# Generalized Point Set Registration with Fuzzy Correspondences Based on Variational Bayesian Inference

Ang Zhang, Zhe Min, Zhengyan Zhang, and Max Q.-H. Meng, *Fellow, IEEE*

***Abstract*—Point set registration (PSR) is an essential problem in surgical navigation and computer-assisted surgery (CAS). In CAS, PSR can be used to map the intra-operative surgical space with the pre-operative volumetric image space. The performances of PSR in real-world surgical scenarios are sensitive to noise and outliers. This paper proposes a novel point set registration approach where the additional features (i.e., the normal vectors) extracted from the point sets are utilized, and the convergence of the algorithm is guaranteed from the theoretical perspective. More specifically, we formulate the PSR with normal vectors by generalizing the Bayesian coherent point drift (BCPD) into the six-dimensional scenario. The proposed algorithm is more accurate and robust to noise and outliers, and the theoretical convergence of the proposed approach is guaranteed. Our contributions of this paper are summarized as follows. (1) The PSR problem with normal vectors is formally formulated through generalizing the BCPD approach; (2) The formulas for updating the parameters during the algorithm's iterations are given in closed forms; (3) Extensive experiments have been done to verify the proposed approach and specifically its significant improvements over the BCPD has been validated.**

***Index Terms*—Computer-assisted surgery (CAS), point set registration, fuzzy correspondence estimation, variational Bayesian inference.**

## I. Introduction

POINT set registration (PSR) is an essential problem in research areas of robotics, computer vision [1], augmented reality [2][3], and computer-assisted surgery (CAS) [4][5]. The PSR aims to estimate the best transformation and correspondences between two point sets (PSs). PSR is a critical step in CAS to map the intra-operative space with the preoperative volumetric image space [6]. In computer-assisted orthopedic surgery (CAOS), the image to patient registration can be broadly classified into two categories: (1) fiducial-based registration; (2) surface-based registration. In fiducial-based registration, artificial fiducials (e.g., either steel implants or landmarks) are first localized in two spaces, and then the 3D position sets with known correspondence are registered together. In surface-based registration, two surfaces usually represented by point sets are registered together. The methods in the first category have their corresponding drawbacks: (1) the artificial implants are invasive to the patients; (2) the anatomical landmarks are error-prone to locate in both spaces; (3) other skin-attached markers can easily move. In contrast, surface-based registration (mainly point set registration) eliminates the additional step of attaching artificial markers or implants to the patient, which brings inconvenience to physicians and also has the potential to enhance the registration accuracy. However, PSR in CAS faces several challenges. First, the intra-operative data is susceptible to noise and outliers which can negatively affect the registration accuracy. Second, the convergence of one PSR algorithm should be guaranteed in both theoretical and practical aspects. The encountered noise and outliers in practical applications influence adversely both the correspondence stage where the correspondences between points in two PSs are constructed, and the transformation stage where the desired transformation is estimated.

Registration can be categorized into rigid registration [7][8] and non-rigid registration [9]–[12]. Due to the rigid nature of bones, rigid registration is used in CAOS. The iterative closest point (ICP) [13] algorithm is widely used to solve the rigid registration problem. It takes an iterative approach: first, to solve the optimal correspondences between points from two PSs; second, to update the rigid transformation. ICP iterates these two steps until a certain convergence condition is satisfied. Despite its many successes, ICP has several drawbacks, including: 1) being sensitive to noises, outliers, and the initial parameter values; 2) being easily trapped into local optimization [14]. We analyze that the above-mentioned drawbacks are attributed to the following: (1) only positional information is used; (2) the hard one-to-one correspondence strategy is used.

Unlike the binary correspondence strategy, fuzzy correspondence estimation can achieve a caution registration [15][16]. In fuzzy sets theory, elements in fuzzy sets have varying degrees of membership between 0 (False) and 1 (True) [17]. To

This project is partially supported by National Key R&D program of China with Grant No. 2019YFB1312400, Hong Kong RGC GRF grants #14211420, and Hong Kong RGC NSFC/RGC Joint Research Scheme #N_CUHK448/17 awarded to Max Q.-H. Meng. *(Ang Zhang and Zhe Min contributed equally to this work.) (Corresponding author: Max Q.-H. Meng.)*

Ang Zhang is with the Department of Electronic Engineering, The Chinese University of Hong Kong, N.T., Hong Kong SAR, China (e-mail: azhang@link.cuhk.edu.hk).

Zhe Min is with the Department of Medical Physics and Bioengineering, University College London, London, United Kingdom (e-mail: z.min@ucl.ac.uk).

Zhengyan Zhang are with the School of Mechanical Engineering and Automation, Harbin Institute of Technology at Shenzhen, Shenzhen 518055, China.

Max Q.-H. Meng is with the Department of Electronic and Electrical Engineering of the Southern University of Science and Technology in Shenzhen, China, on leave from the Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong, and also with the Shenzhen Research Institute of the Chinese University of Hong Kong in Shenzhen, China (e-mail: max.meng@ieee.org).

overcome the challenges in PSR for CAS and the drawbacks of ICP-based methods, we extract the features (the normal vectors) from the raw point set and utilize them in both the fuzzy correspondence and transformation estimation stages. More specifically, the fuzzy point correspondences are constructed in a soft manner, which means one point in the first PS $\mathbf{X}$ may correspond to many points in the second PS $\mathbf{Y}$. The potential benefit of using more useful information from the raw PS is that the transformation and the fuzzy correspondences can be recovered in a more precise manner. The soft-assignment technique makes the algorithm more robust to noise and outliers.

In this paper, we propose a novel rigid PSR approach where the normal vectors extracted from the raw PSs are utilized, and the algorithm's convergence is guaranteed in the theoretical aspect. A generalized PS is composed of positional points and corresponding normal vectors. In real-world scenarios where generalized PSs are registered, the normal vectors can be readily obtained with approaches like these: 1) using the PCA method for a point and its neighbors [18]; 2) using a tracked probe with a force/torque sensor or other approaches for range sensing [19]. Experimental results indicate that the normal vectors can help significantly to make the PSR more robust and accurate. Particularly, we formulate the generalized PSR as a variational Bayesian inference (VBI) problem [20] by utilizing the Bayesian coherent point drift (BCPD) framework [21]. Naturally, the proposed method is named generalized BCPD (GBCPD). A hybrid mixture model (HMM) is presented in GBCPD. HMM is composed of a Gaussian Mixture Model (GMM) and a Fisher Mixture Model (FMM) [22]. It is a probabilistic model to estimate fuzzy correspondences. Thereinto, the FMM is defined to represent the error distribution relating to the normal vectors, while th GMM is chosen to model the distribution associated with the positional localization error. The maximum of the posterior probability of the overall HMM is approximated by an alternative distribution using VBI. The important benefit over treating PSR as a maximum likelihood estimation (MLE) problem is that our GBCPD can guarantee theoretical convergence under variational inference [21].

In this paper, we cast the generalized point set registration problem into a VBI framework for the first time. To summarize, the main contributions of this paper include:

1) The PSR problem with normal vectors is formally formulated through generalizing the state-of-the-art Bayesian Coherent Point Drift (BCPD) approach, whose convergence is guaranteed in a theoretical manner;
2) The formulas to update the relevant parameters during the algorithm's iterations are derived and given in closed forms;
3) We have conducted extensive experiments on human femur bone models to verify the effectiveness of the proposed approach and its significant improvements over the state-of-the-art registration methods.

## II. Related Work

Researchers have put in a lot of effort to tackle the PSR problem. We briefly review those approaches and their contributions in the following three parts.

### A. ICP and Its Variants

ICP is the most commonly used and effective method in rigid registration. Two steps (correspondence estimation and update of transformation) are involved in the ICP method. With the one-to-one correspondence strategy and limited information, ICP relies on the stringent assumptions of small disturbance (noise and outliers) and an acceptable initial alignment which are different to meet in practical applications. Lots of research has been presented to improve the ICP's performance [23]. Two previous studies using the ICP framework combine both positional points with orientational vectors in a probabilistic model to improve the robustness of the 3D/3D rigid registration problem [19][24]. Moreover, some algorithms aim to solve a globally optimal solution. Go-ICP [25] utilizes Branch-and-Bound (BnB) method [26] to find the globally optimal result while it will improve computational cost. Recently, Anderson acceleration (AA) [27] has been adopted to improve the convergence speed of ICP. However, because those methods belong to the ICP framework, they still suffer heavily from disturbance.

### B. Probabilistic Algorithms

A fuzzy assignment strategy is adopted in the category of probabilistic PSR algorithms, where the fuzzy correspondences in two PS are represented with probabilities in $[0, 1]$. Coherent Point Drift (CPD) [28] is one typical probabilistic PSR method that formulates PSR as a maximum likelihood problem based on a GMM in CPD. The fuzzy correspondences are built based upon the motion coherence theory [29]. The EM algorithm is adopted for the maximum likelihood problem, which means CPD alternates between the calculation of the posterior of latent variables in E step and updates of the alignment in M step. A generative model is presented to jointly register multiple PSs in JRMPC [30]. Its performance exceeds CPD and other probabilistic models such as GMMReg [31].

Very recently, one variant of CPD, named Bayesian CPD (BCPD), adopted the VBI method to reformulate CPD as an analytical approximation problem [21]. The motion coherence theory of CPD is replaced by variational inference. The performances of the convergence, acceleration, and parameter tuning are improved. Without extra features, those methods, including CPD, JRMPC, and BCPD, are not robust to noise and outliers.

### C. Deep Learning Methods

In recent years, deep-learning-based methods have been applied to PS problems, such as PointCNN [32] and DGCNN [33]. Registration is also one of the challenging topics that use deep learning. Generally, point sets will be cast to a high-dimensional space to learn features. Then the correspondences can be generated from learned features, and the best transformations will be estimated. PR-Net [34] estimates the soft correspondence from features learned by DGCNN and optimizes the alignment in an end-to-end manner. PointDSC [35] introduces the constraint of the pairwise similarity to estimate correspondences.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2022.3159099, IEEE Transactions on Fuzzy Systems



Although deep PSR methods bring many improvements in different aspects, this strategy still has severe limitations. First, even the SOTA deep PSR approaches are difficult to generate acceptable inlier ratios for real-world applications [36] due to the equal treatment between inliers and outliers in the correspondence estimation; second, many scenarios exist untagged point cloud data, which cannot satisfy many supervised approaches; third, for surgical scenarios, deep PSR methods are unable to give error bounds either theoretically or practically, which is important for surgeons. By comparison, the probabilistic methods can solve effectively the PSR problems with different inliers rates and noise levels. More importantly, there are complete methods (*e.g.*, Target Registration Error (TRE) [37] or Total TRE (TTRE) [38] model) that can precisely compute the error metric, given the spatial distributions of the fiducial localization error. Due to the above reasons, we focus on the probabilistic registration algorithms in both theoretical and experimental sections in this article.

### D. Motivations

In this paper, we concentrate specifically on the rigid PSR problem in CAOS applications. There exist noises (*e.g.*, measuring errors) and outliers in the actual medical applications. With the directional information extracted from the raw PSs, the proposed method can improve registration accuracy and robustness even when the PSs are contaminated by noises and outliers. We also introduce the BCPD backbone architecture to reformulate the 6D generalized PSR problem. The BCPD guarantees convergence from the theoretical level and also demonstrates improved performances.

## III. Notations and Preliminaries

We obey the following notations in this paper:

- $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_N] \in \mathbb{R}^{3\times N}$ - the target positional PS.
- $\hat{\mathbf{X}} = [\hat{\mathbf{x}}_1, \cdots, \hat{\mathbf{x}}_N] \in \mathbb{R}^{3\times N}$ - the corresponding target normal vector set.
- $\mathbf{D_x} = [\mathbf{X}; \hat{\mathbf{X}}] \in \mathbb{R}^{6\times N}$ - the target generalized PS, where the $n$th column vector is represented by $\mathbf{d}_\mathbf{x}^\mathbf{n}$.
- $\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_M] \in \mathbb{R}^{3\times M}$ - the source positional PS.
- $\hat{\mathbf{Y}} = [\hat{\mathbf{y}}_1, \cdots, \hat{\mathbf{y}}_M] \in \mathbb{R}^{3\times M}$ - the corresponding source normal vector set.
- $\mathbf{D_y} = [\mathbf{Y}; \hat{\mathbf{Y}}] \in \mathbb{R}^{6\times N}$ - the source generalized PS, where the $m$th column vector is represented by $\mathbf{d}_\mathbf{y}^\mathbf{m}$.
- $\sigma^2 \in \mathbb{R}$ - isotropic variances of GMMs.
- $\kappa \in \mathbb{R}$ - concentration parameter of FMMs.
- $\alpha = (\alpha_1, \cdots, \alpha_M)^T \in [0,1]^M$ - the membership probabilities which meets $\sum_{m=1}^{M} \alpha_m = 1$, where the element $\alpha_m$ represents mixing proportion of HMM components.
- $c = (c_1, \cdots, c_N)^T \in \{0,1\}^N$ - $c_n = 1$ represents that the $n$th point in $\mathbf{X}$ is a inlier, and takes 0 otherwise.
- $v = (v_1, \cdots, v_N)^T$ - the index symbols that $v_n = m$ means $\mathbf{y}_m$ corresponds to $\mathbf{x}_n$.
- $\Theta$ - the related parameter set $\{\alpha, c, v, \mathbf{R}, \mathbf{t}, \sigma^2, \kappa\}$.

We briefly introduce a basic understanding of the variational Bayesian inference (VBI) framework [20]. VBI aims to estimate the latent variables and parameters, $\mathbf{\Theta}$, given the data, $\mathbf{D}$. Therefore, the posterior $p(\mathbf{\Theta}|\mathbf{D})$ is required. Because of the expensive computational cost of the posterior distribution, it is always unworkable in many actual models. To simplify the computational complexity, we can find a simple distribution $q(\mathbf{\Theta})$ to approximate the real posterior distribution $p(\mathbf{\Theta}|\mathbf{D})$. We first introduce the log-marginal probability:

$$\ln p(\mathbf{D}) = \underbrace{\int q(\mathbf{\Theta}) \ln\left\{\frac{p(\mathbf{D},\mathbf{\Theta})}{q(\mathbf{\Theta})}\right\} \mathrm{d}\mathbf{\Theta}}_{\mathcal{L}(q)} \underbrace{- \int q(\mathbf{\Theta}) \ln\left\{\frac{p(\mathbf{\Theta} \mid \mathbf{D})}{q(\mathbf{\Theta})}\right\} \mathrm{d}\mathbf{\Theta}}_{\mathrm{KL}(q\|p)}$$

Next, we want to minimize the KL divergence to solve the approximate distribution $q(\mathbf{\Theta})$. Given the data, we need to convert the minimization problem into the maximization of the evidence lower bound (ELBO) $\mathcal{L}(q)$. For this optimization problem, $\mathbf{\Theta}$ is decomposed as mutually independent components $\mathbf{\Theta}_i$, which satisfy $q(\mathbf{\Theta}) = \prod_{i=1}^{N} q_i(\mathbf{\Theta}_i)$. Then, we can obtain the general solution of $q_j^\star(\mathbf{\Theta}_j)$:

$$\ln q_j^\star(\mathbf{\Theta}_j) = E_{i\neq j}[\ln p(\mathbf{D},\mathbf{\Theta})] + const, \tag{1}$$

where $E_{i\neq j}[\ln p(\mathbf{D},\mathbf{\Theta})] = \int \ln p(\mathbf{D},\mathbf{\Theta}) \prod_{i(\neq j)}^{N} q_i \mathrm{d}\mathbf{\Theta}_i$. Finally, each $q_j$ can be updated alternately using the coordinate ascent algorithm, which makes $\mathcal{L}(q)$ increase monotonically and guarantees convergence.

## IV. Methods

### A. Problem Formulation

Given two 6D generalized PSs $(\mathbf{D_x}, \mathbf{D_y})$, the positional points and normal vectors in $\mathbf{D_y}$ are regarded as the centroids of the Gaussian Mixture Model (GMM) and the mean directions of the Fisher Mixture Model (FMM) respectively. The positional vectors and normal vectors in generalized target PS $\mathbf{D_x}$ are sampled from GMMs and FMMs respectively. The rigid point set registration (PSR) aims to find the optimal rotation matrix $\mathbf{R} \in SO(3)$ and the translation vector $\mathbf{t}$ for the best alignment between $\mathbf{D_x}$ and $\mathbf{D_y}$. Given $v_n = m$ ($n$th generalized target point is generated from the $m$-th HMM component), the probability density function (PDF) of $\mathbf{d}_\mathbf{x}^\mathbf{n}$ is defined as follows:

$$\begin{aligned} &p\left(\mathbf{d}_\mathbf{x}^\mathbf{n} \mid v_n = m; \mathbf{R}, \mathbf{t}, \sigma^2, \kappa\right) = \\ &\underbrace{\frac{\kappa}{2\pi\left(e^{\kappa} - e^{-\kappa}\right)} e^{\kappa(\mathbf{R}\hat{\mathbf{y}}_m)^\top \hat{\mathbf{x}}_n}}_{\hat{\mathbf{x}}_n \sim \mathcal{F}(\boldsymbol{\mu}_o(\hat{\mathbf{y}}_m, \mathbf{R}), \kappa)} \cdot \underbrace{\frac{1}{(2\pi\sigma^2)^{\frac{3}{2}}} e^{-\frac{1}{2\sigma^2}\|\mathbf{x}_n - (\mathbf{R}\mathbf{y}_m + \mathbf{t})\|^2}}_{\mathbf{x}_n \sim \mathcal{N}\left(\boldsymbol{\mu}_p(\mathbf{y}_m, (\mathbf{R}, \mathbf{t})), \sigma^2\right)} \\ &= \frac{\kappa}{(2\pi\sigma^2)^{\frac{3}{2}} \cdot 2\pi\left(e^{\kappa} - e^{-\kappa}\right)} e^{\kappa(\mathbf{R}\hat{\mathbf{y}}_m)^\top \hat{\mathbf{x}}_n - \frac{1}{2\sigma^2}||\mathbf{x}_n - (\mathbf{R}\mathbf{y}_m + \mathbf{t})\|^2}, \end{aligned} \tag{2}$$

where $\mathcal{F}(\boldsymbol{\mu_o}, \kappa)$ indicates the vMF distribution while $\mathcal{N}(\boldsymbol{\mu_p}, \sigma^2)$ represents the Gaussian distribution. We use the symbol $\varphi_{mn}$ to represent Eq. (2). Different from the setting of CPD, we add a constraint for the outlier distribution: the integral of $p_{out}(x_n)$ over $\mathbb{R}^3$ should be one. In CPD, this integral will be approximately zero when the number of points in target PS becomes larger. Therefore, we assume that $V$ is the minimum volume that includes all points in $\mathbf{D_x}$. The outlier distribution is defined as $p_{out}(\mathbf{x}_n) = 1/V$ to avoid the normalization problem.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2022.3159099, IEEE Transactions on Fuzzy Systems



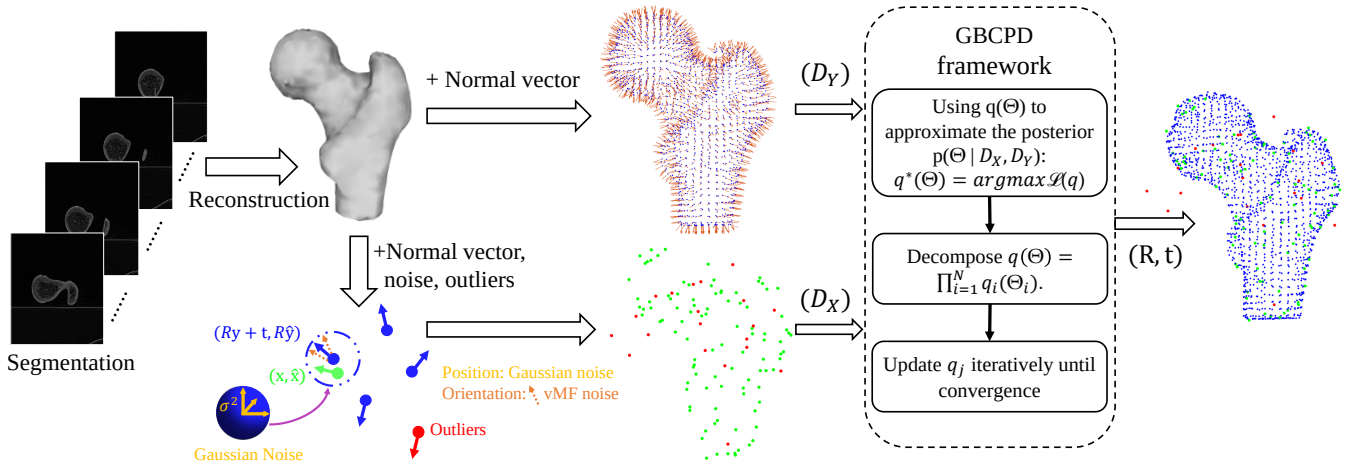


Fig. 1. The GBCPD framework in CAOS. First, the bone (here we use the human femur as an example) is segmented from the medical images. Afterward, the femur model, which can be represented with point sets, is reconstructed from the segmented images. Meanwhile, the normal vectors are calculated from the raw PS. In the target PS, both the positional and normal vectors are usually disturbed by noise and outliers. To model the error distributions with positional and normal vectors, we use the Gaussian and von-mises Fisher distributions. Then the 6D PSs are inputted into the GBCPD. Finally, transformation parameters are obtained after the optimization process.

In addition, to make the point correspondence clearer, we introduce a new definition $\delta_m(v_n)$ where $\delta_m(v_n)$ takes 1 if $v_n = m$, and is 0 otherwise. Then, the joint probability density function of $p(\mathbf{d}_\mathbf{x}^\mathbf{n}, c_n, v_n)$ given $(\mathbf{D_y}, \alpha, \mathbf{R}, \mathbf{t}, \sigma^2, \kappa)$ is defined as:

$$\begin{aligned} &p\left(\mathbf{d}_\mathbf{x}^\mathbf{n}, v_n, c_n \mid \mathbf{D_y}, \alpha, \mathbf{R}, \mathbf{t}, \sigma^2, \kappa\right) \\ &= \left\{\omega p_{\text{out}}\left(\mathbf{x}_n\right)\right\}^{1-c_n} \left\{(1-\omega)\prod_{m=1}^{M}\left(\alpha_m \varphi_{mn}\right)^{\delta_m(v_n)}\right\}^{c_n}, \end{aligned} \tag{3}$$

which includes two cases depending on whether $\mathbf{d}_\mathbf{x}^\mathbf{n}$ is an outlier. $\omega \in [0,1]$ is the outlier probability that $\mathbf{d}_\mathbf{x}^\mathbf{n}$ belongs to outliers. The whole process of our proposed registration framework for CAOS is shown in Fig. 1.

In what follows, we will integrate the HMM into the BCPD framework and solve the optimal transformation by the VBI method.

### B. Variational Bayesian Formulation

Our purpose is to solve the parametric optimization problem. In this article, we adopt the VBI method to obtain a distribution $q(\Theta)$ which can represent the real posterior distribution $p(\Theta|\mathbf{D_x}, \mathbf{D_y})$ approximatively. We first introduce the prior distribution of the parameters in $\Theta$ and then integrate the prior into the HMM to formulate the joint distribution which is used to make the variational processing for the proposed model.

*1) Prior Distribution:*

To control the mixture proportion of HMM components, a Dirichlet distribution is defined as the prior $p(\alpha)$:

$$p(\alpha) = \mathbf{Dir}(\alpha \mid \lambda \mathbf{1}_M) = C(\lambda \mathbf{1}_M) \prod_{m=1}^{M} \alpha_m^{\lambda-1},$$

where $\mathbf{1}_M$ is an $M \times 1$ all-ones vector, and $C(\lambda \mathbf{1}_M)$ can be regarded as a normalization constant. $\lambda$ is the effective prior number. With bigger $\lambda$, comes more influence by the prior rather than by the data. For the simplicity of the variational model, we don't introduce the priors over the source generalized PS $\mathbf{D_y}$ and other parameters in $\Theta$.

*2) Joint Probability Distribution:* Integrating the prior into the HMM, the full joint probability distribution is formulated:

$$p(\mathbf{D_x}, \mathbf{D_y}, \Theta) \propto p(\alpha) \prod_{n=1}^{N} p\left(\mathbf{d}_\mathbf{x}^\mathbf{n}, c_n, v_n \mid \mathbf{D_y}, \alpha, \mathbf{R}, \mathbf{t}, \sigma^2, \kappa\right) \tag{4}$$

### C. Variational Bayesian Approximate Posteriors

In this part, we derive the approximate posterior distributions based on the variational Bayesian formulation mentioned in the previous Section B. Considering the mean field theory and the conditional independence relation between parameters, we define a variational distribution which can be factorized as:

$$q(\Theta) = q_1(\alpha) q_2(c, v) q_3(\mathbf{R}, \mathbf{t}, \sigma^2, \kappa). \tag{5}$$

which can solve a tractable practical solution for our HMM. Then we will update these three factors using the general solution (1).

*1) $q_1(\alpha)$:*

According to the product rule for probabilities and the general result (1), given $q_2(c, v)$ and $q_3(\mathbf{R}, \mathbf{t}, \sigma^2, \kappa)$, we can derive the log of the optimized solution for $q_1^\star(\alpha)$:

$$\ln q_1^\star(\alpha) = E_{q_2, q_3}[\ln p(\mathbf{D_x}, \mathbf{D_y}, \Theta)] + const. \tag{6}$$

Substituting the decomposition (4) into the above equation and absorbing the terms that don't depend on $\alpha$ into the normalization constant, we obtain

$$\begin{aligned} \ln q_1^\star(\alpha) &= \sum_{n=1}^{N}\sum_{m=1}^{M} E_{q_2, q_3}[c_n \delta_m(v_n) \ln(\alpha_m \varphi_{mn})] \\ &\quad + \sum_{m=1}^{M} \ln \alpha_m^{\lambda-1} + const \\ &= \sum_{m=1}^{M} \ln \alpha_m^{\lambda-1+\rho_m} + const, \end{aligned} \tag{7}$$

where $\rho_m = \sum_{n=1}^{N} p_{mn}$. We define $p_{mn}$ as the posterior probability of HMM components which represents the fuzzy correspondence probability between $\mathbf{y}_m$ and $\mathbf{x}_n$, and it is defined as $p_{mn} = E[c_n \delta_m(e_n)] \in [0, 1]$. Taking the exponential of $\ln q_1^\star(\alpha)$, we obtain $q_1^\star(\alpha)$ which also follows a Dirichlet distribution:

$$q_1^\star(\alpha) = \mathbf{Dir}(\alpha \mid \lambda \mathbf{1}_M + \rho) \tag{8}$$

where $\rho = \mathbf{P}\mathbf{1}_N$, $\mathbf{P} = (p_{mn})^{M \times N}$ is the probability matrix.

*2) $q_2(c, v)$:*

Next, we continue to update $q_2(c, v)$ which represents the fuzzy shape correspondences between two PSs. Same as the previous update of $q_1(\alpha)$, the log of the optimized solution for $q_2(c, v)$ is written as follows:

$$\begin{aligned} \ln q_2^\star(c, v) = \sum_{n=1}^{N} \Big[ & \ln \left\{ \omega p_{\text{out}}(\mathbf{x}_n) \right\}^{(1-c_n)} \\ & + \sum_{m=1}^{M} \ln \left\{ (1-\omega) \langle \alpha_m \rangle \langle \varphi_{mn} \rangle \right\}^{c_n \delta_m(v_n)} \Big] + const, \end{aligned} \tag{9}$$

where $\langle \bullet \rangle$ represents the operator $\exp(E[\ln \bullet])$. We have solved $q_1(\alpha)$ following a Dirichlet distribution. According to the standard solutions of the Dirichlet distribution, we can find:

$$\langle \alpha_m \rangle = exp[\psi(\rho_m + \lambda) - \psi(N_{\mathbf{p}} + \lambda M)], \tag{10}$$

where $\psi(\bullet)$ represents the digamma function and $N_{\mathbf{p}} = \sum_{n,m=1}^{N,M} p_{mn}$. In this paper, we focus on the rigid PSR with isotropic positional assumption, so $\langle \varphi_{mn} \rangle$ can be simplified as follows:

$$\langle \varphi_{mn} \rangle = \varphi_{mn}. \tag{11}$$

We can rewrite $q_2^\star(c, v)$ into $q_2^\star(c, v) = \prod_{n=1}^{N} q_2^{\star(n)}(c_n, v_n)$, where

$$\begin{aligned} & q_2^{\star(n)}(c_n, v_n) \propto \\ & \left\{ \omega p_{\text{out}}(\mathbf{x}_n) \right\}^{(1-c_n)} \prod_{m=1}^{M} \left\{ (1-\omega) \langle \alpha_m \rangle \langle \varphi_{mn} \rangle \right\}^{c_n \delta_m(v_n)}. \end{aligned} \tag{12}$$

We define the normalization constant of the $q_2^{\star(n)}$ above as $A = \omega p_{\text{out}}(\mathbf{x}_n) + (1-\omega) \sum_{m=1}^{M} \langle \alpha_m \rangle \langle \varphi_{mn} \rangle$, which includes all pairs of $v_n$ and $c_n$. For each pair of $v_n$ and $c_n$, only one component is available. Then, we derive the closed-form solution as the following:

$$q_2^\star(c, e) = \prod_{n=1}^{N} (1 - \rho_n')^{1-c_n} \left\{ \rho_n' \prod_{m=1}^{M} \left( \frac{p_{mn}}{\rho_n'} \right)^{\delta_m(e_n)} \right\}^{c_n}, \tag{13}$$

where $\rho_n' = \sum_{m=1}^{M} p_{mn}$ represents the posterior probability that $\mathbf{x}_n$ belongs to an inlier. The probability vector $\rho' = (\rho_1', \cdots, \rho_N')$ can be represented as $\mathbf{P}^T \mathbf{1}_M$.

The fuzzy correspondence probability $p_{mn}$ is defined as:

$$p_{mn} = \frac{(1-\omega) \langle \alpha_m \rangle \langle \varphi_{mn} \rangle}{\omega p_{\text{out}}(x_n) + (1-\omega) \sum_{k=1}^{M} \langle \alpha_k \rangle \langle \varphi_{kn} \rangle}. \tag{14}$$

Note that these three representations about $p_{mn}$ are consistent, which means $p_{mn} = q_2^{\star(n)}(c_n = 1, v_n = m) = E[c_n \delta_m(v_n)]$. From Eq. (13), we can easily observe that $q_2^\star(c, v)$ is composed of a Bernoulli distribution (the posterior marginal distribution of $c_n$) and a categorical distribution (the posterior conditional distribution of $v_n$ when $c_n$ is 1). This part about $q_2(c, v)$ guarantees that the update of the posterior probability matrix $\mathbf{P}$ improves the ELBO. Meanwhile, other parameters associated with $p_{mn}$ (*e.g.,* $N_{\mathbf{p}}$, $\rho$ and $\rho'$) are updated.

*3) $q_3(\mathbf{R}, \mathbf{t}, \sigma^2, \kappa)$:*

Different from the VBI method to update $q_1$ and $q_2$, for $q_3$ we update the $\Theta_3 = (\mathbf{R}, \mathbf{t}, \sigma^2, \kappa)$ by directly maximizing the ELBO $\mathcal{L}(q)$. Here, we use the Dirac delta function [39] to represent $q_3$, which ensures that it is characterized by its mode. By maximizing the ELBO over the parameters of $\Theta_3$, we can obtain a point mass which can improve the lower bound. With $q_1$ and $q_2$, the ELBO can be defined as:

$$\mathcal{L}(q) = \int q_3 \cdot E_{q_1, q_2}[\ln p(\mathbf{D}_{\mathbf{x}}, \mathbf{D}_{\mathbf{y}}, \Theta)] d\Theta_3 + const. \tag{15}$$

The derivation can be found in Appendix-C. Moreover, we denote the expected log joint probability distribution by $\mathbf{Q}(\Theta)$. Then this optimization problem [20] is simplified to solve the maximization of $\mathbf{Q}(\Theta)$:

$$\begin{aligned} \mathbf{Q}(\Theta) = & E_{q_1, q_2}[\ln p(\mathbf{D}_{\mathbf{x}}, \mathbf{D}_{\mathbf{y}}, \Theta)] = \\ & - \sum_{n=1}^{N} \sum_{m=1}^{M} p_{mn} \left( \frac{1}{2\sigma^2} (\mathbf{x}_n - \mu_p)^\top (\mathbf{x}_n - \mu_p) \right. \\ & \left. - \kappa ((\mathbf{R}\widehat{\mathbf{y}}_m)^{\mathrm{T}} \widehat{\mathbf{x}}_n) \right) - \frac{3}{2} N_{\mathbf{P}} \log \sigma^2 \\ & - N_{\mathbf{P}} \log \left( e^{\kappa} - e^{-\kappa} \right) + N_{\mathbf{P}} \log \kappa + const. \end{aligned} \tag{16}$$

where $\mu_p = \mathbf{t} + \mathbf{R}\mathbf{y}_{\mathbf{m}}$ and $N_{\mathbf{p}} = \sum_{n,m=1}^{N,M} p_{mn}$. For the maximization $\mathbf{Q}(\Theta)$ with respect to each parameter in $\Theta_3$, we first compute the partial derivative of (16) over $\sigma^2$ and set it to zero. Then the optimized $\sigma^{2\star}$ can be solved as:

$$(\sigma^2)^\star = \frac{\sum_{n=1}^{N} \sum_{m=1}^{M} p_{mn} \left( \left\| \mathbf{x}_n - \mu_p \right\|^2 \right)}{3 N_{\mathbf{p}}} \tag{17}$$

Then, to find the optimal $\mathbf{t}^\star$, we set Eq. (17) as a function of $t$, denoted by $\xi(\mathbf{t})$. Observing Eqs. (16) and (17), we can find that minimization of $\xi(\mathbf{t})$ is to maximize the $\mathbf{Q}(\Theta)$ and $\mathcal{L}(q)$. The minimization problem $\mathbf{t}^\star = \arg\min_t \xi(\mathbf{t})$ can be solved by differentiating $\xi(\mathbf{t})$ with respect to $t$ and setting it to zero. After simple manipulations, $\mathbf{t}^\star$ can be solved as follows:

$$\mathbf{t}^\star = \bar{\mathbf{x}} - \mathbf{R}\bar{\mathbf{y}}, \tag{18}$$

where

$$\begin{aligned} \bar{\mathbf{x}} &= \tfrac{1}{N_{\mathrm{p}}} \textstyle\sum_{n,m=1}^{N,M} p_{mn} \mathbf{x}_n = \tfrac{1}{N_{\mathbf{p}}} \mathbf{X}^{\mathrm{T}} \mathbf{P}^{\mathrm{T}} \mathbf{1}_M, \\ \bar{\mathbf{y}} &= \tfrac{1}{N_{\mathbf{p}}} \textstyle\sum_{n,m=1}^{N,M} p_{mn} \mathbf{y}_m = \tfrac{1}{N_{\mathbf{p}}} \mathbf{Y}^{\mathrm{T}} \mathbf{P} \mathbf{1}_N. \end{aligned}$$

$\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$ represent the expectations of two positional PSs respectively.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2022.3159099, IEEE Transactions on Fuzzy Systems



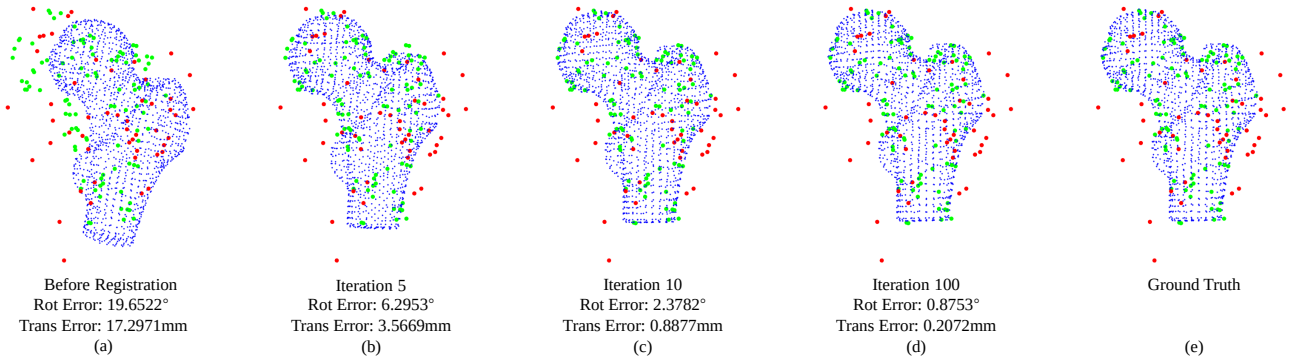


Fig. 2. Registration process using proposed GBCPD with 50% outliers and 1mm/1° noise. (a)~(d): the registration result before registration, 5th, 10th, and 100th iteration respectively; (e): ground truth. The points in the target set X, points in the source set Y, and outliers are represented by green, blue, and red dots respectively.

To solve optimal $\mathbf{R}^\star$ that maximizes the ELBO, we should solve the constrained maximization problem as follows:

$$\begin{aligned}\mathbf{R}^\star &= \arg\max_{\mathbf{R}} \mathbf{Q}(\Theta) \\ &= \arg\max_{\mathbf{R}} -\sum_{n=1}^{N}\sum_{m=1}^{M} p_{mn}\left(\frac{1}{2\sigma^2}\left\|\mathbf{x}_n-(\mathbf{R}\mathbf{y}_m+\mathbf{t})\right\|^2\right. \\ &\left.-\kappa\left(\left(\mathbf{R}\widehat{\mathbf{y}}_m\right)^{\mathrm{T}}\widehat{\mathbf{x}}_n\right)\right),\ \text{s.t. } \mathbf{R}^{\mathrm{T}}\mathbf{R}=\mathbf{I}_3,\ \det(\mathbf{R})=1\end{aligned} \tag{19}$$

To succinctly represent the calculation process of Eq. (19), we introduce these notations:

$$\begin{aligned}&\widetilde{\mathbf{x}}_n=\mathbf{x}_n-\bar{\mathbf{x}}, \widetilde{\mathbf{y}}_m=\mathbf{y}_m-\bar{\mathbf{y}},\\ &\mathbf{H}_o=\kappa\textstyle\sum_{n,m=1}^{N,M} p_{mn}\widehat{\mathbf{y}}_m\widehat{\mathbf{x}}_n^{\mathrm{T}},\\ &\mathbf{H}_p=\frac{1}{\sigma^2}\textstyle\sum_{n,m=1}^{N,M} p_{mn}\widetilde{\mathbf{y}}_m\widetilde{\mathbf{x}}_n^{\mathrm{T}},\\ &\mathbf{H}=\mathbf{H}_o+\mathbf{H}_p.\end{aligned}$$

Then substituting the optimized $\mathbf{t}^\star$ (18) into $\mathbf{R}^\star$ (19), we can rewrite $\mathbf{R}^\star$ as follows:

$$\begin{aligned}\mathbf{R}^\star &= \arg\max_{\mathbf{R}}\left(\frac{1}{\sigma^2}\sum_{n=1}^{N}\sum_{m=1}^{M} p_{mn}\tilde{\mathbf{x}}_n^{\mathrm{T}}\mathbf{R}\tilde{\mathbf{y}}_m\right.\\ &\left.+\kappa\sum_{n=1}^{N}\sum_{m=1}^{M} p_{mn}(\mathbf{R}\widehat{\mathbf{y}}_m)^{\mathrm{T}}\widehat{\mathbf{x}}_n\right)\\ &= \arg\max_{\mathbf{R}}\ \mathrm{Tr}\left(\mathbf{R}(\mathbf{H}_p+\mathbf{H}_o)\right),\end{aligned} \tag{20}$$

where Tr(•) represents the trace of a matrix. Using the lemma proposed in [40], we can obtain the optimal $\mathbf{R}^\star$:

$$\mathbf{R}^\star = \mathbf{V}\mathrm{d}(1,1,\det(\mathbf{V}\mathbf{U}^{\mathrm{T}}))\mathbf{U}^{\mathrm{T}}, \tag{21}$$

where $\mathbf{U}$ and $\mathbf{V}$ are from the factorization of the form $\mathbf{H}=\mathbf{U}\mathbf{H}'\mathbf{V}^{\mathrm{T}}$ after performing the SVD of matrix $\mathbf{H}$, and d(•) represents the diagonal matrix of a vector. Then applying the optimal $\mathbf{R}^\star$ into (17) and (18), we have:

$$\begin{aligned}&\mathbf{t}^\star=\bar{\mathbf{x}}-\mathbf{R}^\star\bar{\mathbf{y}},\\ &(\sigma^2)^\star=\frac{\sum_{n=1}^{N}\sum_{m=1}^{M} p_{mn}\left(\left\|\mathbf{x}_n-(\mathbf{R}^\star\mathbf{y}_m+\mathbf{t}^\star)\right\|^2\right)}{3N_{\mathbf{p}}}.\end{aligned} \tag{22}$$

Finally, we adopt the method proposed in [22] to update $\kappa$. According to the different sources, $\kappa$ is divided into two parts: positional error part and orientational error part. The part caused by the positional error is defined as:

$$err_p=\frac{\sum_{n,m=1}^{NM} p_{mn}\tilde{\mathbf{x}}_n^{\mathrm{T}}\mathbf{R}^\star\tilde{\mathbf{y}}_m}{\sum_{n,m=1}^{NM} p_{mn}\left\|\mathbf{R}^\star\tilde{\mathbf{y}}_m\right\|\left\|\tilde{\mathbf{x}}_n\right\|}. \tag{23}$$

The other part generated from the orientational error is represented as:

$$err_o=\frac{1}{N_{\mathbf{p}}}\sum_{n,m=1}^{NM} p_{mn}\left(\mathbf{R}^\star\widehat{\mathbf{y}}_m\right)^{\mathrm{T}}\widehat{\mathbf{x}}_n. \tag{24}$$

We combine these two parts and obtain $err=\varepsilon err_p+(1-\varepsilon)err_o$ with $\varepsilon=0.5$. Then the optimal $\kappa^\star$ can be updated as follows:

$$\kappa^\star=\frac{err(3-err^2)}{1-err^2}. \tag{25}$$

### D. Implementation Details

According to the closed-form solutions above, $q_1, q_2, q_3$ are updated alternately until convergence. The detailed procedures of our GBCPD method are summarized in Algorithm 1. We first illustrate the initialization of the model parameters. we empirically initialize $\kappa$= 10, and $\omega$ = 0.5. We also initialize $\langle\alpha_m\rangle$ to $1/M$ and $\lambda$ to infinity to keep consistent with BCPD and CPD. One important constraint is that the upper bound $\kappa$ is set to 100 in each iteration to guarantee the computability of $e^\kappa$. We adopt the following criteria as conditions of convergence: 1) the change $\Delta\sigma^2$ between successive iterations is smaller than $10^{-6}$; 2) $\sigma^2$ is below the threshold value $10^{-6}$; 3) the maximum iterations exceed 100 times.

## V. Experiments

Two series of experiments are designed to validate our GBCPD method. As shown in Figs. 1 and 2, the noise-free source PS $\mathbf{Y}$ is sampled from the CT model of the femur. The corresponding unit normal vector set $\hat{\mathbf{Y}}$ is extracted using the PCA technique. The noise-free target generalized PS $\mathbf{D_x}$ is generated by sampling from the source generalized PS $\mathbf{D_y}$. We add noise vectors sampled from the two different noise levels into $\mathbf{D_x}$ in two series of experiments respectively. Moreover, each group further consists of nine cases

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2022.3159099, IEEE Transactions on Fuzzy Systems



TABLE I
ROTATION AND TRANSLATION ERRORS UNDER THE LOW NOISE LEVEL. DIFFERENT RATES OF OUTLIERS ARE INJECTED INTO $\mathbf{D_x}$.

| Error Type | Method | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% |
|---|---|---|---|---|---|---|---|---|---|---|
| Rotation (°) | ICP [13] | 2.6877 | 3.6509 | 3.9594 | 4.5431 | 4.8543 | 5.2485 | 5.3575 | 4.8144 | 5.2281 |
| | CPD [28] | 1.5168 | 1.1019 | 1.1783 | 1.4665 | 1.8922 | 2.8537 | 3.9650 | 4.6353 | 5.3339 |
| | BCPD [21] | 1.6861 | 1.8341 | 1.0474 | 2.5165 | 2.8246 | 2.9246 | 3.4831 | 3.2950 | 3.4021 |
| | Proposed Method | **0.9530** | **1.0261** | **1.0353** | **1.0143** | **1.1096** | **1.0123** | **0.9228** | **1.0080** | **1.1005** |
| Translation (mm) | ICP [13] | 1.0479 | 1.4309 | 1.6958 | 1.7979 | 1.8395 | 2.0176 | 2.1533 | 2.1878 | 2.0304 |
| | CPD [28] | 3.3426 | 3.4244 | 3.3414 | 3.3038 | 3.5482 | 3.7009 | 3.8305 | 4.2901 | 4.0380 |
| | BCPD [21] | 3.2149 | 3.1650 | 3.1210 | 2.9859 | 3.3414 | 3.1810 | 3.2779 | 3.4989 | 3.4497 |
| | Proposed Method | **0.4804** | **0.5800** | **0.5228** | **0.5771** | **0.5204** | **0.5696** | **0.5633** | **0.5968** | **0.4781** |

TABLE II
ROTATION AND TRANSLATION ERRORS UNDER THE HIGH NOISE LEVEL. DIFFERENT RATES OF OUTLIERS ARE INJECTED INTO $\mathbf{D_x}$.

| Error Type | Method | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% |
|---|---|---|---|---|---|---|---|---|---|---|
| Rotation (°) | ICP [13] | 3.5297 | 4.2803 | 4.5789 | 5.0979 | 5.3169 | 4.9937 | 5.2821 | 5.9766 | 5.1078 |
| | CPD [28] | 2.3523 | 2.2863 | 2.6430 | 3.1757 | 2.7845 | 3.9739 | 4.7553 | 5.2052 | 5.1097 |
| | BCPD [21] | 2.5191 | 2.3988 | 2.7055 | 3.2340 | 3.0436 | 3.4137 | 3.5485 | 3.7751 | 3.7763 |
| | Proposed Method | **2.2005** | **2.2277** | **2.5437** | **2.3418** | **2.2074** | **2.3365** | **2.1846** | **2.4130** | **2.8599** |
| Translation (mm) | ICP [13] | 1.3525 | 1.7002 | 1.9964 | 1.9823 | 2.1776 | 2.0680 | 2.0100 | 2.1745 | 2.0908 |
| | CPD [28] | 3.0710 | 3.3535 | 3.4965 | 3.4642 | 3.5874 | 3.7885 | 4.1659 | 4.4055 | 3.8958 |
| | BCPD [21] | 3.1502 | 3.3460 | 3.2847 | 3.3272 | 3.1086 | 3.2112 | 3.3821 | 3.3514 | 3.1342 |
| | Proposed Method | **1.1780** | **1.1585** | **1.1992** | **1.2345** | **1.1735** | **1.1423** | **1.2638** | **1.3381** | **1.2843** |

**Algorithm 1** Robust Generalized Bayesian Coherent Point Drift

1: **Initialization**: $\mathbf{R} = \mathbf{I}_3$, $\mathbf{t} = \mathbf{0}_{3\times 1}$, $\omega$, $\kappa$, $\lambda$, $\sigma^2 = \frac{1}{3MN}\sum_{n,m=1}^{NM} \|\mathbf{x}_n - \mathbf{y}_m\|^2$, $\langle \alpha_m \rangle = \frac{1}{M}$.
2: **repeat**
3: - Update $\langle \alpha_m \rangle$, $\langle \varphi_{mn} \rangle$ and $\mathbf{P} = (p_{mn})^{M\times N}$ by (10), (11) and (14) respectively
4: - Update the transformation $\mathbf{R}$ and $\mathbf{t}$ by (21) and (22)
5: - Update HMM's parameters $\sigma^2$ and $\kappa$ by (22) and (25)
6: **until** Convergence.
7: **return** $\mathbf{R}^\star$ and $\mathbf{t}^\star$.

where the different ratios of outliers are injected into $\mathbf{D_x}$ to generate the *disturbed* $\mathbf{D_x}$. We conduct 100 registration trials in each case and set the number of inliers in $\mathbf{D_x}$ to 100. The ground truth of the rigid transformation $[\mathbf{R_{true}}, \mathbf{t_{true}}]$ in each case is randomly sampled from the range $[10°, 25°]$ and $[10\text{mm}, 25\text{mm}]$ respectively. Note that our generalized BCPD method is intended to solve the local fine registration problem in computer-assisted orthopedic surgery (CAOS). In the general procedure of CAOS, the coarse registration that utilizes the biological anatomical landmarks will be finished before the fine registration. Therefore, the misalignment is set in a small range. Then the final *misaligned* source generalized PS $\mathbf{D_y}$ is generated by performing the real $[\mathbf{R_{true}}, \mathbf{t_{true}}]$ to $(\mathbf{Y}, \hat{\mathbf{Y}})$. In each registration trial, the *misaligned* source PS $\mathbf{D_y}$ and *disturbed* target PS $\mathbf{D_x}$ (with further noise injected) are registered together.

The rotation error and translation error are chosen as registration metrics: $\theta_{error} = \arccos\left[\left(\text{tr}(\mathbf{R_{true}}\mathbf{R}_{cal}^\top) - 1\right)/2\right]$, and $\mathbf{t}_{error} = \|\mathbf{t}_{cal} - \mathbf{t_{true}}\|_2$, where $\mathbf{R}_{cal}$ and $\mathbf{t}_{cal}$ are the estimated transformation between *disturbed* $\mathbf{D_x}$ and *misaligned* $\mathbf{D_y}$. We calculate the mean error values of 100 registration trials in each case. The performance of our GBCPD method is evaluated by comparing it with several SOTA registration algorithms: ICP [13], CPD [28], and BCPD [21].

### A. Low Noise Level

In the first group of experiments, the final *disturbed* $\mathbf{D_x}$ are further processed by two steps: 1) adding the noise vectors sampled from the zero-mean Gaussian (positional) noise with 1mm standard deviations (std) and vMF (orientation) noise with 1° std to PSs $\mathbf{X}$ and $\hat{\mathbf{X}}$ respectively, where $\kappa = 3200$ corresponds to 1° std [19]; 2) injecting outliers with different rates from 10% to 90% with an interval of 10%. As we have mentioned, the number of inliers is set to 100 in each case. The outliers are produced by applying displacement vectors uniformly sampled from $[20\text{mm}, 30\text{mm}]$ to points randomly sampled from the CT model PS. Then the final *disturbed* $\mathbf{D_x}$ is obtained with $N = 110$ to 190.

### B. High Noise Level

Like the setting of the low-level noise group, $\kappa$ is equated to 800 to obtain 2° std noise. Then both positional and orientational noises sampled from 2mm/2° stds are injected into $\mathbf{D_x}$. After outliers are also injected, we obtain the second group of *disturbed* $\mathbf{D_x}$.

## VI. RESULTS AND DISCUSSION

### A. Two Noise Levels

*1) Quantitative Results:* The mean values of rotation and translation errors with different ratios of outliers under low-level and high-level noise are summarized in Tables I and II respectively. In the first group, the low-level noise vectors are injected into the target PS $\mathbf{D_x}$. Table I indicates that in ICP [13], CPD [28], and BCPD [21] methods, both rotation and translation error values increase as the rate of outliers increases. In contrast, the performance of our GBCPD method is still stable with different outliers. In each case, our proposed

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2022.3159099, IEEE Transactions on Fuzzy Systems



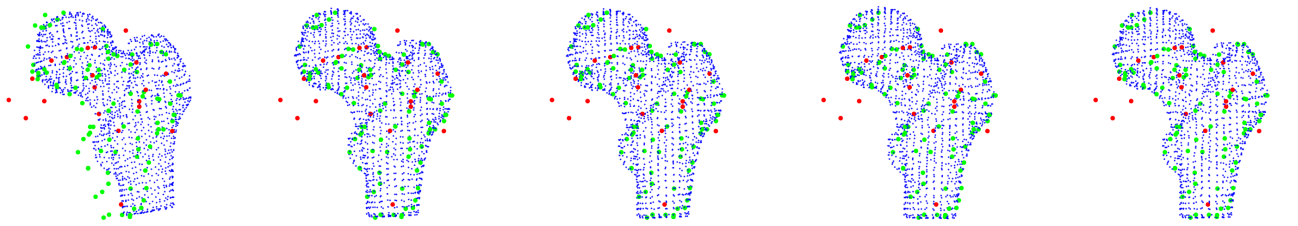


Fig. 3. Subfigure (a) represents the result before the registration. Subfigures (b)~(d) represent three registration results using ICP [13], CPD [28], BCPD [21], and our method respectively. 1mm/1° noise and 20% outliers are injected into $\mathbf{D_x}$. The points in the source PS, points in the target PS, and outliers are represented by blue, green, and red dots respectively.

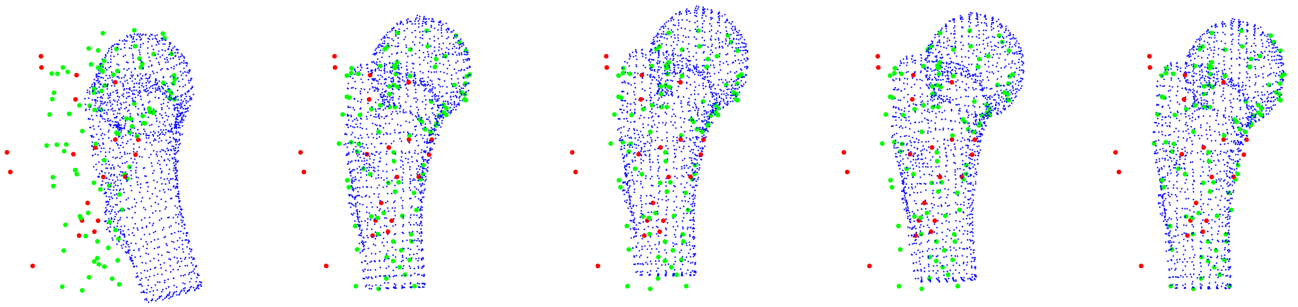


Fig. 4. Subfigure (a) represents the result before the registration. Subfigures (b)~(d) represent three registration results using ICP, CPD, BCPD, and our method respectively. 1mm/1° noise and 90% outliers are injected into $\mathbf{D_x}$. The points in the source PS $\mathbf{D_y}$, points in the target PS $\mathbf{D_x}$ and outliers are represented by blue, green, and red dots respectively.

method outperforms ICP, CPD, and BCPD, especially in the cases with larger percentages of outliers. With 90% outliers injected, the difference of rotation error is up to about 4° compared with ICP and CPD. Meanwhile, the difference of translation error is more than 3mm compared with CPD and BCPD. Our method achieves the smallest rotation and translation errors, where the mean values of nine cases are 1.0202° and 0.5432mm respectively.

In the second group, the high-level noise is injected into $\mathbf{D_x}$. Compared with Table I, Table II shows that all methods perform worse with high-level noise, especially for rotation results. However, when the outliers increase, our GBCPD method is still stable like in Table I and performs best among these four algorithms, which demonstrates the accuracy and robustness of the GBCPD method.

Through these two sets of experiments, we can also find that CPD and BCPD perform badly on the estimation of the translation vector, even worse than ICP. By utilizing the normal vectors in the registration process, the performance of estimating the translation vector is significantly improved, which greatly exceeds the accuracy of the ICP algorithm.

We further use the ***ttest*** function in MATLAB to compute the $p$-value of our algorithm compared with ICP, CPD, and BCPD respectively. Almost all $p$-values are not more than 0.05, which represents a 0.05 significance level. It indicates that the results in our experiments are statistically significant.

*2) Qualitative Results:* Figs. 3, 4, 5, and 6 show the qualitative registration results using ICP, CPD, BCPD, and our methods respectively. For each noise level, we choose 20% and 90% outliers to show the results. As it can be seen from these figures, the green points which represent the intra-operative inlier points are well registered to their corresponding blue points after registration using our method even with a large percentage of outliers (90% outliers) in the target set $\mathbf{D_x}$.

In Fig. 4, a sagittal view is adopted to obtain a more intuitive display. The other three figures are observed from the coronal view. We can find that the condition of noise and the number of outliers have a great influence on the registration performances. Compared results under low-level noise (Fig. 3 and Fig. 4) with high-level noise (Fig. 5 and Fig. 6), high-level noise leads to worse registration performances for all methods, while our methods are still in a small range. Compared Fig. 3 with Fig. 5, we can clearly observe that two PSs(denoted with green and blue dots respectively) are successfully aligned using our method with a large percentage of outliers.

### B. Convergence Speed

Fig. 2(a)-(d) shows the process of PSR in 0th (before registration), 5th, 10th, and 100th iteration using the GBCPD method. Fig. 2(e) presents the registration result with the ground truth $[\mathbf{R_{true}}, \mathbf{t_{true}}]$. We can see that the performance in the 10th iteration is pretty close to the performance in

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2022.3159099, IEEE Transactions on Fuzzy Systems




Fig. 5. Subfigure (a) represents the result before the registration. Subfigures (b)~(d) represent three registration results using ICP, CPD, BCPD, and our method respectively. 2mm/2° noise and 20% outliers are injected into $\mathbf{D_x}$. The points in the source PS, points in the target PS, and outliers are represented by blue, green, and red dots respectively.

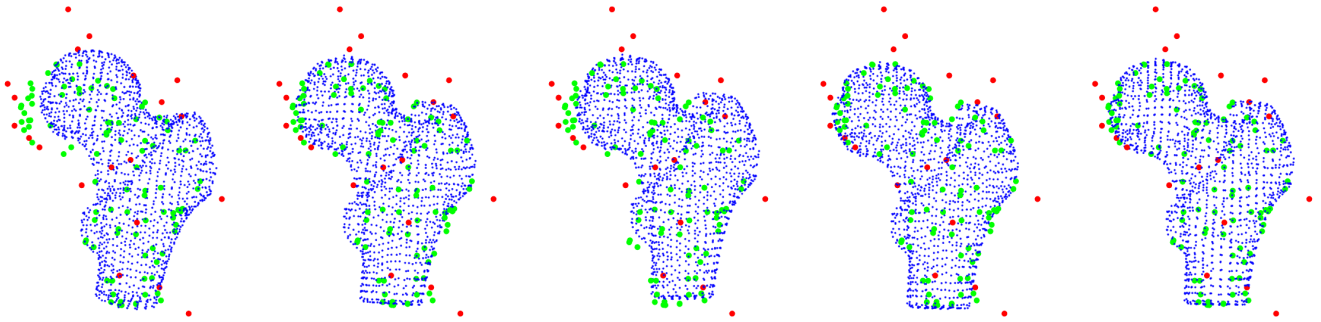


Fig. 6. Subfigure (a) represents the result before the registration. Subfigures (b)~(d) represent three registration results using ICP, CPD, BCPD, and our method respectively. 2mm/2° noise and 90% outliers are injected into $\mathbf{D_x}$. The points in the source PS $\mathbf{D_y}$, points in the target PS $\mathbf{D_x}$, and outliers are represented by blue, green, and red dots respectively.

the 100th iteration. Furthermore, we choose one case which is injected 1mm/1° noise and 50% outliers to evaluate the convergence speed of four algorithms.

Fig. 7 indicates the convergence speeds of the rotation error and translation error respectively. Compared with the other three methods, our GBCPD method achieves the smallest error with the fastest convergence speed. BCPD also converges fast at first, but after about the 10th iteration it becomes slower and the final results are worse in terms of both rotation and translation errors. The convergence speed of CPD is similar to that of BCPD, and its result is only a little lower than that of BCPD. ICP has a slower convergence speed and the biggest errors for the rotation error. For the translation error, there are fluctuations after about the 10th iteration and the final translation error is bigger than ours and smaller than that of CPD and BCPD.

### C. Robustness to Parameter $\omega$

The parameter $\omega$ denotes the weight of $p_{out}(\mathbf{x}_n)$ which also means the prior that $\mathbf{x}_n$ is an outlier. Five different $\omega$ from 0.1 to 0.9 with an interval of 0.2 are chosen to evaluate the effect of $\omega$ on the PSR results. We also perform 100 trials in each case. Fig. 8 shows the rotation errors and translation errors with 1mm/1° noise and 50% outliers injected into the target PS $\mathbf{D_x}$. We find that $\omega$ has a small influence on the final registration results under different values. Finally, we choose $\omega = 0.5$ in our experiments.

### D. Discussion

To summarize, our proposed GBCPD method exceeds the SOTA methods in the following aspects: 1) registration accuracy; 2) robustness to noise and outliers; 3) convergence speed.

This paper generalizes the recently proposed Bayesian Coherent Point Drift (BCPD) method to the six-dimensional scenario. The posterior distribution of the model parameters is factorized into the product of three components: (1) that of the membership probability of both inliers' and outliers' distributions; (2) that of the fuzzy correspondence variables; (3) that of the rotational matrix, translational vector, the variance with the positional localization error, and the concentration parameter with the estimation of the normal vectors. With the variational Bayesian setting, the convergence of the algorithm is guaranteed in the theoretical aspect. In terms of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2022.3159099, IEEE Transactions on Fuzzy Systems



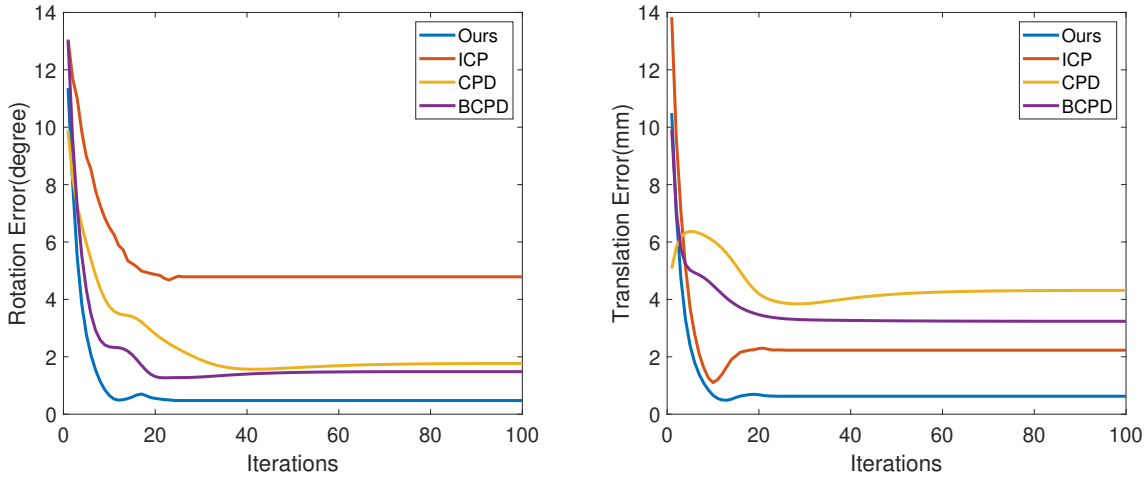


Fig. 7. Convergence speed of four methods with respect to the iterations. 1mm/1° noise and 50% outliers are injected into the generalized target PS $\mathbf{D_x}$. Left: rotation error. Right: translation error.

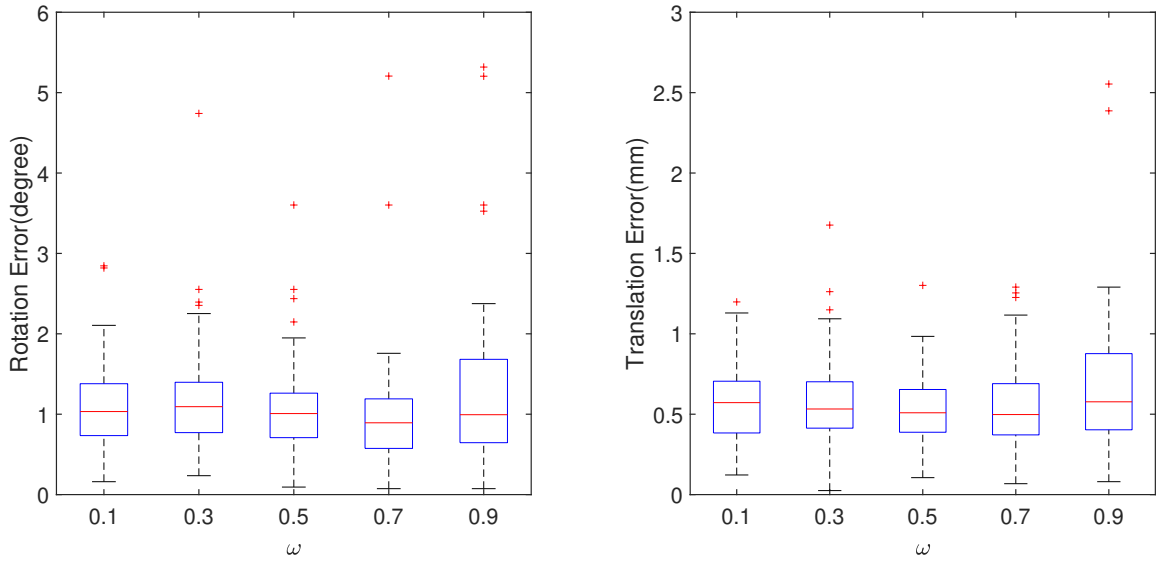


Fig. 8. Boxplots of rotation errors (left) and translation errors (right). 1mm/1° noise and 50% outliers are injected into the generalized target PS $\mathbf{D_x}$.

the performances, compared to the BCPD, both the rotation and translation error values are reduced with the additional normal vectors. More importantly, the proposed method is quite robust to increasing rates of outliers.

We look forward to inspiring more advanced PSR methods that utilize orientational features. Three main limitations exist in the proposed GBCPD method: (1) GBCPD is still to solve the local rigid PSR which means the coarse registration should be performed before using our GBCPD. We will develop the global method to get rid of the dependence on anatomical landmarks in the coarse registration. (2) The features of local structures that can help to estimate the correspondences [41] are not considered in our GBCPD. These local structures can be lightly incorporated into our GBCPD. (3) The positional error distribution is assumed to be isotropic. However, anisotropic uncertainty is very common in biomedical applications. In the future, we will introduce the anisotropic positional error to satisfy the actual situation.

## VII. Conclusions

In this paper, we propose a robust variational bayesian inference based registration method under the hybrid mixture model (HMM) framework. In addition to the positional information in the raw point sets, the normal vectors generated from the raw point sets are also used in the registration process. Our method demonstrates a better performance on robustness and accuracy compared with the SOTA registration approaches. At the same time, the proposed GBCPD method guarantees theoretical convergence. The results demonstrate great clinical potentials in computer-assisted surgery (CAS) and other point set registration (PSR) applications.

The work can be further improved in the following aspects.

Firstly, the prior distributions over $\sigma$, $\kappa$, or the variables $\mathbf{R}, \mathbf{t}$ can be introduced to make the generative model more accurate, such as the Gaussian-Wishart distribution and the Gaussian-Gamma distribution [20]. Secondly, we explore several ways to accelerate the complex model [42]. Thirdly, we will further consider the anisotropic cases in both rotation and translation errors to reduce related errors motivated by practical computer-assisted surgery (CAS) applications. The anisotropic uncertainty is very common in biomedical applications. For example, there are larger standard deviations in different directions of an optical tracking system (OTS) used for surgical instrument tracking [19].

## APPENDIX

### A. Introduction to vMF Distribution

The vMF distribution is widely used to model the directional features in high-dimensional data [22]. In general, the mean and concentration parameters ($\boldsymbol{\mu}_o \; and \; \boldsymbol{\kappa}$) of FMM should be estimated. Based on the vMF distribution, the 3-dimensional normal vectors $\widehat{\mathbf{x}}_n$ are distributed on the unit hypersphere. Its PDF is formulated as:

$$p(\widehat{\mathbf{x}}_n \mid \boldsymbol{\mu}_o, \boldsymbol{\kappa}) = \frac{\kappa}{2\pi \left(e^{\kappa} - e^{-\kappa}\right)} e^{\kappa \boldsymbol{\mu}_o^{\top} \widehat{\mathbf{x}}_n} \tag{26}$$

where $||\widehat{\mathbf{x}}_n|| = 1$, and $\boldsymbol{\mu}_o$ is the mean direction.

### B. Derivation of Eq. (9)

Given $q_1^{\star}(\alpha)$ and $q_3^{\star}(\mathbf{R}, \mathbf{t}, \sigma^2, \kappa)$, the closed-form solution of $q_2^{\star}(c, v)$ can be obtained using the general result of VBI in Eq. (1) as follows:

$$\ln q_2^{\star}(c, v) = E_{q_1, q_3}[\ln p(\mathbf{D}_{\mathbf{x}}, \mathbf{D}_{\mathbf{y}}, \Theta)] + const. \tag{27}$$

Then, substituting the full joint probability density function Eq. (4) into this equation, we have:

$$\begin{aligned} Eq.(27) = E_{q_1, q_3}\Bigg[ \ln \Bigg( & p(\alpha) \cdot \\ & \prod_{n=1}^{N} p\left(\mathbf{d}_{\mathbf{x}}^{\mathbf{n}}, c_n, v_n \mid \mathbf{D}_{\mathbf{y}}, \alpha, \mathbf{R}, \mathbf{t}, \sigma^2, \kappa\right) \Bigg) \Bigg] + const. \end{aligned} \tag{28}$$

Substitute the joint probability density function Eq. (3) and remove the terms that are independent of $(c, v)$ into the normalization constant as follows,

$$\begin{aligned} & Eq.(28) \\ & = E_{q_1, q_3}\Bigg[ \ln \Bigg( p(\alpha) \prod_{n=1}^{N} \left\{\omega p_{\text{out}}\left(\mathbf{x}_n\right)\right\}^{1-c_n} \\ & \quad \left\{(1-\omega) \prod_{m=1}^{M} \left(\alpha_m \varphi_{mn}\right)^{\delta_m(v_n)} \right\}^{c_n} \Bigg) \Bigg] + const. \\ & = E_{q_1, q_3}\Bigg[ \sum_{n=1}^{N} \left[ \ln \left\{\omega p_{\text{out}}\left(\mathbf{x}_n\right)\right\}^{(1-c_n)} \right] + \sum_{n=1}^{N} \ln(1-\omega)^{(c_n)} \\ & \quad + \sum_{n=1}^{N} c_n \sum_{m=1}^{M} \delta_m\left(v_n\right) \ln\left(\alpha_m \varphi_{mn}\right) \Bigg] + const. \\ & = \sum_{n=1}^{N} \Bigg\{ \ln \left\{\omega p_{\text{out}}\left(\mathbf{x}_n\right)\right\}^{(1-c_n)} + c_n \ln(1-\omega) \\ & \quad + \sum_{m=1}^{M} c_n \delta_m\left(v_n\right) E_{q_1, q_3}\left[\ln\left(\alpha_m \varphi_{mn}\right)\right] \Bigg\} + const. \end{aligned} \tag{29}$$

By definition, we have $c_n = \sum_{m=1}^{M} c_n \delta_m\left(v_n\right)$, and denote $\langle \alpha_m \rangle = \exp(E[\ln \alpha_m])$ and $\langle \varphi_{mn} \rangle = \exp(E[\ln \varphi_{mn}])$. Then, the Eq. (29) can be written as follows:

$$\begin{aligned} \ln q_2^{\star}(c, v) = \sum_{n=1}^{N} \Bigg[ & \ln \left\{\omega p_{\text{out}}\left(\mathbf{x}_n\right)\right\}^{(1-c_n)} \\ & + \sum_{m=1}^{M} \ln \left\{(1-\omega) \left\langle \alpha_m \right\rangle \left\langle \varphi_{mn} \right\rangle\right\}^{c_n \delta_m(v_n)} \Bigg] + const, \end{aligned} \tag{30}$$

which is the equation in Eq. (9). Then, following the instructions in the section IV.C.2, the final closed-form solution of $q_2^{\star}(c, v)$ can be obtained.

### C. Derivation of Eq. (15)

We denote $\Theta_{12} = (\alpha, c, v), \Theta_3 = (\mathbf{R}, \mathbf{t}, \sigma^2, \kappa)$ in this part. Given $q(\Theta_{12})$, we have:

$$\begin{aligned} \mathcal{L}(q) &= \int q(\Theta) \cdot \ln \frac{p(D_x, D_y, \Theta)}{q(\Theta)} d\Theta \\ &= \int q\left(\Theta_{12}\Theta_3\right) \cdot \ln \frac{p\left(D_x, D_y, \Theta_{12}, \Theta_3\right)}{q\left(\Theta_{12}\right) \cdot q\left(\Theta_3\right)} d\Theta_{12} d\Theta_3 \\ &= \int q\left(\Theta_{12}\Theta_3\right) \ln \frac{p\left(D_x, D_y, \Theta_{12}, \Theta_3\right)}{q\left(\Theta_{12}\right)} d\Theta_{12} d\Theta_3 \\ &\quad - \int q\left(\Theta_{12}\right) q(\Theta_3) \ln q(\Theta_3) d\Theta_{12} d\Theta_3 \\ &= \int q\left(\Theta_3\right) E_{q(\Theta_{12})}[\ln p\left(D_x, D_y, \Theta_{12}, \Theta_3\right)] d\Theta_3 \\ &\quad - E_{q(\Theta_{12})}\left[\ln q\left(\Theta_{12}\right)\right] - \int q(\Theta_3) \ln q(\Theta_3) d\Theta_3 \\ &= \int q\left(\Theta_3\right) E_{q(\Theta_{12})}[\ln p\left(D_x, D_y, \Theta_{12}, \Theta_3\right)] d\Theta_3 \\ &\quad - \int q(\Theta_3) \ln q(\Theta_3) d\Theta_3 + const, \end{aligned} \tag{31}$$

Because $q_3\left(\Theta_3\right)$ is a Dirac delta function, we can remove the entropy term $-\int q(\Theta_3) \ln q(\Theta_3) d\Theta_3$. Then Eq. (15) is derived.

## REFERENCES


[1] A. Fan, J. Ma, X. Jiang, and H. Ling, "Efficient deterministic search with robust loss functions for geometric model fitting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[2] A. L. Fuhrmann, R. Splechtna, and J. Přikryl, "Comprehensive calibration and registration procedures for augmented reality," in *Immersive Projection Technology and Virtual Environments 2001*. Springer, 2001, pp. 219–227.

[3] L. Qian, J. Y. Wu, S. P. DiMaio, N. Navab, and P. Kazanzides, "A review of augmented reality in robotic-assisted surgery," *IEEE Transactions on Medical Robotics and Bionics*, vol. 2, no. 1, pp. 1–16, 2019.

[4] K. Van Wyk and J. A. Marvel, "Strategies for improving and evaluating robot registration performance," *IEEE Transactions on Automation Science and Engineering*, vol. 15, no. 1, pp. 320–328, 2017.

[5] K. Wu, Z. J. Daruwalla, K. L. Wong, D. Murphy, and H. Ren, "Development and selection of asian-specific humeral implants based on statistical atlas: toward planning minimally invasive surgery," *International journal of computer assisted radiology and surgery*, vol. 10, no. 8, pp. 1333–1345, 2015.

[6] Z. Yaniv, "Registration for orthopaedic interventions," in *Computational radiology for orthopaedic interventions*. Springer, 2016, pp. 41–70.

[7] R. Horaud, F. Forbes, M. Yguel, G. Dewaele, and J. Zhang, "Rigid and articulated point registration with expectation conditional maximization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 3, pp. 587–602, 2010.

[8] L. Zhou, S. Wang, and M. Kaess, "A fast and accurate solution for pose estimation from 3d correspondences," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 1308–1314.

[9] J. Ma, W. Qiu, J. Zhao, Y. Ma, A. L. Yuille, and Z. Tu, "Robust $l\_\{2\}e$ estimation of transformation for non-rigid registration," *IEEE Transactions on Signal Processing*, vol. 63, no. 5, pp. 1115–1129, 2015.

[10] J. Ma, X. Jiang, A. Fan, J. Jiang, and J. Yan, "Image matching from handcrafted to deep features: A survey," *International Journal of Computer Vision*, vol. 129, no. 1, pp. 23–79, 2021.

[11] X. Jiang, J. Ma, J. Jiang, and X. Guo, "Robust feature matching using spatial clustering with heavy outliers," *IEEE Transactions on Image Processing*, vol. 29, pp. 736–746, 2019.

[12] A. Fan, X. Jiang, Y. Ma, X. Mei, and J. Ma, "Smoothness-driven consensus based on compact representation for robust feature matching," *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[13] P. J. Besl and N. D. McKay, "Method for registration of 3-d shapes," in *Sensor fusion IV: control paradigms and data structures*, vol. 1611. International Society for Optics and Photonics, 1992, pp. 586–606.

[14] B. Maiseli, Y. Gu, and H. Gao, "Recent developments and trends in point set registration methods," *Journal of Visual Communication and Image Representation*, vol. 46, pp. 95–106, 2017.

[15] F.-L. Chung, Z. Deng, and S. Wang, "An adaptive fuzzy-inference-rule-based flexible model for automatic elastic image registration," *IEEE Transactions on Fuzzy Systems*, vol. 17, no. 5, pp. 995–1010, 2009.

[16] Q. He, J. Zhou, S. Xu, Y. Yang, R. Yu, and Y. Liu, "Adaptive hierarchical probabilistic model using structured variational inference for point set registration," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 11, pp. 2784–2798, 2020.

[17] G. Klir and B. Yuan, *Fuzzy sets and fuzzy logic*. Prentice hall New Jersey, 1995, vol. 4.

[18] C. Raposo and J. P. Barreto, "Using 2 point+ normal sets for fast registration of point clouds with small overlap," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 5652–5658.

[19] S. Billings and R. Taylor, "Generalized iterative most likely oriented-point (g-imlop) registration," *International journal of computer assisted radiology and surgery*, vol. 10, no. 8, pp. 1213–1226, 2015.

[20] C. M. Bishop, *Pattern recognition and machine learning*. springer, 2006.

[21] O. Hirose, "A bayesian formulation of coherent point drift," *IEEE transactions on pattern analysis and machine intelligence*, 2020.

[22] A. Banerjee, I. S. Dhillon, J. Ghosh, S. Sra, and G. Ridgeway, "Clustering on the unit hypersphere using von mises-fisher distributions." *Journal of Machine Learning Research*, vol. 6, no. 9, 2005.

[23] J. Hermans, D. Smeets, D. Vandermeulen, and P. Suetens, "Robust point set registration using em-icp with information-theoretically optimal outlier handling," in *CVPR 2011*. IEEE, 2011, pp. 2465–2472.

[24] S. Billings and R. Taylor, "Iterative most likely oriented point registration," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2014, pp. 178–185.

[25] J. Yang, H. Li, D. Campbell, and Y. Jia, "Go-icp: A globally optimal solution to 3d icp point-set registration," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 11, pp. 2241–2254, 2015.

[26] C. Olsson, F. Kahl, and M. Oskarsson, "Branch-and-bound methods for euclidean registration problems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 5, pp. 783–794, 2008.

[27] A. L. Pavlov, G. W. Ovchinnikov, D. Y. Derbyshev, D. Tsetserukou, and I. V. Oseledets, "Aa-icp: Iterative closest point with anderson acceleration," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 3407–3412.

[28] A. Myronenko and X. Song, "Point set registration: Coherent point drift," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 12, pp. 2262–2275, 2010.

[29] A. L. Yuille and N. M. Grzywacz, "A mathematical analysis of the motion coherence theory," *International Journal of Computer Vision*, vol. 3, no. 2, pp. 155–175, 1989.

[30] G. D. Evangelidis and R. Horaud, "Joint alignment of multiple point sets with batch and incremental expectation-maximization," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 6, pp. 1397–1410, 2017.

[31] B. Jian and B. C. Vemuri, "Robust point set registration using gaussian mixture models," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 8, pp. 1633–1645, 2010.

[32] Y. Li, R. Bu, M. Sun, W. Wu, X. Di, and B. Chen, "Pointcnn: Convolution on $\chi$-transformed points," in *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, 2018, pp. 828–838.

[33] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, "Dynamic graph cnn for learning on point clouds," *Acm Transactions On Graphics (tog)*, vol. 38, no. 5, pp. 1–12, 2019.

[34] Y. Wang and J. Solomon, "Prnet: Self-supervised learning for partial-to-partial registration," *NeurIPS*, 2019.

[35] X. Bai, Z. Luo, L. Zhou, H. Chen, L. Li, Z. Hu, H. Fu, and C.-L. Tai, "Pointdsc: Robust point cloud registration using deep spatial consistency," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 15 859–15 869.

[36] H. Yang, J. Shi, and L. Carlone, "Teaser: Fast and certifiable point cloud registration," *IEEE Transactions on Robotics*, 2020.

[37] Z. Min and M. Q.-H. Meng, "General first-order target registration error model considering a coordinate reference frame in an image-guided surgical system," *Medical & Biological Engineering & Computing*, vol. 58, no. 12, pp. 2989–3002, 2020.

[38] Z. Min, H. Ren, and M. Q.-H. Meng, "Statistical model of total target registration error in image-guided surgery," *IEEE Transactions on Automation Science and Engineering*, vol. 17, no. 1, pp. 151–165, 2019.

[39] J. Bernardo, M. Bayarri, J. Berger, A. Dawid, D. Heckerman, A. Smith, M. West *et al.*, "The variational bayesian em algorithm for incomplete data: with application to scoring graphical model structures," *Bayesian statistics*, vol. 7, no. 453-464, p. 210, 2003.

[40] A. Myronenko and X. Song, "On the closed-form solution of the rotation matrix arising in computer vision problems," *arXiv preprint arXiv:0904.1613*, 2009.

[41] J. Ma, J. Zhao, and A. L. Yuille, "Non-rigid point set registration by preserving global and local structures," *IEEE Transactions on image Processing*, vol. 25, no. 1, pp. 53–64, 2015.

[42] O. Hirose, "Acceleration of non-rigid point set registration with down-sampling and gaussian process regression," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 8, pp. 2858–2865, 2020.