and Rubin, D. B., "Maximum likelihood estimation via the ECM algorithm: a general framework," *Biometrika* **80**, 267–278 (1993).

[8] Audet, C. and Jr., J. E. D., "Analysis of generalized pattern searches," *SIAM Journal on Optimization* **13**(3), 889–903 (2003).

[9] David, P., DeMenthon, D., Duraiswami, R., and Samet, H., "SoftPOSIT: Simultaneous pose and correspondence determination," *International Journal of Computer Vision* **59**, 259–284 (Sept. 2004).

[10] Gold, S. and Rangarajan, A., "A graduated assignment algorithm for graph matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **18**, 377–388 (1996).

[11] Gold, S., Rangarajan, A., Lu, C.-P., Pappu, S., and Mjolsness, E., "New algorithms for 2d and 3d point matching: pose estimation and correspondence," *Pattern Recognition* **31**(8), 1019 – 1031 (1998).

[12] Dementhon, D. F. and Davis, L. S., "Model-based object pose in 25 lines of code," *International Journal of Computer Vision* **15**, 123–141 (1995). 10.1007/BF01450852.
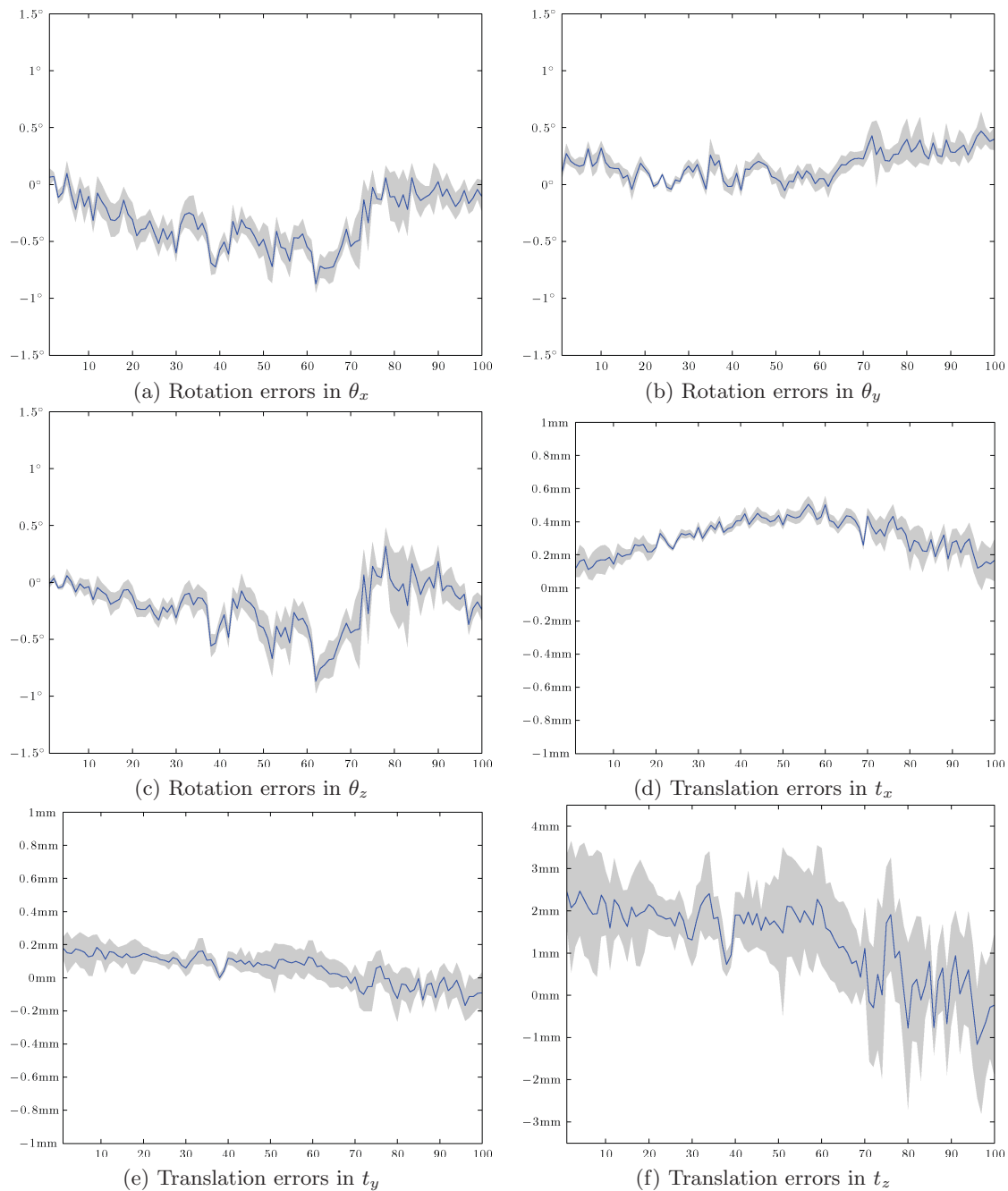
(a) Rotation errors in θ_x

(b) Rotation errors in θ_y

(c) Rotation errors in θ_z

(d) Translation errors in t_x

(e) Translation errors in t_y

(f) Translation errors in t_z

Figure 4: The statistics of registration errors in rotations about X-, Y-, and Z-axis and in translations along X-, Y-, and Z-axis.