Robust and Accurate Point Set Registration with Generalized Bayesian Coherent Point Drift

Ang Zhang[1,†], Zhe Min[2,†], Jin Pan[1], and Max Q.-H. Meng[3,§], *Fellow, IEEE*

***Abstract*— Point set registration (PSR) is an essential problem in surgical navigation and image-guided surgery (IGS). It can help align the pre-operative volumetric images with the intra-operative surgical space. The performances of PSR are susceptible to noise and outliers, which are the cases in real-world surgical scenarios. In this paper, we provide a novel point set registration method that utilizes the features extracted from the PSs and can guarantee the convergence of the algorithm simultaneously. More specifically, we formulate the PSR with normal vectors by generalizing the bayesian coherent point drift (BCPD) into the six-dimension scenario. Our contributions can be summarized as follows. (1) The PSR problem with normal vectors is formulated by generalizing the Bayesian coherent point drift (BCPD) approach; (2) The updated parameters during the algorithm's iterations are given in closed-forms; (3) Extensive experiments have been done to verify the proposed approach and its significant improvements over the BCPD has been validated. We have validated our proposed registration approach on both the human femur model. Results demonstrate that our proposed method outperforms the state-of-the-art registration methods and the convergence is guaranteed at the same time.**

I. Introduction

Point set registration (PSR) is an essential element in medical robotics, surgical navigation and computer-assisted surgery (CAS) [1] [2]. The aim of PSR is to recover the misalignment or the best transformation between two point sets (PS). PSR is generally used in CAS to map the preoperative volumetric image space with the intraoperative patient space [3].

The iterative closest point (ICP) [4] algorithm is the most classical and well-known algorithm for solving PSR problem. An iterative framework is used in ICP to find the optimal correspondences between points from two point sets in the first step, and the rigid transformation is updated in the next step. ICP iterates the correspondence steps and registration steps until the convergence condition is met. Although ICP achieves great success in many cases. there are still disadvantages inlcuding: 1) is susceptible to noises and outliers; 2) is sensitive to initial parameter values; 3) easily trapped in a local minimum [5]. Plenty of work in recent years has been proposed to improve the ICP's performances in variant aspects [6] [7].

This work is motivated by solving the rigid PSR problem for real-world applications, such as the surgical scenarios where the PSs are easily disturbed by noises and outliers. In this paper, we provide a novel PSR method that utilizes the features extracted from the PSs and can guarantee the convergence of the algorithm. Besides the positional information of each point, our method also utilizes the orientation information provided by the normal vectors which can be readily obtained in these ways in practice: 1) using the PCA method in the neighborhood of a point [8]; 2) utilizing a calibrated probe with a force/torque sensor or other methods for range imaging [9]. The orientational information can help to improve the robustness and accuracy of the registration. An intuitive understanding is that with great information comes great result. More specifically, we formulate the point set registration problem with normal vectors into a variational Bayesian inference (VBI) problem [10] using a hybrid mixture model (HMM) which consists of a von-Mises-Fisher (vMF) mixture model (FMM) [11] and a Gaussian mixture model (GMM). The GMM is chosen to model the positional information, while the FMM is used to represent the orientational information (the distribution of normal vectors). By VBI, we use an alternative distribution to approximate the maximum of the posterior probability of the overall HMM, which means that two point sets are optimally registered. The significant advantage over formulating the registration as a maximum likelihood estimation problem and solving with the expectation maximization framework is, the convergence can be well guaranteed in the Bayesian setting [12].

In this work, we cast the pairwise PSR problem with orientational information into a VBI frame for the first time. Our contributions to this paper can be summarized as follows.

1) The PSR problem with normal vectors is formulated by generalizing the Bayesian coherent point drift (BCPD) approach;
2) The updated parameters during the algorithm's iterations are given in closed-forms;
3) Extensive experiments have been done to verify the proposed approach and its significant improvements over the BCPD has been validated.

†Equal contribution.

[1]Ang Zhang and Jin Pan are with the Department of Electronic Engineering, The Chinese University of Hong Kong, N.T., Hong Kong SAR, China. azhang,jpan@link.cuhk.edu.hk

[2]Zhe Min is with Wellcome/EPSRC Centre for Surgical and Interventional Sciences, University College London, UK. z.min@ucl.ac.uk

[3]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.

[§]Corresponding author, max.meng@cuhk.edu.hk

II. Related Work

In this section, we first discuss the ICP algorithm and its variants. Then we review the probabilistic methods, especially based on the variational Bayesian inference framework and general EM framework. Finally, we discuss the deep-learning-based methods and their flaws and limitations.

A. ICP and Its Variants

As we have mentioned in the previous section, ICP is the most commonly used approach for rigid PSR. However, the assumptions including appropriate initial transformation matrix and less noise and outliers, cannot be always satisfied in real-word application. There are many variants [7] [13] of ICP to improve the PSR performance. In the category of 3D-3D registrations, two previous studies were conducted to incorporate both positional and orientational information into a probabilistic framework to solve the registration Two studies [9] [14] under the ICP framework propose to integrate both positional and orientational information into a probabilistic system to solve PSR problem. Billings *et al.* [9] utilize the Gaussian and Kent distributions to represent the anisotropic positional and orientational data respectively. However, because of their ICP framework, these improved algorithms still suffer from noise and outliers heavily.

B. Probabilistic Algorithms

In this paradigm, probabilistic PSR algorithms generally adopt a soft assignment strategy which means the correspondence confidences in two PS are modeled by the probabilistic distributions. Coherent point drift (CPD) [15] is one of the most representative probabilistic methods. In CPD, PSR problem is formulated as a probability density estimation problem based on a GMM. EM algorithm is applied to solve the likelihood estimation problem, in which E step calculates the posterior of latent variables and M step updates the parameters of the transformation. JRMPC [16] proposes a generative model that can jointly registers multiple point clouds. Its performance exceeds or is comparable to the classic and SOTA methods, such as GMMReg [17] and CPD [18]. Moreover, some probabilistic algorithms which utilize the normal vectors [19]–[22] can improve performances compared with existing methods. However, these methods are under the CPD framework and exist some problems on the convergence, parameter tuning, and robustness.

Very recently, Hirose has proposed a method named BCPD that reformulates CPD using VBI and achieves rigid and non-rigid registration in one algorithm [12]. It uses variational inference to substitute the motion coherence theory of CPD and improves several problems about the convergence, parameter tuning, a restricted acceleration scheme.

C. Deep Learning Methods

Along with the development of deep learning on 3D point clouds, such as PointNet [23], PointCNN [24] and DGCNN [25], more and more deep-learning-based registration methods are proposed and have been successful in many PSR problems. These methods usually map the point clouds to high-dimensional space to learn features, then learn to find the correspondences and optimize the transformations for the best alignment. PointNetLK [26] learn the feature representations using PointNet and then estimate the alignment using LK algorithm. PRNet [27] matches the correspondence based on DGCNN feature and achieves the point set registration using an end-to-end fashion.

Although deep learning provides some successful and effective methods to solve the PSR problem, there are still flaws and limitations. For example, 1) even the state-of-the-art deep-learning-based approaches still struggle to produce acceptable inlier rates in real problems [28]; 2) For the application of computer assisted surgery, deep-learning methods cannot guarantee the desired error bounds to surgeons in both theoretical and practical aspects. In the contrast, some state-of-the-art probabilistic algorithms can handle the registration with different noise levels and outliers rates. Moreover, given the measurement error of the markers such as fiducial localization error (FLE), there exists developed theory to accurately estimated the registration error such as Target Registration Error (TRE) [29] or Total Target Registration Error (TTRE) [30] model.

III. Notations and Preliminaries

Throughout this paper, we obey the following conventions:

- $\mathbf{X} = (\mathbf{x}_1, \cdots, \mathbf{x}_N) \in \mathbb{R}^{3\times N}$ - the measured (target) position point set .
- $\hat{\mathbf{X}} = (\hat{\mathbf{x}}_1, \cdots, \hat{\mathbf{x}}_N) \in \mathbb{R}^{3\times N}$ - the measured (target) orientational vector set , each vector corresponds to one point in $\mathbf{X}$ with the same subscript.
- $\mathbf{Y} = (\mathbf{y}_1, \cdots, \mathbf{y}_M) \in \mathbb{R}^{3\times M}$ - the original (source) position point set , which represents the centroids of GMM.
- $\hat{\mathbf{Y}} = (\hat{\mathbf{y}}_1, \cdots, \hat{\mathbf{y}}_M) \in \mathbb{R}^{3\times M}$ - the original (source) orientation unit vector set , which represents the mean directions of FMM.
- $\kappa \in \mathbb{R}$ - concentration parameter of the FMM.
- $\sigma^2 \in \mathbb{R}$ - isotropic variances of the GMM.
- $\alpha = (\alpha_1, \cdots, \alpha_M)^T \in [0,1]^M$ - the membership probabilities which represents mixing proportion of HMM components, s.t. $\sum_{m=1}^{M} \alpha_m = 1$.
- $c = (c_1, \cdots, c_N)^T \in \{0,1\}^N$ - $c_n = 0$ if the nth point in $\mathbf{X}$ is a outlier, otherwise equals 1.
- $e = (e_1, \cdots, e_N)^T$ - the index vector that $e_n = m$ represents $\mathbf{x}_n$ corresponds to $\mathbf{y}_m$.
- Θ - the set of all latent variables and parameters $\{\alpha, c, e, \mathbf{R}, \mathbf{t}, \sigma^2, \kappa\}$.
- $p_{mn} \in [0,1]$ - the posterior probability which represents the correspondence probability between $\mathbf{y}_m$ and $\mathbf{x}_n$, and is defined as $p_{mn} = E[c_n \delta_m(e_n)]$.
- $\mathbf{P} = (p_{mn}) \in [0,1]^{M\times N}$ - the M×N probability matrix.

In this section, we will review the variational Bayesian inference (VBI) [10]. The target of VBI is to estimate the latent variables and parameters, $\mathbf{Z}$, given the observed set, $\mathbf{D}$. For this purpose, the posterior distribution $p(\mathbf{Z}|\mathbf{D})$ or the expectation $E[\mathbf{Z}]$ with respect to $p(\mathbf{Z}|\mathbf{D})$ is necessary. However, it is always impracticable in many practical models due

to the difficult calculation of the expectation and the posterior distribution. Therefore, we need to apply an approximation method such as VBI to reduce the computational difficulty. A tractable distribution $q(\mathbf{Z})$ is chosen to approximate the true posterior distribution $p(\mathbf{Z}|\mathbf{D})$. Generally, we can define the marginal probability of the model evidence $p(\mathbf{D})$ as follows:

$$\ln p(\mathbf{D}) = \underbrace{\int q(\mathbf{Z}) \ln \left\{ \frac{p(\mathbf{D}, \mathbf{Z})}{q(\mathbf{Z})} \right\} \mathrm{d}\mathbf{Z}}_{\mathcal{L}(q)} \underbrace{- \int q(\mathbf{Z}) \ln \left\{ \frac{p(\mathbf{Z} \mid \mathbf{D})}{q(\mathbf{Z})} \right\} \mathrm{d}\mathbf{Z}}_{\mathrm{KL}(q \| p)} \quad (1)$$

Then, VBI can be regarded as the minimization of the KL divergence above. It is equivalent to maximizing the evidence lower bound (ELBO) $\mathcal{L}(q)$. To solve the problem of parametric optimization, $q(\mathbf{Z})$ is decomposed as $q(\mathbf{Z}) = \prod_{i=1}^{N} q_i(\mathbf{Z}_i)$, where each $\mathbf{Z}_i$ is independent. After some manipulations, the general solution of posterior approximate distribution $q_j^\star(\mathbf{Z}_j)$ is obtained:

$$\ln q_j^\star(\mathbf{Z}_j) = E_{i \neq j}[\ln p(\mathbf{D}, \mathbf{Z})] + const, \quad (2)$$

where $E_{i \neq j}[\ln p(\mathbf{D}, \mathbf{Z})] = \int \ln p(\mathbf{D}, \mathbf{Z}) \prod_{i(\neq j)}^{N} q_i \, \mathrm{d}\mathbf{Z}_i$ and the additive constant can be set by normalizing the distribution $q_j^\star(\mathbf{Z}_j)$. VBI uses coordinate ascent algorithm to update q_j, which guarantees monotonically increasing of ELBO and the convergence of VBI.

IV. Methods

A. Problem Formulation

Given two point sets (PSs) $\mathbf{X}, \mathbf{Y}$ and corresponding normal vector sets $\hat{\mathbf{X}}$, $\hat{\mathbf{Y}}$, the points in $\mathbf{Y}$ are regarded as the centroids of GMM while the mean directions of the FMM is defined as the vectors from $\hat{\mathbf{Y}}$. Moreover, the points and normal vectors in the target point set $\mathbf{X}$ and orientation unit vector set $\hat{\mathbf{X}}$ are generated from the GMM and the FMM respectively. The ultimate goal of our PSR is to find the optimal rigid transformation matrix that matches $(\mathbf{X}, \hat{\mathbf{X}})$ and $(\mathbf{Y}, \hat{\mathbf{Y}})$. Under the framework of HMM, we define the probability density function of $\mathbf{x}_n$ given $e_n = m$ (nth observed generalized point is assigned to the mth HMM component) as follows:

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

where $\mathcal{F}(\hat{\boldsymbol{\mu}}, \kappa)$ and $\mathcal{N}(\boldsymbol{\mu}, \sigma^2)$ represent a vMF and a normal distribution respectively. We denote φ_{mn} as a symbol of Eq. (3). The outliers is defined as a uniform distribution $p_{out}(\mathbf{x}_n) = 1/V$ to satisfy the normalization condition of $p_{out}(\mathbf{x}_n)$, where V is the minimum volume that can contain all points in $\mathbf{X}$. ω is used to denote the weight of this uniform distribution, where $\omega \in [0, 1]$. We use an explicit definition $\delta_m(e_n)$ to indicate the correspondence where if $e_n = m$, $\delta_m(e_n)$ is 1, otherwise 0.

Therefore, the joint probability density function of $p(\mathbf{x}_n, \hat{\mathbf{x}}_n, c_n, e_n)$ given $(\mathbf{Y}, \hat{\mathbf{Y}}, \alpha, \mathbf{R}, \mathbf{t}, \sigma^2, \kappa)$ can be written as follows:

$$\begin{aligned} &p\left(\mathbf{x}_n, \hat{\mathbf{x}}_n, e_n, c_n \mid \mathbf{Y}, \hat{\mathbf{Y}}, \alpha, \mathbf{R}, \mathbf{t}, \sigma^2, \kappa\right) \\ &= \left\{\omega p_{\text{out}}(\mathbf{x}_n)\right\}^{1 - c_n} \left\{(1 - \omega) \prod_{m=1}^{M} \left(\alpha_m \varphi_{mn}\right)^{\delta_m(e_n)}\right\}^{c_n}, \end{aligned} \quad (4)$$

which consists of two mixture distributions: the left item have two components, and the right item have M components.

Then we will introduce the approach to find the optimal parameters $\mathbf{R}, \mathbf{t}, \sigma^2, \kappa$ using VBI algorithm under the HMM framework described above.

B. Variational Bayesian Formulation

In this paper, Bayesian inference is used to solve the problem of parametric optimization. Our task is to find a distribution $q(\Theta)$ to approximate the true posterior distribution $p(\Theta|\mathbf{X}, \hat{\mathbf{X}}, \mathbf{Y}, \hat{\mathbf{Y}})$. In this section, we introduce the prior distribution over the items in Θ and the joint probability distribution to formulate the variational treatment of HMM.

1) Prior Distribution:

In this part, we introduce priors over the parameters. We define a Dirichlet distribution over the membership probabilities α as follows:

$$p(\alpha) = \mathbf{Dir}(\alpha \mid \lambda \mathbf{1}_M) = C(\lambda \mathbf{1}_M) \prod_{m=1}^{M} \alpha_m^{\lambda - 1}, \quad (5)$$

where $\mathbf{1}_M$ is a column vector of all ones with size M. $C(\lambda \mathbf{1}_M)$ is the normalization constant. The Dirichlet distribution can be used to control the mixture proportion of HMM components. With smaller λ, the data will further influence the posterior, and the prior will bring less impact on the posterior.

To simplify the variational model, we don't introduce the priors over the original point and normal vector set $(\mathbf{Y}, \hat{\mathbf{Y}})$, the variance parameter σ, the concentration parameter κ, transformation parameters $\mathbf{R}, \mathbf{t}$.

2) Joint Probability Distribution: After combining the prior distribution into the HMM, we can write down the full joint probability distribution:

$$\begin{aligned} p(\mathbf{X}, \hat{\mathbf{X}}, \mathbf{Y}, \hat{\mathbf{Y}}, \Theta) \propto \\ p(\alpha) \prod_{n=1}^{N} p\left(\mathbf{x}_n, \hat{\mathbf{x}}_n, c_n, e_n \mid \mathbf{Y}, \hat{\mathbf{Y}}, \alpha, \mathbf{R}, \mathbf{t}, \sigma^2, \kappa\right) \end{aligned} \quad (6)$$

C. Variational Bayesian Approximate Posteriors

In this section, we will derive the approximated posterior distributions based on the variational Bayesian inference described in section III. According to mean field theory and the conditional independence relation between variables and parameters, $q(\Theta)$ can be factorized as follows:

$$q(\Theta) = q_1(\alpha) q_2(c, e) q_3(\mathbf{R}, \mathbf{t}, \sigma^2, \kappa). \quad (7)$$

This factorization is applied to guarantee our Bayesian hybrid mixture model to have a reasonably computable solution. We update $q(\Theta)$ in each iteration.

1) $q_1(\alpha)$:

Using the general solution (2) and the product rule for probabilities, we can obtain the optimal solution for $q_1^\star(\alpha)$ as follows:

$$\ln q_1^\star(\alpha) = E_{q_2,q_3}[\ln p(\mathbf{X}, \hat{\mathbf{X}}, \mathbf{Y}, \hat{\mathbf{Y}}, \Theta)] + const. \tag{8}$$

Substituting the decomposition (6) into the solution and removing the terms that are independent of α into the normalization constant, we obtain

$$\begin{aligned}\ln q_1^\star(\alpha) &= \sum_{n=1}^{N}\sum_{m=1}^{M} E_{q_2,q_3}[c_n\delta_m(e_n)\ln(\alpha_m\varphi_{mn})] \\ &\quad + \sum_{m=1}^{M}\ln\alpha_m^{\lambda-1} + const \\ &= \sum_{m=1}^{M}\ln\alpha_m^{\lambda-1+\rho_m} + const,\end{aligned} \tag{9}$$

where $\rho_m = \sum_{n=1}^{N} p_{mn}$, and $p_{mn} = E[c_n\delta_m(e_n)]$ which is the posterior that $\mathbf{x}_n$ corresponds to $\mathbf{y}_m$. Taking the exponential of the equation above, we obtain $q_1^\star(\alpha)$ which follows a Dirichlet distribution:

$$q_1^\star(\alpha) = \mathbf{Dir}(\alpha \mid \lambda\mathbf{1}_M + \rho) \tag{10}$$

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

2) $q_2(c, e)$:

Similar to the previous step, we continue to derive the optimal solution of $q_2(c, e)$ which represents the shape correspondence between two point sets. Substituting the joint distribution (6) into the general solution (2) and removing the terms that are independent of (c, e) into the normalization constant, we obtain the optimal $q_2^\star(c, e)$ as follows:

$$\begin{aligned}\ln q_2^\star(c,e) = \sum_{n=1}^{N}\Big[&\ln\left\{\omega p_{\text{out}}\left(\mathbf{x}_n\right)\right\}^{(1-c_n)} \\ &+ \sum_{m=1}^{M}\ln\left\{(1-\omega)\left\langle\alpha_m\right\rangle\left\langle\varphi_{mn}\right\rangle\right\}^{c_n\delta_m(e_n)}\Big] + const,\end{aligned} \tag{11}$$

where $\langle\alpha_m\rangle$ = exp$(E[\ln\alpha_m])$ and $\langle\varphi_{mn}\rangle$ = exp$(E[\ln\varphi_{mn}])$. In the last part, we prove $q_1(\alpha)$ is a Dirichlet distribution. Therefore, making use of the standard results of the Dirichlet distribution, we obtain :

$$\langle\alpha_m\rangle = exp[\psi(\lambda+\rho_m) - \psi(\lambda M + N_{\mathbf{p}})], \tag{12}$$

where we define $N_{\mathbf{p}} = \sum_{n=1}^{N}\sum_{m=1}^{M} p_{mn}$ and $\psi(\cdot)$ is the digamma function. Because the registration is rigid in this paper and the variance is isotropic, then we have :

$$\langle\varphi_{mn}\rangle = \varphi_{mn}. \tag{13}$$

Observing the Eq. (11), $q_2^\star(c, e)$ can be written in this form $q_2^\star(c,e) = \prod_{n=1}^{N} q_2^{\star(n)}(c_n, e_n)$, where

$$q_2^{\star(n)}(c_n, e_n) = \frac{1}{C}\cdot\left\{\omega p_{\text{out}}\left(\mathbf{x}_n\right)\right\}^{(1-c_n)}\prod_{m=1}^{M}\left\{(1-\omega)\left\langle\alpha_m\right\rangle\left\langle\varphi_{mn}\right\rangle\right\}^{c_n\delta_m(e_n)}. \tag{14}$$

Because only one component is effective for one pair of e_n and c_n, the normalization constant C is defined as the sum of all components in Eq. (14) which can be written as $C = \omega p_{\text{out}}\left(\mathbf{x}_n\right) + (1-\omega)\sum_{m=1}^{M}\left\langle\alpha_m\right\rangle\left\langle\varphi_{mn}\right\rangle$. Finally, we obtain the solution of $q_2^\star(c, e)$ as the following:

$$q_2^\star(c,e) = \prod_{n=1}^{N}\left(1-\rho_n'\right)^{1-c_n}\left\{\rho_n'\prod_{m=1}^{M}\left(\frac{p_{mn}}{\rho_n'}\right)^{\delta_m(e_n)}\right\}^{c_n}, \tag{15}$$

where $\rho' = \mathbf{P}^T\mathbf{1}_M$ and $\rho_n' = \sum_{m=1}^{M} p_{mn}$ is the posterior that $\mathbf{x}_n$ belongs to a non-outlier. The posterior probability p_{mn} represents the correspondence probability between $\mathbf{y}_m$ and $\mathbf{x}_n$, and is computed as follows:

$$p_{mn} = \frac{(1-\omega)\left\langle\alpha_m\right\rangle\left\langle\varphi_{mn}\right\rangle}{\omega p_{\text{out}}\left(x_n\right) + (1-\omega)\sum_{m=1}^{M}\left\langle\alpha_m\right\rangle\left\langle\varphi_{mn}\right\rangle}. \tag{16}$$

Due to $p_{mn} = q_2^{\star(n)}(c_n, e_n) = E[c_n\delta_m(e_n)]$, it is consistent with previous definition. Observing the Eq. (15), we can find $q_2^\star(c, e)$ consists of a categorical distribution and a Bernoulli distribution. This part ensures the update of $\mathbf{P}$ and the variables associated with $\mathbf{P}$ such as ρ and ρ', with the Eq. (16) improves the lower bound.

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

In this part, we assume q_3 as a Dirac delta function [31] which models an idealized point mass at $(\mathbf{R}, \mathbf{t}, \sigma^2, \kappa)$. It means that q_3 is only described by the first moment. Therefore, for the update of q_3, we choose to maximize the ELBO $\mathcal{L}(q)$ directly instead of applying the general solution (2) of VBI. Given the approximated posterior distribution $q_1(\alpha)$ and $q_2(c, e)$, we obtain the evidence lower bound as follows:

$$\begin{aligned}\mathcal{L}(q) = &E[\ln p(\mathbf{X}, \hat{\mathbf{X}}, \mathbf{Y}, \hat{\mathbf{Y}}, \Theta)] + const = \\ &-\frac{3}{2}N_{\mathbf{P}}\log\sigma^2 - N_{\mathbf{P}}\log\left(e^{\kappa} - e^{-\kappa}\right) + N_{\mathbf{P}}\log\kappa \\ &-\sum_{n=1}^{N}\sum_{m=1}^{M} p_{mn}\left(\frac{1}{2\sigma^2}\left\|\mathbf{x}_n - (\mathbf{R}\mathbf{y}_m + \mathbf{t})\right\|^2\right. \\ &\left.-\kappa\left(\left(\mathbf{R}\widehat{\mathbf{y}}_m\right)^{\mathrm{T}}\widehat{\mathbf{x}}_n\right)\right) + const.\end{aligned} \tag{17}$$

To maximize the ELBO with respect to $\mathbf{R}, \mathbf{t}, \sigma^2$ and κ, the first step we need is solving the partial derivative of Eq. (17) with respect to σ^2 and setting it to 0. Then the optimal σ^2 can be written as:

$$(\sigma^2)^\star = \frac{\sum_{n=1}^{N}\sum_{m=1}^{M} p_{mn}\left(\left\|\mathbf{x}_n - (\mathbf{R}\mathbf{y}_m + \mathbf{t})\right\|^2\right)}{3N_{\mathbf{p}}} \tag{18}$$

We define the $\xi(\mathbf{t}) = (\sigma^2)^\star$. According to Eqs. (17) and (18), we can see that minimizing $\xi(\mathbf{t})$ is maximizing the

$\mathcal{L}(q)$. Then, the optimal $\mathbf{t}^\star = \arg\min_t \xi(\mathbf{t})$ can be found by solving $\partial\xi(\mathbf{t})/\partial\mathbf{t} = 0$. After calculations, we can obtain $\mathbf{t}^\star$ as follows:

$$\mathbf{t}^\star = \bar{\mathbf{x}} - \mathbf{R}\bar{\mathbf{y}}, \tag{19}$$

where

$$\begin{aligned} \bar{\mathbf{x}} &= \tfrac{1}{N_\mathrm{p}} \textstyle\sum_{n=1}^{N} \sum_{m=1}^{M} p_{mn}\mathbf{x}_n = \tfrac{1}{N_\mathbf{p}}\mathbf{X}^\mathrm{T}\mathbf{P}^\mathrm{T}\mathbf{1}_M, \\ \bar{\mathbf{y}} &= \tfrac{1}{N_\mathbf{p}} \textstyle\sum_{n=1}^{N} \sum_{m=1}^{M} p_{mn}\mathbf{y}_m = \tfrac{1}{N_\mathbf{p}}\mathbf{Y}^\mathrm{T}\mathbf{P}\mathbf{1}_N, \end{aligned} \tag{20}$$

represent the expectations of two position point sets respectively.

To find optimal $\mathbf{R}^\star$ which maximizes the lower bound, we should solve the optimization problem as follows:

$$\begin{aligned} \mathbf{R}^\star &= \arg\max_{\mathbf{R}} \mathcal{L}(q) \\ &= \arg\max_{\mathbf{R}} -\sum_{n=1}^{N}\sum_{m=1}^{M} p_{mn}\left(\frac{1}{2\sigma^2}\left\|\mathbf{x}_n - (\mathbf{R}\mathbf{y}_m + \mathbf{t})\right\|^2\right. \\ &\left. -\kappa\left((\mathbf{R}\widehat{\mathbf{y}}_m)^\mathrm{T}\widehat{\mathbf{x}}_n\right)\right), \text{ s.t. } \mathbf{R}^\mathrm{T}\mathbf{R} = \mathbf{I}_3, \ \det(\mathbf{R}) = 1 \end{aligned} \tag{21}$$

Let us introduce some notations at first as follows:

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

Then substitute the optimal $\mathbf{t}^\star$ in Eq. (19) into $\mathbf{R}^\star$ in Eq. (23), using the above notations, we obtain:

$$\begin{aligned} \mathbf{R}^\star &= \arg\max_{\mathbf{R}} \left(\frac{1}{\sigma^2}\sum_{n=1}^{N}\sum_{m=1}^{M} p_{mn}\tilde{\mathbf{x}}_n^\mathrm{T}\mathbf{R}\tilde{\mathbf{y}}_m \right. \\ &\left. +\kappa\sum_{n=1}^{N}\sum_{m=1}^{M} p_{mn}(\mathbf{R}\widehat{\mathbf{y}}_m)^\mathrm{T}\widehat{\mathbf{x}}_n\right) \\ &= \arg\max_{\mathbf{R}} \ \mathrm{Tr}\left(\mathbf{R}(\mathbf{H}_p + \mathbf{H}_o)\right), \end{aligned} \tag{22}$$

where $\mathrm{Tr}(\cdot)$ denotes the trace of a matrix. Applying the lemma in [18], the optimal $\mathbf{R}^\star$ can be solved as follows:

$$\mathbf{R}^\star = \mathbf{V}\mathrm{d}(1, 1, \det(\mathbf{V}\mathbf{U}^\mathrm{T}))\mathbf{U}^\mathrm{T}, \tag{23}$$

where $\mathbf{U}$ and $\mathbf{V}$ can be obtained by SVD of $\mathbf{H}$ as $\mathbf{H} = \mathbf{U}\mathbf{H}'\mathbf{V}^\mathrm{T}$, and $\mathrm{d}(\cdot)$ is the diagonal matrix of a vector.
Then we can update $\mathbf{t}^\star$ and $(\sigma^2)^\star$ by

$$\begin{aligned} \mathbf{t}^\star &= \bar{\mathbf{x}} - \mathbf{R}^\star\bar{\mathbf{y}}, \\ (\sigma^2)^\star &= \frac{\sum_{n=1}^{N}\sum_{m=1}^{M} p_{mn}\left(\left\|\mathbf{x}_n - (\mathbf{R}^\star\mathbf{y}_m + \mathbf{t}^\star)\right\|^2\right)}{3N_\mathbf{p}}. \end{aligned} \tag{24}$$

For the update of κ, we use the method described in [11] which divides κ into two parts. One part is produced by the position error err_p which is written as:

$$err_p = \frac{\sum_{n=1}^{N}\sum_{m=1}^{M} p_{mn}\tilde{\mathbf{x}}_n^\mathrm{T}\mathbf{R}\tilde{\mathbf{y}}_m}{\sum_{n=1}^{N}\sum_{m=1}^{M} p_{mn}\left\|\mathbf{R}\tilde{\mathbf{y}}_m\right\|\left\|\tilde{\mathbf{x}}_n\right\|}. \tag{25}$$

The other part is produced because of orientation error err_o:

$$err_o = \frac{1}{N_\mathbf{p}}\sum_{n=1}^{N}\sum_{m=1}^{M} p_{mn}\left(\mathbf{R}\widehat{\mathbf{y}}_m\right)^\mathrm{T}\widehat{\mathbf{x}}_n. \tag{26}$$

Then we define $err = \varepsilon err_p + (1-\varepsilon)err_o$ with $\varepsilon = 0.5$. Then we can obtain the optimal $\kappa^\star$ as the following:

$$\kappa^\star = \frac{err(3 - err^2)}{1 - err^2}. \tag{27}$$

D. Implementation Details

The detailed procedures of our method are shown in Algorithm 1. The model parameters should be first initialized. Besides the parameters which have been illustrated in Algorithm 1, we set κ to 10 empirically, and the value of ω is initialized to 0.5. The λ is set to infinity to unify the related setting of BCPD. One significant consideration is that κ cannot be large to ensure the calculability of e^κ. Therefore, an upper bound is defined for κ as 100 during the update. We also use the following conditions to determine whether the algorithm converges: 1) the difference $\Delta\sigma^2$ between two iterations is smaller than 10^{-6}; 2) the maximum iterations is 100; 3) σ^2 is smaller than the threshold value 10^{-6}.

Algorithm 1 Robust Generalized Point Set Registration Based on Variational Bayesian Inference

1: **Initialization**: $\mathbf{R} = \mathbf{I}_3$, $\mathbf{t} = \mathbf{0}_{3\times1}$, ω, κ, λ, $\sigma^2 = \frac{1}{3MN}\sum_{n=1}^{N}\cdot\sum_{m=1}^{M}\left\|\mathbf{x}_n - \mathbf{y}_m\right\|^2$, $\langle\alpha_m\rangle = \frac{1}{M}$.
2: **repeat**
3: - Update $\langle\alpha_m\rangle$, φ_{mn} and $\mathbf{P} = (p_{mn})$ by (12), (13) and (16) respectively
4: - Update $\mathbf{R}$ and $\mathbf{t}$ by (23) and (24)
5: - Update σ^2 and κ by (24) and (27)
6: **until** Convergence.
7: **return** $\mathbf{R}^\star$ and $\mathbf{t}^\star$.

V. Experiments

Two groups of experiments are conducted to validate the robustness and accuracy of our proposed method. The source point sets $(\mathbf{Y}, \hat{\mathbf{Y}})$ are generated from the CT model of the femur bone which is shown in Fig.1. The target point $(\mathbf{X}, \hat{\mathbf{X}})$ are produced by sampling from the CT model point set $(\mathbf{Y}, \hat{\mathbf{Y}})$. We then add two different noise levels to the points of $(\mathbf{X}, \hat{\mathbf{X}})$ in two groups of experiments respectively. Furthermore, each group is divided into nine cases where the different ratio outliers are added to the target point to produce the final *disturbed* $(\mathbf{X}, \hat{\mathbf{X}})$. We do 100 trials in every case and define the number of X as 100 (excluding outliers). The ground-truth value of rigid transformation matrix $[\mathbf{R}_\mathbf{true}, \mathbf{t}_\mathbf{true}]$ in each case is randomly chosen from $[10°, 25°]$ and $[10\mathrm{mm}, 25\mathrm{mm}]$. Then the final source point set *misaligned* $(\mathbf{Y}, \hat{\mathbf{Y}})$ are generated by applying the real $[\mathbf{R}_\mathbf{true}, \mathbf{t}_\mathbf{true}]$ to $(\mathbf{Y}, \hat{\mathbf{Y}})$.

We use the rotation error and translation error as registration errors and compute the mean errors among 100 trials in each case to evaluate the performance. We define the rotation error and translation error as follows: $\theta_{error} = \arccos\left[\mathrm{tr}\left(\mathbf{R}_\mathbf{true}\mathbf{R}_{cal}^\top - 1\right)/2\right]$, and $\mathbf{t}_{error} = \left\|\mathbf{t}_{cal} - \mathbf{t}_\mathbf{true}\right\|_2$, where $\mathbf{R}_{cal}$ and $\mathbf{t}_{cal}$ are the transformation matrix between *disturbed* $(\mathbf{X}, \hat{\mathbf{X}})$ and *misaligned* $(\mathbf{Y}, \hat{\mathbf{Y}})$ obtained by our method.

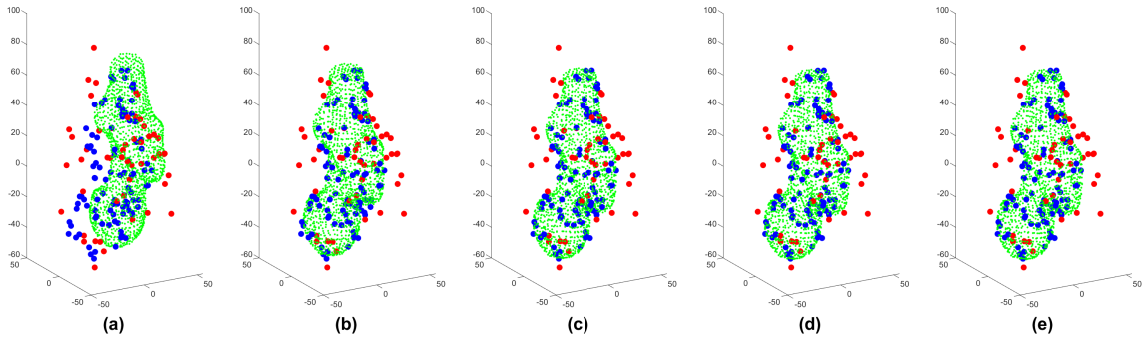

Fig. 1. Registration Process with 50% outliers and 1mm/1° noise. The unit of measure is mm. (a)-(d): the registration result in 0th, 5th, 10th and 100th iteration using the proposed method; (e): ground truth. Blue: inliers in X; Red: outliers in X; Green: points in Y.

TABLE I

ROTATION AND TRANSLATION ERRORS UNDER LOW NOISE LEVEL. DIFFERENT RATIOS OF OUTLIERS ARE ADDED TO $(\mathbf{X}, \hat{\mathbf{X}})$.

Error Type	Method	10%	20%	30%	40%	50%	60%	70%	80%	90%
Rotation (Degree)	ICP	2.6877	3.6509	3.9594	4.5431	4.8543	5.2485	5.3575	4.8144	5.2281
Rotation (Degree)	BCPD	0.9949	1.1307	1.0474	1.5096	3.1661	3.4587	4.1637	3.7359	4.1753
Rotation (Degree)	Proposed Method	**0.9530**	**1.0261**	**1.0353**	**1.0143**	**1.1096**	**1.0123**	**0.9228**	**1.0080**	**1.1005**
Translation (mm)	ICP	1.0479	1.4309	1.6958	1.7979	1.8395	2.0176	2.1533	2.1878	2.0304
Translation (mm)	BCPD	1.1007	0.6440	0.5283	0.8482	0.6397	0.5914	0.5633	0.6825	0.6329
Translation (mm)	Proposed Method	**0.4804**	**0.5800**	**0.5228**	**0.5771**	**0.5204**	**0.5696**	**0.5581**	**0.5968**	**0.4781**

TABLE II

ROTATION AND TRANSLATION ERRORS UNDER HIGH NOISE LEVEL. DIFFERENT RATIOS OF OUTLIERS ARE ADDED TO $(\mathbf{X}, \hat{\mathbf{X}})$.

Error Type	Method	10%	20%	30%	40%	50%	60%	70%	80%	90%
Rotation (Degree)	ICP	3.5297	4.2803	4.5789	5.0979	5.3169	4.9937	5.2821	5.9766	5.1078
Rotation (Degree)	BCPD	2.3523	2.5753	2.8529	3.2450	3.6830	4.2149	3.9536	4.3385	4.3145
Rotation (Degree)	Proposed Method	**2.2005**	**2.2277**	**2.5437**	**2.3418**	**2.2074**	**2.3365**	**2.1846**	**2.4130**	**2.8599**
Translation (mm)	ICP	1.3525	1.7002	1.9964	1.9823	2.1776	2.0680	2.0100	2.1745	2.0908
Translation (mm)	BCPD	1.2486	1.1849	1.8629	1.2522	1.7430	1.9102	1.8588	1.8648	1.8916
Translation (mm)	Proposed Method	**1.1780**	**1.1585**	**1.1992**	**1.2345**	**1.1735**	**1.1423**	**1.2638**	**1.3381**	**1.2843**

A. Low Noise Level

In the first experiment, the *disturbed* $(\mathbf{X}, \hat{\mathbf{X}})$ are generated in two steps: 1) add the low noise 1mm/1° standard deviations (std) to $\mathbf{X}$ and $\hat{\mathbf{X}}$ respectively. To obtain 1° std, we define $\kappa = 3200$ following the illustration in [9]. 2) Add nine ratios of outliers from 10% to 90% with an interval of 10%. The number of inlier points is fixed at 100. The outliers are generated by adding a uniformly distributed displacement within $[20\text{mm}, 30\text{mm}]$ to the randomly sampled points from the CT model point set. Then we obtain the *disturbed* $(\mathbf{X}, \hat{\mathbf{X}})$ with $N = 110$ to 190. We compare our method with ICP [4] and BCPD [12]. One hundred trials are conducted for each case.

B. High Noise Level

Similar with the last part, we set κ as 800 to obtain the std of 2°. Then the high noise 2mm/2° stds are added to $(\mathbf{X}, \hat{\mathbf{X}})$. Following the same process described above, we obtain the second group of *disturbed* $(\mathbf{X}, \hat{\mathbf{X}})$. In this part, we also evaluate the performance of the proposed method compared with ICP and BCPD.

VI. RESULTS AND DISCUSSION

A. Two Noise Levels

We summarize the results of the two experiments in Table I and II respectively. In the first group, the low noise is added to the target sets $(\mathbf{X}, \hat{\mathbf{X}})$. As shown in Table I, both rotation and translation errors increase for ICP and BCPD methods as the ratio of outliers becomes larger. On the contrary, the performance of our method is still stable with the increasing of the outliers. In each case, our proposed method exceeds ICP and BCPD, especially in the cases with more outliers. In the second group, the high level of noise is added to $(\mathbf{X}, \hat{\mathbf{X}})$. We can find that every method performs worse with larger noise by comparing Table II with Table I. However, the proposed method is still stable like in Table I with the increasing of the outliers and performs best in the three methods. It indicates the accuracy and robustness of our proposed method. We also use the ***ttest*** function in MATLAB to calculate the p-value of our algorithm compared with ICP and BCPD respectively. Almost all p-values are not more than 0.05 which means 0.05 significance level. It demonstrates that the errors in our experiments are statistically significant.

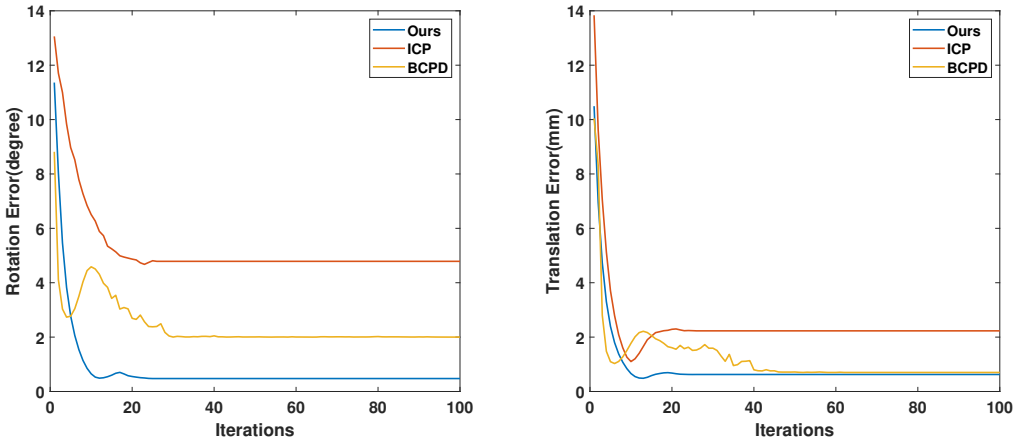

Fig. 2. Convergence speed of three methods with respect to the iterations. 1mm/1° noise and 50% outliers are chosen. Left: rotation error. Right: translation error.

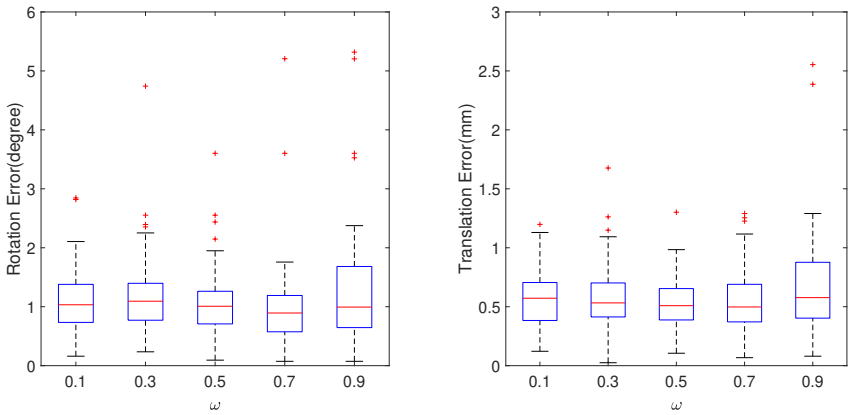

Fig. 3. Convergence speed of three methods with respect to the iterations. 1mm/1° noise and 50% outliers are chosen. Left: rotation error. Right: translation error.

B. Convergence Speed

Fig. 1(a)-(d) presents the process of registration in 0th, 5th, 10th and 100th iteration using the proposed method. The Fig. 1(e) shows the ground truth of the registration with $[\mathbf{R}_{\mathbf{true}}, \mathbf{t}_{\mathbf{true}}]$. We can see that the $[\mathbf{R}_{cal}, \mathbf{t}_{cal}]$ in 10th iteration is almost same as the result of the 100th iteration. We also choose one trial which is added 1mm/1° noise and 50% outliers to test the convergence speed of three methods. Fig. 2 shows the convergence speed of the rotation error and translation error directly. We can see that our method obtains the smallest error with a faster convergence speed compared with the other two methods. ICP also converges fast, but the final error is bigger. BCPD is just the opposite of ICP, which has a slower speed with smaller errors.

C. Parameter ω

The parameter ω represents the weight of $p_{out}(\mathbf{x}_n)$ which can be regarded as the prior that $\mathbf{x}_n$ belongs to an outlier. To find the influence of the ω on the registration results, we set nine different ω from 0.1 to 0.9 with an interval of 0.1. For each case, we also conduct 100 trials. As shown in Fig. 3, we can find the parameter ω have a very small influence on the final registration error in different ω. Therefore, we set $\omega = 0.5$ in the initialization.

To sum up, our proposed algorithm can outperform the state-of-the-art algorithms in 1) robustness to noise and outliers; 2) registration accuracy; 3) convergence speed.

VII. Conclusions

In this paper, we proposed a robust VBI-based registration method under the HMM framework. Besides the position information of points, we also consider the orientation information. Our method has a better performance on robustness and accuracy compared with the state-of-the-art methods. It has great potential in computer-assisted surgery (CAS) and other PSR applications. For future work, we will try to introduce the prior distributions over σ and κ, such as Gaussian-Wishart distribution and Gaussian-Gamma distribution, and explore the acceleration method for the complexity model. Besides, we will further consider the anisotropic cases to reduce related errors.

APPENDIX

A. Derivation of Eq.(17)

To simplify writing, we denote $x = (\mathbf{X}, \hat{\mathbf{X}}), y = (\mathbf{Y}, \hat{\mathbf{Y}}), \theta_1 = \alpha, \theta_2 = (c, e), \theta_3 = (\mathbf{R}, \mathbf{t}, \sigma^2, \kappa)$. Given $q(\theta_1)$ and $q(\theta_2)$, we have:

$$\begin{aligned}
\mathcal{L}(q) &= \int q(\theta) \cdot \ln \frac{p(x,y,\theta)}{q(\theta)} d\theta \\
&= \int q\left(\theta_1\theta_2\theta_3\right) \cdot \ln \frac{p\left(x,y,\theta_1,\theta_2,\theta_3\right)}{q\left(\theta_1\theta_2\right) \cdot q\left(\theta_3\right)} d\theta_1 d\theta_2 d\theta_3 \\
&= \int q\left(\theta_1\theta_2\theta_3\right) \ln \frac{p\left(x,y,\theta_1,\theta_2,\theta_3\right)}{q\left(\theta_1 \cdot \theta_2\right)} d\theta_1 d\theta_2 d\theta_3 \\
&\quad - \int q\left(\theta_1\theta_2\right) q(\theta_3) \ln q(\theta_3) d\theta_1 d\theta_2 d\theta_3 \\
&= E_{q(\theta_1\theta_2\theta_3)}[\ln p(x,y,\theta_1,\theta_2,\theta_3)] - E_{q(\theta_1,\theta_2)}\left[\ln q\left(\theta_1\theta_2\right)\right] \\
&\quad - \int q(\theta_3) \ln q(\theta_3) d\theta_3 \\
&= E_{q(\theta_1\theta_2\theta_3)}[\ln p(x,y,\theta_1,\theta_2,\theta_3)] - \int q(\theta_3) \ln q(\theta_3) d\theta_3 \\
&\quad + const,
\end{aligned} \tag{28}$$

Because $q_3\left(\theta_3\right)$ is a Dirac delta function, we can drop the entropy term $-\int q(\theta_3) \ln q(\theta_3) d\theta_3$. Then we obtain the form of Eq. (17).

ACKNOWLEDGMENT

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.

REFERENCES

[1] 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.

[2] 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.

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

[4] 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.

[5] 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.

[6] S. D. Billings, E. M. Boctor, and R. H. Taylor, "Iterative most-likely point registration (imlp): A robust algorithm for computing optimal shape alignment," *PloS one*, vol. 10, no. 3, p. e0117688, 2015.

[7] J. Yang, H. Li, and Y. Jia, "Go-icp: Solving 3d registration efficiently and globally optimally," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1457–1464.

[8] 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.

[9] 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.

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

[11] 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.

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

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] Z. Min, J. Wang, and M. Q.-H. Meng, "Joint rigid registration of multiple generalized point sets with hybrid mixture models," *IEEE Transactions on Automation Science and Engineering*, vol. 17, no. 1, pp. 334–347, 2019.

[20] Z. Min, D. Zhu, and M. Q.-H. Meng, "Robust and accurate 3d curve to surface registration with tangent and normal vectors," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 9953–9959.

[21] Z. Min, J. Liu, L. Liu, and M. Q.-H. Meng, "Generalized coherent point drift with multi-variate gaussian distribution and watson distribution," *IEEE Robotics and Automation Letters*, 2021.

[22] Z. Min, J. Wang, J. Pan, and M. Q.-H. Meng, "Generalized 3-d point set registration with hybrid mixture models for computer-assisted orthopedic surgery: From isotropic to anisotropic positional error," *IEEE Transactions on Automation Science and Engineering*, 2020.

[23] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 652–660.

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

[25] 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.

[26] Y. Aoki, H. Goforth, R. A. Srivatsan, and S. Lucey, "Pointnetlk: Robust & efficient point cloud registration using pointnet," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 7163–7172.

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

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

[29] 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.

[30] 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.

[31] 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.