

# SCIENTIFIC REPORTS



OPEN

## Accurate and Robust Non-rigid Point Set Registration using Student's-t Mixture Model with Prior Probability Modeling

Zhiyong Zhou<sup>1</sup>, Jianfei Tu<sup>2</sup>, Chen Geng<sup>1</sup>, Jisu Hu<sup>1</sup>, Baotong Tong<sup>1</sup>, Jiansong Ji<sup>2</sup> & Yakang Dai<sup>1</sup>

A new accurate and robust non-rigid point set registration method, named DSMM, is proposed for non-rigid point set registration in the presence of significant amounts of missing correspondences and outliers. The key idea of this algorithm is to consider the relationship between the point sets as random variables and model the prior probabilities via Dirichlet distribution. We assign the various prior probabilities of each point to its correspondences in the Student's-t mixture model. We later incorporate the local spatial representation of the point sets by representing the posterior probabilities in a linear smoothing filter and get closed-form mixture proportions, leading to a computationally efficient registration algorithm comparing to other Student's-t mixture model based methods. Finally, by introducing the hidden random variables in the Bayesian framework, we propose a general mixture model family for generalizing the mixture-model-based point set registration, where the existing methods can be considered as members of the proposed family. We evaluate DSMM and other state-of-the-art finite mixture models based point set registration algorithms on both artificial point set and various 2D and 3D point sets, where DSMM demonstrates its statistical accuracy and robustness, outperforming the competing algorithms.

Registration of point sets is of great importance in many computer vision tasks, such as medical image registration, image analysis, computer graphics, and pattern recognition. Many problems in these fields can be solved by point set registration algorithms operating on points or landmarks extracted from the input images. In medical image analysis, point set registration is necessary to match points or landmarks in 3D images for disease diagnosis, motion models, point-set-based image registration, image fusion, and construction of image atlases. In this paper, we focus on the registration model which is a key problem in non-rigid point set registration.

Mathematically, the point set registration problem can be described as follows. Let  $X \in IR^D$  and  $Y \in IR^D$  be two point sets to be registered, where  $X$  is a  $D$ -dimensional point set labeled as the target set, while  $Y$  is a  $D$ -dimensional point set labeled as the template set. The general approach of point set registration is to estimate a mapping  $T$  from  $IR^D$  to  $IR^D$ , which yields the best transformation between the target set  $X$  and the template set  $Y$ .

Many algorithms have been proposed for point set registration. Iterative Closest Point (ICP) algorithm<sup>1</sup> is the most popular algorithm owing to its low computation complexity. The traditional ICP algorithm defines the correspondences based on a closest neighbor principle. ICP finds a closest point  $y_i$  in  $Y$  for each point  $x_i$  in  $X$ . It subsequently estimates a transformation which best aligns  $X$  to  $Y$  by using a least-squares method. ICP iterates the cycle of correspondences establishment and alignment until it reaches the local minimum. However, the traditional ICP requires the initial position of the two point sets to be adequately close. ICP has been received a lot of attentions and several improved algorithms have been proposed<sup>2-5</sup>. Liu reviewed the improvements over ICP<sup>6</sup>.

Instead of aligning a one-to-one correspondence based on a closest distance criterion, the Robust Point Matching (RPM) algorithm<sup>7</sup> proposed by Gold *et al.* and its variants<sup>8,9</sup>, alternatively estimate soft-assignment of correspondences and transformation, leading to allowing for fuzzy correspondences, and<sup>9</sup> subsequently used Thin-Plate-Spline (TPS) to re-parameterize the transformation that resulted into the TPS-RPM algorithm. Tsin and Kanade<sup>10</sup> proposed a kernel-correlation-based point set registration approach, considering the non-rigid

<sup>1</sup>Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou, 215163, China.

<sup>2</sup>Lishui Central Hospital, Lishui, 323000, China. Correspondence and requests for materials should be addressed to J.J. (email: [jstcty@sina.com](mailto:jstcty@sina.com)) or Y.D. (email: [daiyk@sibet.ac.cn](mailto:daiyk@sibet.ac.cn))

point set registration as an alignment between two distributions. This approach parameterizes the point sets using explicit TPS parameterizations, which is equivalent to a regularization of second order derivatives of the transformation. Their algorithm attempts to align the given two point sets without explicitly estimating the correspondences, leading to a more robust algorithm against degeneration (such as missing correspondences and outliers).

Chui *et al.*<sup>11</sup> pointed out that the processing of alternative correspondence estimate and transformation in the RPM algorithm is equivalent to the Expectation Maximization (EM) framework for Gaussian mixture model (GMM), in which one point set is considered as GMM centroids and the other one is considered as data<sup>12</sup>. GMM is a well-known mixture model, widely used to formulate non-rigid point set registration as it is a natural and simple way to describe the given point sets. Revow *et al.*<sup>13</sup> represented the contour-like point sets using splines and modeled them by the probabilistic GMM formulation, where GMM centroids were uniformly positioned along the contours. This algorithm allows non-rigid transformation for point sets. Similar to<sup>9</sup>, Myronenko *et al.*<sup>14</sup> proposed a robust point set registration framework. Myronenko *et al.*<sup>15</sup> later introduced the Coherent Point Drift (CPD) algorithm, which enforced the points drift coherently by regularizing the transformation following the Motion Coherence Theory (MCT)<sup>16,17</sup>. The major difference between the two algorithms proposed in<sup>9,15</sup> is that<sup>9</sup> re-parameterizes the transformation via TPS, while<sup>15</sup> re-parameterizes the transformation by using Gaussian radial basis functions (GRBF). However, the CPD algorithm aligns a same mixture proportion for all mixture components and introduce an additional uniform distribution in mixture model for improving robustness against outliers, noise and occlusion<sup>18</sup>. Jian and Vemuri<sup>19</sup> modeled both point sets using GMM and introduced a general robust framework involving the minimization of the  $L_2$  distance between Gaussian mixtures. Tustison *et al.*<sup>20</sup> also represented point sets by using a GMM with an anisotropy covariance. In addition, features such as mutual information<sup>21</sup> and shape<sup>22–24</sup> extracted from images or point sets are incorporated into point set registration. Wang *et al.*<sup>25</sup> generalized a  $L_2$  divergence and obtained closed-form solutions for registration. Subsequently, Wang *et al.*<sup>26</sup> used a similar model to simultaneously align multiple point sets. However, it is well known that the GMM-based non-rigid point set algorithms are sensitive to significant amounts of outliers and missing correspondences since they use an additional component to represent the heavy tail of the mixture model<sup>27</sup>.

There are also several algorithms that attempt to align two point sets using the Student's-t mixture model (SMM) to improve the accuracy and robustness against outliers and missing correspondences. SMM has been introduced as an alternative to GMM, providing an effective and non-heuristic mean to handle degradations such as missing correspondences and outliers<sup>28</sup>. It is worth to point out that, mathematically, the Student's-t distribution corresponds to a Gaussian distribution when the degree of freedom (DoF)  $\gamma \rightarrow \infty$ , making the Gaussian mixture model be a special case of the Student's-t mixture model<sup>27</sup>. The Student's-t mixture model has heavily tails, leading to a natural and elegant model for modeling the given point sets with degradations<sup>29</sup>. Gerogiannis *et al.*<sup>30,31</sup> proposed a SMM-based rigid point set registration algorithm which was more robust than the GMM-based algorithms. However, it is regretful that the proposed algorithm is limited to rigid point set registration. In previous work, we introduced a SMM-based non-rigid point set registration method (called pSMM in this paper) for contour-like and surface-like point sets<sup>32</sup>, subsequently, we apply it for matching surface-like points<sup>33</sup>. Unfortunately, pSMM utilized EM framework to directly calculate the prior probability, which is a least-square-based method for fitting parameters, whose lack of robustness is well known. Moreover, it is an arduous task to get closed-form solutions for the SMM-based non-rigid point set registration in the EM framework<sup>34</sup>. To overcome this problem, Peel and McLachlan considered<sup>34</sup> SMM as an infinite mixture model of the scaled GMM integral form to get the closed-form solutions in EM framework<sup>35,36</sup>. Liu and Rubin indicated that convergence of estimating parameters of SMM in EM framework is slow, they subsequently extended the EM framework in the form of ECM and ECME algorithms<sup>37,38</sup>. Recently, the Student's-t distribution and the Student's-t mixture model also demonstrate their accuracy and robustness against outliers in various applications, such as data cluster<sup>39,40</sup>, data classification<sup>41</sup>, and image segmentation<sup>42–44</sup>. However, the prior distribution of SMM does not depend on the given point sets and the a same mixture proportion is assigned to all data in the existing approaches<sup>29,31,40</sup>. Additionally, the existing point set registration approaches do not take into account the local spatial representation of the input point sets. In order to overcome the lack of local spatial representation, Ma *et al.*<sup>45</sup> introduced a novel transformation estimation method using  $L_2E$  estimator for building robust sparse and dense correspondences. Some feature descriptors, such as shape context, are utilized for establish rough correspondences in their work. Ma *et al.*<sup>46</sup> considered point set registration as the estimation of a mixture of density, where the local feature is used to assign the membership probability of the mixture model.

In this paper, we proposed a more accurate and robust non-rigid point set algorithm, called DSMM, by using Dirichlet distribution in the Student's-t mixture model to formulate the various mixture proportion and assign them to corresponding mixture components, instead the same value in the existing methods. Comparing with the existing state-of-the-art point set registration algorithms (include pSMM), the key contributions of our work are: (1) We introduce the idea of considering the mixture component label vector as random variables, which is a major difference from the existing point set registration, where the mixture proportions are considered as discrete labels. We consequently utilize the Dirichlet distribution as a natural model for formulating the mixture proportion in the Student's-t mixture model, and assign various mixture proportion  $w_{mm}$  for each observation  $x_m$  belonging to corresponding component  $y_n$ . It is worth to point out that the main difference between DSMM and pSMM is that pSMM mathematically use a least-squared method to estimate the prior probabilities, while DSMM utilizes an Dirichlet distribution for modeling it, which is detailed in subsection 2.2. (2) We further propose a general mixture model family for point set registration based on the hidden variables in the Bayesian framework, which reveals the relationship of DSMM and the existing methods in subsection 2.3. We consider the Student's-t mixture model as infinite mixture of scaled Gaussian mixture model as Peel and McLachlan did<sup>34</sup>, and subsequently parameterize the hidden variables using Dirichlet distribution. (3) In order to incorporate the local spatial relationship between neighboring points, we further formulate the mixture proportions by the parameters of Dirichlet distribution by representing the posterior probabilities in a linear smoothing filter.

The rest of this paper is organized as follows. In the section 2, we present the main idea of the Dirichlet distribution for modeling the mixture proportions of the Student's-t mixture model, and further propose a general mixture model family for point set registration, where DSMM and existing approaches can be considered as its member. Section 3 contains some qualitative and quantitative evaluations on 2D and 3D point sets with outliers and missing correspondences. Finally, we present a discussion in section 4 and a conclusion in section 5.

## Method

**Student's-t mixture model for registration.** In this section, we start with briefly reviewing our previous work on point set registration based on Student's-t mixture model<sup>32</sup>. Let  $X_{M \times D} = (x_1, \dots, x_M)^T$  denotes a  $D$ -dimension point set considered as an observation,  $Y_{N \times D} = (y_1, \dots, y_N)^T$  denotes the other  $D$ -dimension point set. Each point  $y_n$  is considered as a component of the Student's-t mixture model. The probability density function of the Student's-t mixture model with  $N$  components is defined as

$$f(x_m|y_n, w_n, \sigma^2, \gamma_n) = \sum_{n=1}^N w_n S(x_m|y_n, \sigma^2, \gamma_n) \quad (1)$$

where  $w_n$  is a prior probability (mixture proportion) for  $y_n$ , satisfying the following constraint

$$0 < w_n < 1, \quad \sum_{n=1}^N w_n = 1 \quad (2)$$

$S(x_m|y_n, \sigma^2, \gamma_n)$  represents a probability density of multivariate Student's-t distribution, which takes the form

$$S(x_m|y_n, \sigma^2, \gamma_n) = \frac{\Gamma((\gamma_n + D)/2)}{\sigma(\gamma_n \sqrt{\pi})^{D/2} \Gamma(\gamma_n/2) (1 + d(x_m, y_n, \sigma^2)/\gamma_n)^{(D+\gamma_n)/2}} \quad (3)$$

In the Eq. (3),  $d(x_m, y_n, \sigma^2) = (x_m - y_n)^T (x_m - y_n) / \sigma^2$  is the Mahalanobis squared distance between  $x_m$  and  $y_n$ , and  $\Gamma(\cdot)$  is a Gamma function. In our registration method, each Student's-t distribution  $S(x_m|y_n, \sigma^2, \gamma_n)$ , which is called a component of the mixture model, has its own parameter set  $\Theta_n = \{y_n, \sigma^2, \gamma_n\}$  with its component centroid  $y_n$ , variance  $\sigma^2$  (or precision  $1/\sigma^2$ ) and degree of freedom  $\gamma_n$ .

Mathematically, the multivariate Student's-t distribution is equivalent to Gaussian distribution when its  $\gamma \rightarrow \infty$ . The Student's-t distribution provides a heavy-tailed model for fit the degradations such as data with longer than normal tails, outliers, and missing correspondences.

**Prior probability modeling with Dirichlet distribution.** The prior probability  $w_n$  in the Eq. (1) represents the mixture proportions of the  $n$ -th component in the mixture model. Unfortunately, in the previous work<sup>14,15,47</sup>, the mixture proportion  $w_n$  is assigned to all correspondences, which is unreasonable as the observations vary in their locations. Moreover, the existing methods estimate the prior probabilities via a least-squared-based method in the EM framework, leading a well-known under-fitting problem for complex point set registration. Another limitation is that each observation is considered as an independent point to its neighbors. Therefore, these methods do not take into account the spatial correlation between the neighboring points in the decision process. In order to overcome the under-fitting problem and improve the robustness to noise, outliers and occlusion, we introduce Dirichlet distribution for modeling the prior probabilities and assign different prior probabilities between the observations and their correspondences.

Firstly, we rewrite the density function of Student's-t mixture model at an observation  $x_m$ , which takes the form

$$f(x_m|y_n, w_n, \sigma^2, \gamma_n) = \sum_{n=1}^N w_{mn} S(x_m|y_n, \sigma^2, \gamma_n) \quad (4)$$

Specially, the parameter  $w_{mn}$  denotes the mixture proportion of the component  $y_n$  belonging to its correspondence  $x_m$ .

Secondly, we introduce the hidden variables<sup>27</sup> in the Bayesian approach to model the prior probabilities in our method. In the Bayesian approach, the complete-data vector, which composes of the hidden variables, is given by

$$v_c = (y_1, \dots, y_N, z_1, \dots, z_N, u_1, \dots, u_N)^T \quad (5)$$

where the discrete label  $z_n = (z_{1n}, \dots, z_{Mn})^T$  denotes the component label vector, which defines the relationships between  $x_m$  and  $y_n$  ( $n = 1, \dots, N$ ;  $m = 1, \dots, M$ ).  $z_{mn}$  is 1 or 0 depending on whether  $x_m$  belongs to the  $n$ -th component

$$z_{mn} = \begin{cases} 1 & x_m \text{ belongs to } n\text{-th component} \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

$u_1, \dots, u_N$  represent the hidden variables associated with the scaling weights of the covariance of the equivalent Gaussian distributions, which is defined as

$$u_n|z_{mn}=1 \sim f_{\Gamma}(\gamma_n/2, \gamma_n/2) \quad (7)$$

where  $f_{\Gamma}(x)$  is the Gamma function. According to the Eq. (7),  $u_1, \dots, u_N$  are independent variables if  $z_1, \dots, z_N$  are given. Consequently,  $x_m$  is a random variable defined as<sup>34</sup>

$$x_m|u_n, z_{mn}=1 \sim f_N(y_n, \sigma^2/u_n) \tag{8}$$

where  $f_N(y_n, \sigma^2/u_n)$  is a Gaussian distribution with the mean  $y_n$  and the covariance  $\sigma^2/u_n$ . We now focus on the hidden variable  $z_n$ , which is considered as an independent variable in pSMM. We now consider  $z_n = (z_{1n}, \dots, z_{Mn})$  as a probable label vector and formulate it by Dirichlet distribution and Dirichlet law<sup>44,48,49</sup> for accurately modeling the prior probabilities. Dirichlet distribution is a natural and power method for modeling complex data by varying its parameters.

According to<sup>29</sup>, we get the conditional probability of the probability label  $z_n$

$$p(z_n|\alpha_n) = \int_0^1 p(z_n|\xi_n)p(\xi_n|\alpha_n)d\xi_n \tag{9}$$

where  $\xi_n = \{\xi_{1n}, \dots, \xi_{Mn}\}$  ( $m = 1, \dots, M$ ) is the Dirichlet parameter in the  $M$ -dimensional probability simplex, satisfying  $0 < \xi_{mn} < 1$  and  $\sum_{m=1}^M \xi_{mn} = 1$ ; and  $\alpha_n = \{\alpha_{1n}, \dots, \alpha_{Mn}\}$ , satisfying  $0 < \alpha_{mn} < 1$ , is the vector of the Dirichlet parameters.  $p(z_n|\xi_n)$  and  $p(\xi_n|\alpha_n)$  take the form of

$$p(z_n|\xi_n) = \frac{\prod_{m=1}^M (\xi_{mn})^{z_{mn}}}{\prod_{m=1}^M (z_{mn})!} \tag{10}$$

$$p(\xi_n|\alpha_n) = \frac{\Gamma(\sum_{m=1}^M \alpha_{mn})}{\prod_{m=1}^M \Gamma(\alpha_{mn})} \prod_{n=1}^N (\xi_{mn})^{\alpha_{mn}-1} \tag{11}$$

Combining the Eqs (9), (10) and (11), the probability label subsequently takes the form

$$\begin{aligned} p(z_n|\alpha_n) &= \int_0^1 \frac{\prod_{n=1}^N (\xi_{mn})^{z_{mn}} \Gamma(\sum_{n=1}^N \alpha_{mn})}{\prod_{n=1}^N (z_{mn})! \prod_{n=1}^N \Gamma(\alpha_{mn})} \prod_{n=1}^N (\xi_{mn})^{\alpha_{mn}-1} d\xi_n \\ &= \frac{1}{\prod_{n=1}^N (z_{mn})! \prod_{n=1}^N \Gamma(\alpha_{mn})} \frac{\prod_{n=1}^N \Gamma(\alpha_{mn} + z_{mn})}{\Gamma(\sum_{n=1}^N (\alpha_{mn} + z_{mn}))} \\ &\times \int_0^1 \frac{\Gamma(\sum_{n=1}^N (\alpha_{mn} + z_{mn}))}{\prod_{n=1}^N \Gamma(\alpha_{mn} + z_{mn})} \prod_{n=1}^N (\xi_{mn})^{\alpha_{mn} + z_{mn} - 1} d\xi_m \end{aligned} \tag{12}$$

According to the property of the probability density function,  $p(\xi_m|\alpha_m)$  always satisfies the following condition

$$\int_0^1 p(\xi_m|\alpha_m)d\xi_m = \int_0^1 \frac{\Gamma(\sum_{n=1}^N \alpha_{mn})}{\prod_{n=1}^N \Gamma(\alpha_{mn})} \prod_{n=1}^N \xi_{mn}^{\alpha_{mn}-1} d\xi_m = 1 \tag{13}$$

Utilizing the Eq. (13) to rewrite the Eq. (12), we could obtain the probability

$$p(z_m|\alpha_m) = \frac{1}{\prod_{n=1}^N (z_{mn})! \prod_{n=1}^N \Gamma(\alpha_{mn})} \frac{\prod_{n=1}^N \Gamma(\alpha_{mn} + z_{mn})}{\Gamma(\sum_{n=1}^N (\alpha_{mn} + z_{mn}))}. \tag{14}$$

We now consider the condition of discrete label  $z_{mn}$  in the Eq. (6). Considering  $\Gamma(x + 1) = x\Gamma(x)$ , the closed-form solution of prior probability  $w_{mn}$  is finally given by

$$w_{mn} = p(z_{mn} = 1|\alpha_m) = \alpha_{mn} / \sum_{n=1}^N \alpha_{mn} \tag{15}$$

However, the components in the mixture model are still assumed to be independent identically distributed, which brings an attendant trouble that there is no neighborhood information for registration process since  $x_n$  is considered as an independent point to its neighbors. In order to solve the problem, we constraint the Dirichlet distribution with local spatial representation via defining parameter  $\alpha_{mn}$  of the Dirichlet distribution as<sup>42</sup>

$$\alpha_{mn} = \exp\left(\frac{\bar{\alpha}}{N_n} \sum_{y_i \in \partial y_n} p(y_i; x_m)\right) = \exp\left(\frac{\bar{\alpha}}{N_n} \sum_{y_i \in \partial y_n} p_{mn}\right) \tag{16}$$

where  $p_{mn}$  is a posterior probability, which is formulated as

$$p_{mn} = \frac{w_{mn}S(x_m; y_n, \sigma^2, \gamma_m)}{\sum_{n=1}^N w_{mn}S(x_m; y_n, \sigma^2, \gamma_m)} \tag{17}$$

$N_n$  stands for the number of neighbors locating in the window around the point  $y_n$ , and  $y_i \in \partial y_n$  represents that  $y_i$  locates in the neighborhood of the given point  $y_n$ .  $\bar{\alpha}$  is a local spatial constraint coefficient of the Dirichlet distribution.  $\alpha_{mn}$  contains the neighborhood information that makes registration has a spatial constraint. Moreover,

only a parameter in the EM framework need to be calculated, not  $M \times N$  parameters  $\alpha_{mn}$  in the traditional Student's-t distribution mixture model, leading our method to be a computationally effective algorithm. We finally accurately model the prior probability  $w_{mn}$  and incorporate the local spatial constraint in a simple way. Combining the Eqs (15) and (16),  $w_{mn}$  gets its closed-form as

$$w_{mn} = \frac{\exp\left(\frac{\bar{\alpha}}{N_n} \sum_{y_i \in \mathcal{Y}_n} p_{mn}^{m_i}\right)}{\sum_{n=1}^N \exp\left(\frac{\bar{\alpha}}{N_n} \sum_{y_i \in \mathcal{Y}_n} p_{mn}^{m_i}\right)} \tag{18}$$

In order to get a solution of  $\bar{\alpha}$ , we separate  $w_{mn}$  from the probability density function (4) and estimate it by minimizing the negative log-likelihood function equivalently.

$$E(\Psi) = \sum_{n=1}^N w_{mn} S(x_m | y_n, \sigma^2, \gamma_n) = E(w_{mn}) + E(\gamma_n) + E(y_n, \sigma^2) \tag{19}$$

We obtain the iterative solution of  $\bar{\alpha}$  by minimizing  $E(w_{mn})$ , or equivalently solve the following equation

$$\sum_{n=1}^N \sum_{m=1}^M p_{mn} \left( \frac{\sum_{y_i \in \mathcal{Y}_n} p_{mi}}{N_n} - \frac{\sum_{m'=1}^M \left( \left( \frac{1}{N_n} \sum_{y_i \in \mathcal{Y}_n} p_{m'i} \right) \exp\left(\frac{\bar{\alpha}}{N_n} \sum_{y_i \in \mathcal{Y}_n} p_{m'i}\right) \right)}{\sum_{m'=1}^M \exp\left(\frac{\bar{\alpha}}{N_n} \sum_{y_i \in \mathcal{Y}_n} p_{m'i}\right)} \right) = 0 \tag{20}$$

Comparing to the mathematical expressions in the MRF method<sup>50</sup>, we find a connection between the proposed Dirichlet-based spatial representation and the MRF method. The energy function  $U_{\text{MRF}}$  in the MRF method in<sup>50</sup> may degenerate to the prior probability in the Eq. (18) of our method if  $U_{\text{MRF}}$  is set up as a diagonal matrix with the diagonal as  $-1$ , which implies that the Dirichlet distribution models the prior probabilities by using a spatial clustering method. A limitation of the previous methods is that they consider each point is independent to its neighbors, which results to the lack of a spatial correlation between the neighboring points.

The parameter set of non-rigid point set registration is defined as  $\Psi = (w_1, \dots, w_M, \gamma_1, \dots, \gamma_M, y_1, \dots, y_M, \sigma^2)$ , where  $w_n = (w_{1n}, \dots, w_{Mn})$  represents the prior probability, whose solution has been discussed above. We subsequently separate parameters of SMM and estimate them by maximizing their log-likelihood, or by minimizing the negative log-likelihood function equivalently for calculating other parameters in the EM method. We now briefly reviews the solution of these parameter, which is detailed in our previous work<sup>32,33</sup>. Firstly, we consider the Eq. (19),  $u_{mn}$  can be calculated from the equation

$$u_{mn} = E_{\Psi}(u_n; x_m, z_{mn} = 1) = \frac{\gamma_n + D}{\gamma_n + d(x_m, y_n, \sigma^2)} \tag{21}$$

The solution of  $\gamma_n$  of  $k$ -iteration could be obtained by minimizing  $E(\gamma_n)$ . The iteration of  $\gamma_n$  is given by

$$-\varphi\left(\frac{\gamma_n}{2}\right) + \ln \frac{\gamma_n}{2} + 1 + \frac{\sum_{n=1}^M p_{mn} (\ln u_{mn} - u_{mn})}{\sum_{n=1}^M p_{mn}} + \varphi\left(\frac{\gamma_n^{(k-1)} + D}{2}\right) - \ln \frac{\gamma_n^{(k-1)} + D}{2} = 0 \tag{22}$$

where  $\gamma^{(k-1)}$  is an optimization solution in  $(k-1)$ -iteration.

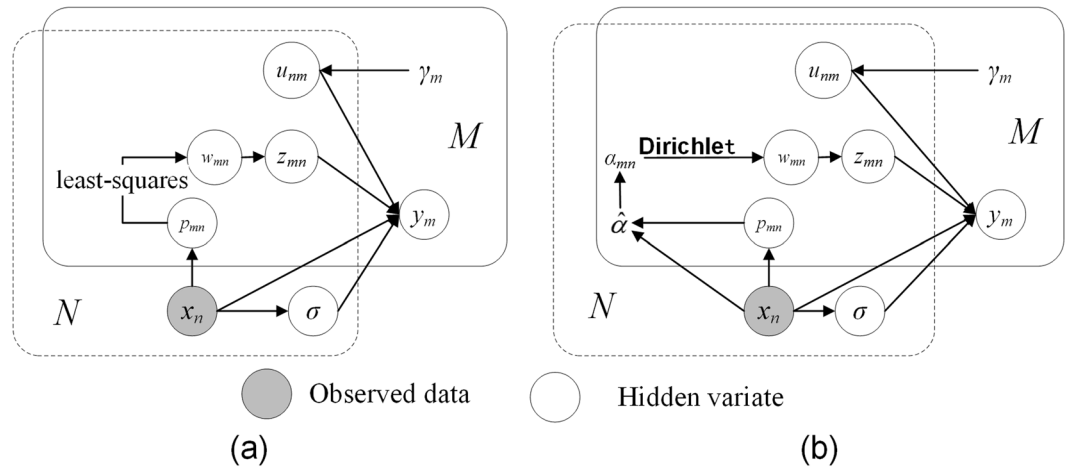
Finally, we calculate the transformation field  $Y^{15}$  as  $Y^{(k)} = Y^{(k-1)} + GW^{(k)}$ , where  $G_{M \times M}$  is a Gaussian kernel matrix with it element  $g_{ij} = \exp(-|y_i - y_j| / (2\beta)^2)$  in order to reduce the oscillating energy at high frequency.  $\beta$  is a width of smoothing Gaussian filter, defining the model of the smoothness regularization.  $G(m;)$  is the column vector of the kernel matrix  $G_{M \times M}$ , and  $W_{M \times D}$  is the weight matrix of  $G_{M \times M}$ . Using  $\partial E(y_n, \sigma^2) / \partial W = 0$ ,  $W$  is given by

$$W = (\text{diag}(\hat{P}\mathbf{I})G + \lambda(\sigma^2)^{(k-1)}\mathbf{I})^{-1}(\hat{P}X - \text{diag}(\hat{P}\mathbf{I})Y^{(k-1)}) \tag{23}$$

where  $\hat{P}$  is a  $M \times N$  matrix with its element  $\hat{p}_{mn} = p_{mn} u_{mn}$ , denoting the posterior probability density corrected by  $u_{mn}$ .  $\mathbf{I}$  is a column vector of all ones;  $\mathbf{I}$  is an identity matrix;  $\text{diag}(\cdot)$  denotes a diagonal matrix.  $\lambda$  represents the trade-off between the goodness of maximum likelihood fit and regularization. Using  $\partial E(y_n, \sigma^2) / \partial (\sigma^2) = 0$ ,  $\sigma^2$  is formulated as

$$\sigma^2 = \frac{\sum_{n=1}^N \sum_{m=1}^M p_{mn} u_{mn} \|x_m - y_n - G(m, \bullet)W\|^2}{D \sum_{n=1}^N \sum_{m=1}^M p_{mn} u_{mn}} \tag{24}$$

Generally speaking, the main advantages of our method are: (1) We model the prior probabilities by Dirichlet distribution in EM framework, which more accurately represents the mixture proportion of each component in the mixture model, leading an excellent method for degenerated point set registration. In the existing methods (includes pSMM), the prior probability  $w_m$  is directly estimated as  $w_m = \sum_{n=1}^N p_{mn} / N$  in  $M$ -step of EM method. Mathematically, it is a least-squared solution, which gives a point estimate to the prior probabilities and disregards the remaining uncertainty in the estimation. Therefore, low robustness and under-fitting problem are introduced into the process of optimization. Dirichlet distribution and its mixture model could automatically determine the number of necessary mixture components based on the data<sup>51</sup>. In DSMM, we introduce the Dirichlet distribution



**Figure 1.** Representation of the Student’s-t mixture model with Dirichlet distribution as a directed graphical model. The  $M$ -box denotes the  $M$  observations  $x_m$ , and the  $N$ -box denotes the  $N$  mixture components. Note that the random variable  $u_{nm}$ ,  $w_{nm}$ , and  $z_{nm}$  belong to both  $M$ -box and  $N$ -box, indicating that they are corresponding random variables for the Student’s-t mixture components and the observations. (a) pSMM for point set registration, (b) DSMM for point set registration.

for modeling the prior probabilities, and then assign various mixture proportions (prior probabilities)  $w_{nm}$  of  $n$ -th component to its  $m$ -th correspondence. Rather than taking a point estimate, we model the prior probabilities using Dirichlet distribution, where Dirichlet distribution gives the posterior probability distribution over all model parameters in E-step of  $(k + 1)$  iteration by using the observed data together with the prior distributions. Subsequently, we utilize these posterior probability distributions to estimate the prior probabilities in M-step. In general, comparing to a least-square-based estimation, the estimate of the prior probabilities via Dirichlet distribution could yield a robust and stable result, by including the resulting uncertainty into the estimation. (2) We incorporate the local spatial relationship between neighboring points into the Dirichlet distribution parameters in a simple and natural way by representing their posterior probabilities in a linear smoothing filter, leading to taking into consideration of the spatial correction in the registration process. Furthermore, it potentially supplies a universal approach to incorporate more ingenious filters for local spatial representation in the mixture model<sup>52</sup>.

In order to summarize the proposed method and theoretically reveal the differences between DSMM and pSMM, we represent the joint distribution of all random variables in our method via a directed graph model, as show in Fig. 1. Moreover, we will further quantitatively estimate performance of DSMM, pSMM and other competitive method in the following experiments, which will more intuitively reflect the power of modeling prior probabilities via Dirichlet distribution.

**Family of the mixture-model-based registration.** We tooe an interesting observation that the mixture-model-based registration methods (included the proposed method) can be generally modeled as infinite Gaussian mixture models at a single observation  $x$  for potential outliers or data with longer than normal tails, which takes the form

$$\int f_N(x|y, \sigma^2/u)dH(u). \tag{25}$$

Now considering an  $N$ -component mixture for point set registration, a general mixture model family for registration is given by

$$f(x_m|y_n, \Psi) = \sum_{m=1}^M w_{nm} \left( \int f_N(x_m|y_n, \sigma^2/u_n)dH(u_n) \right) \tag{26}$$

where  $f_N$  is a general symbol for denoting a probability density function of Gaussian.

We now assume that  $H$  is a chi-squared distribution with the degree of freedom  $\gamma_n$  and its random variable  $u_n \sim (u|\alpha, \beta) = \alpha^\beta u^\alpha e^{-\beta u}$ , where  $G(u|\alpha, \beta)$  is a symbol of Gamma distribution. In our method, we choose  $\alpha = \beta = \gamma_n/2$ . According to<sup>27,34</sup>, it is obvious that we can rewrite Student’s-t distribution as an infinite mixture of scaled Gaussian mixture model. Therefore, we conclude that the Student’s-t mixture model is a member of the general mixture model family.

We subsequently simplify the infinite mixture to a finite mixture with two different components by placing the mass  $\varepsilon$  at the point  $u_n = 1$  and mass  $(1 - \varepsilon)$  at the point  $u_n = 1/c$ . The Eq. (25) therefore transforms to a Gaussian scaled mixture that takes the form as

$$\varepsilon f_N^1(x_m|y_n, c\sigma^2) + (1 - \varepsilon) f_N^2(x_m|y_n, \sigma^2) \tag{27}$$

$\sum_{m=1}^M w_{mn} \left( \int \phi(x_m   y_n, \sigma^2   u_n) dH(u_n) \right)$				
	$w_{mn}$	$\phi$	$H$	$u$
DSMM	$p(z_{mn} = 1   \xi_n) p(\xi_n   \alpha_n)$	Gaussian mixture	chi-squared	$u \sim (u   \gamma_n/2, \gamma_n/2)$
SMM	$p(z_{mn} = 1   x_m) = \{0, 1\}$	Gaussian mixture	chi-squared	$u \sim (u   \gamma_n/2, \gamma_n/2)$
PR-GLS	$\tau/ I $ or $(1-\tau)/(N-I)$	Gaussian mixture	discrete value	constant
CPD	$1/M$	Gaussian mixture	discrete value	constant
RPM-RBF	$1/M$	GRBF	discrete value	constant
RPM-TPS	$1/M$	thin plate spline	discrete value	constant

**Table 1.** Relationship between the general family of mixture model registration and the existing methods.

where  $f_N(x|y, \sigma^2)$  denotes the Gaussian distribution with its mean  $y$  and variance  $\sigma^2$ ;  $\varepsilon$  is a small value, representing the small proportion of observation in the mixture and  $c$  is a relatively large value for representing the potential degeneration that has a relatively large variance. In the two components mixture, the first term denotes the probability density of potential degeneration, while the second term denotes the probability density of normal data. Comparing to the Student's-t mixture model, the major limitation of Gaussian scaled mixture is lack of robustness to degeneration due to its additional Gaussian components to capture the tail of the distribution, as shown in the Eq. (27).

We further simplify the Gaussian scaled mixture model. We now assume that  $\phi_1$  is a uniform distribution, which is given by  $\phi_1 = N/M$ ;  $\phi_2$  is a Gaussian distribution, and simultaneously fix  $w_{mn}$  as a constant, satisfying  $w_{mn} = 1/M$ . The Eq. (26) finally transforms to  $\varepsilon/N + (1-\varepsilon)(f_N(x|y, \sigma^2))/M$ , which takes the same form as CPD. It is obviously to find that CPD is a member of the large family, which is formulated by the Eq. (26). Moreover, it is worthy to point out that RPM-based registration methods, such as RPM-TPS and RPM-RBF are mathematically equivalent to CPD in the EM framework, which leads RPM-based methods to be members of the mixture model family. Theoretically, the discrete latent variable  $z_{mn}$  specifies which component of the Student's-t mixture model generates the observation  $x_n$ , and the continuous latent variable  $u_{mn}$  specifies the scaling of the corresponding equivalent Gaussian distribution. Consequently, pSMM will transform to CPD if  $z_{mn} = 1$ ,  $u_{mn} = 1$ , and  $\gamma_n \rightarrow \infty$  simultaneously. The degrees of freedom  $\gamma_n$  is a trade-off between robustness and efficiency. A small DoF  $\gamma_n$  can appropriately assign a small weight to the outliers or missing correspondences depending on the input data, while a relative larger value of DoF tends to fit a Gaussian mixture model to the data. Actually, the degree of freedom reflects the assumption on the amount of noise in the point sets, which plays an important role in point matching. For the initialization of the degrees of freedom, we always use the value 1 (multivariate Student's-t distribution reduces to Cauchy distribution when  $\gamma = 1$ ) to maximize the robustness at the beginning of registration process.

In the existing methods, the major disadvantage is that the parameter  $z_{mn}$  is considered as a discrete label  $z_{mn} = \{0, 1\}$ . Another limitation of the existing mixture-model-based method is their under-fitting for prior probabilities. It is easily understood by recalling the maximization of prior probabilities in the EM framework. The estimation of prior probabilities in these methods mathematically is a least-square solution, leading to a well-known under-fitting problem. In order to get a more precise model, we consider the label  $z_{mn}$  as a random variable following a multinomial distribution with its probability vector  $\xi_n = \{\xi_{1n}, \dots, \xi_{Mn}\}$ . According to the multinomial definition, the conditional distribution takes the form as

$$p(z_n | \xi_n) = \frac{K!}{\prod_{m=1}^M (z_{mn})!} \prod_{m=1}^M (\xi_{mn})^{z_{mn}} \quad (28)$$

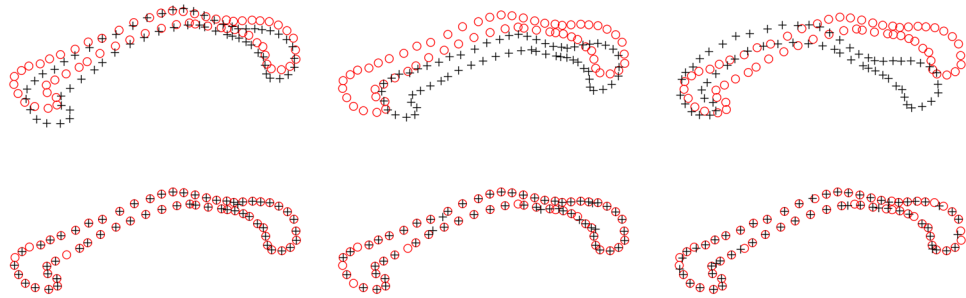
where  $\xi_{mn} > 0$  and  $\sum_n \xi_{mn} = 1$ . The multinomial model represents the probability  $\xi_{mn}$  of observation  $x_m$  belonging to the component  $y_n$  with  $K$  realizations, satisfying  $K = \sum_{m=1}^M z_{mn}$ . When the multinomial distribution is used to generate the correspondences, the distribution of the number of emissions (i.e., counts) of an individual component follows a binomial law<sup>53,54</sup>

$$p(z_{mn} | \xi_{mn}) = \binom{K}{z_{mn}} (\xi_{mn})^{z_{mn}} (1 - \xi_{mn})^{K - z_{mn}}. \quad (29)$$

The above equation reveals that it is a small probability to a point corresponding to multi-component under the multinomial model, since the count of a single point corresponding to components decays exponentially. A better approach is hierarchical: the probabilities of correspondences between point  $x_m$  and component  $y_n$  is generated by multinomial, whose parameters are formulated by Dirichlet distribution, which is also called Dirichlet compound multinomial<sup>55</sup>. As discussed in the subsection 2.2, we finally formulate the mixture proportion by using parameters of Dirichlet distribution. Jian *et al.*<sup>19</sup> revealed the relationship between point set registration methods from the view of the divergence function.

Generally, we generalize a family of mixture-model-based point set registration from the view of hidden variables in the Bayesian framework, and summarize a relationship between DSMM and the existing mixture-model methods in the Table 1.

**Data availability statement.** All data was obtained from public data collections, including dir-lab (<https://www.dir-lab.com/index.html>) and ADNI (<http://www.adni-info.org/>), all these database allow researches reproduce their images and data.



**Figure 2.** Performance of DSMM on 2D Corpus Callosum data. The target points are denoted as red “o” and the template points are denoted as black “+”. Each point set contains 63 points respectively. We align the black “+” to the red “o” by using DSMM. The point sets before registration is superimposed in the top row, and the performance of DSMM is shown in the bottom row.

**Ethical approval.** All data used in our experiments are from public image bases, and permit researches use images for algorithm research. All clinical data has been approved by the Medical Ethics Committee of Lishui Central Hospital and Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Science, and has been allowed for carrying out experiments.

## Results

In this section, we qualitatively and quantitatively evaluate DSMM on various point sets, such as artificial data, points extracted from various medical images, and points from surface scan models. These point sets have various shapes, including 2D contour-like point sets, 3D cloud-like and surface-like point sets. In order to show the performance of our method, we compare DSMM with other state-of-the-art non-rigid point set registration (PR-GLS<sup>46</sup>, pSMM<sup>32</sup>, GMM-L2<sup>19</sup>, CPD<sup>15</sup>, RPM-TPS<sup>9</sup>, and its variety RPM-RBF) in the following evaluations. The performance of DSMM and pSMM will be intuitively shown in these evaluations. It is worth to point out that we directly perform DSMM on all point sets without any preprocessing (including rigid registration initialization), except data normalization. We only simply set  $\beta = 2$ , initial value of DoF  $\gamma = 1$  in all tests, which is also a reflection of the robustness of DSMM.

**Qualitative evaluations.** We firstly demonstrate the qualitative evaluation of DSMM on 2D contour-like point sets. Specifically, Fig. 2 shows three examples of 2D contour-like Corpus Callosum (CC), which are from <http://www.nitrc.org/>. Each point set contains 63 points extracted from outer contour of CC in brain MR images of several normal subjects. The top row in Fig. 2 shows three pairs of Corpus Callosum point sets before registration and figures on bottom row show the performance of DSMM.

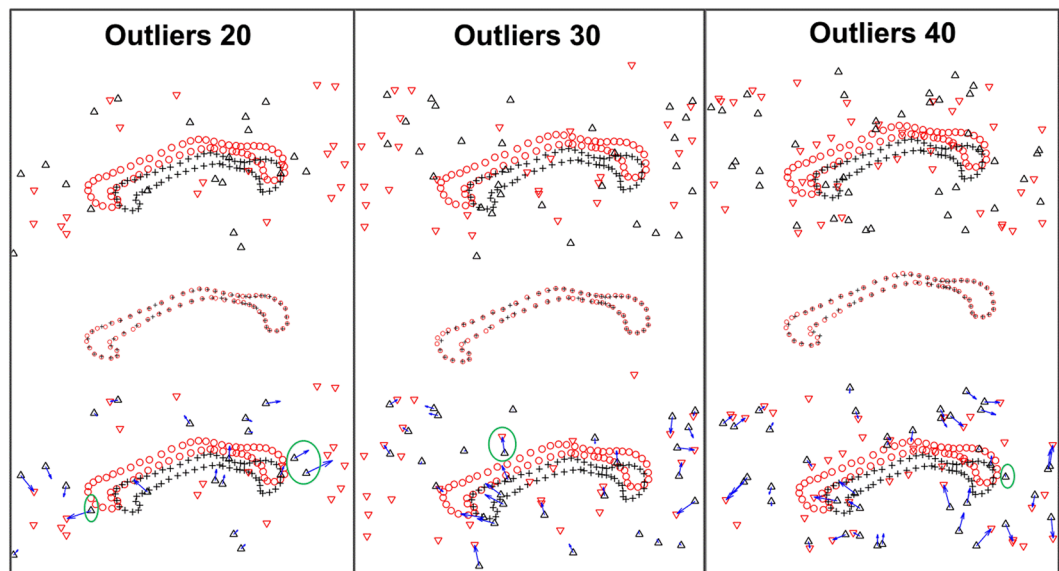
We add various numbers of additional random outliers with uniform distribution. Examples of such point sets (with additional 32%, 48% and 63% outliers) are respectively shown in the top row of Fig. 3. The middle row shows the final registration results, which demonstrates the data points accurately match to their correspondences, resisting the impact of the outliers. In order to intuitively show the displacement vector of outliers, we overlap the warp of outlier on the point sets before registration, which demonstrates the transformation maps the most outliers to the sound positions, except few points who are much closed to the data points.

**Quantitative evaluations.** We perform quantitative evaluations on 2D counter-like datasets, 3D cloud-like and 3D surface-like datasets for DSMM and other competing non-rigid point set registration algorithms. To take quantitative evaluations, we use the mean 3D Euclidean magnitude distance and standard deviation between correspondences as a statistical measure. In the quantitative evaluations, we show the performance of DSMM, PR-GLS<sup>46</sup>, pSMM<sup>32</sup>, GMM-L2<sup>19</sup>, CPD<sup>15</sup>, RPM-TPS<sup>9</sup>, and its variety RPM-RBF. Comparing to the existing methods, the major difference of our method is that DSMM models the prior probabilities by using Dirichlet distribution and assigns the various prior probability values for components, while the existing methods estimate a prior probability by a least-squared solution. PR-GLS assigns the membership probability  $w_{mn}$  based on shape context feature, so that the local structure information can also be used to achieve good performance.

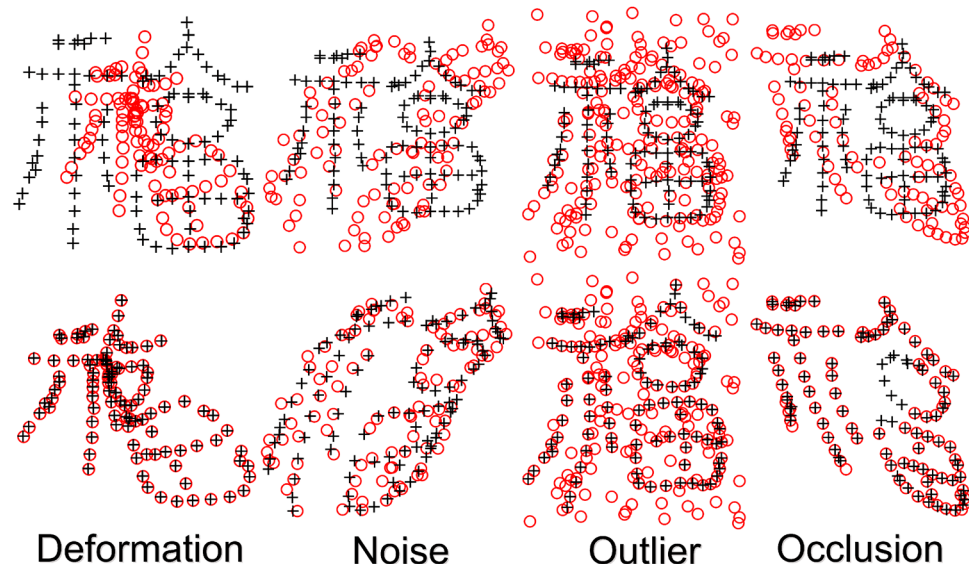
We perform the first quantitative evaluation on 2D Chinese characters<sup>46</sup> with deformation, noise, outliers and occlusion (the ratio of noise, outliers, and occlusion is from 10 to 50%). Each point set contains 105 normal points. The superimposed points of Chinese character are respectively shown in the top row of Fig. 4. The goal of our experiments are to align the template points (black “+”) to their correspondences in the red point set (red “o”). The performance of DSMM seems good, which accurately and robustly matches the correspondences. The registration results are intuitively shown in the bottom of Fig. 4. Figure 5 shows the statistical registration results of DSMM and other completing methods. The y axis of bar in Fig. 5 indicates the mean registration error of each method, where a small error value indicates a good performance. We break the one-to-one correspondence by add noise, outliers, and removing points in these datasets. Benefitting from Dirichlet distribution and Student’s-t mixture model, the statistical results show that DSMM performs the best results, which are slightly better than PR-GLS and significantly better than other five methods.

The second quantitative evaluation is performed on 20 samples of real 3D cloud-like lung datasets with 10 point sets extracted from thoracic 4D CT images<sup>55,56</sup> and the other 10 point sets extracted from COPD images<sup>56</sup>,



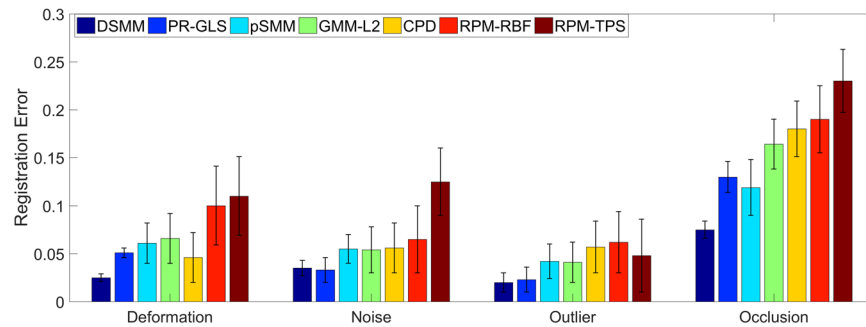


**Figure 3.** Performance of DSMM on 2D Corpus Callosum data with outliers. The data has been break up one-to-one correspondence by adding different number of additional uniform distribution outliers both in the template set and the target set. In the top row, the red “o” represents the data points in the target set, and the black “+” represents the correspondence in the template sets. For clarity, we denote outliers in the target set with red “▽”, and the outliers in the template set with black “△”. The figures of middle row show the transformations that map all data points in the template set to their correspondences, resisting the influence of the outliers in the target set. In the bottom row, we overlap the transformations of outliers on the template set on the degenerated points before alignment, demonstrating that our method has an ability to handle most outlier, except few points very closed to the black “+”. The results intuitively show that DSMM is accurate and robust against significant number of outliers.

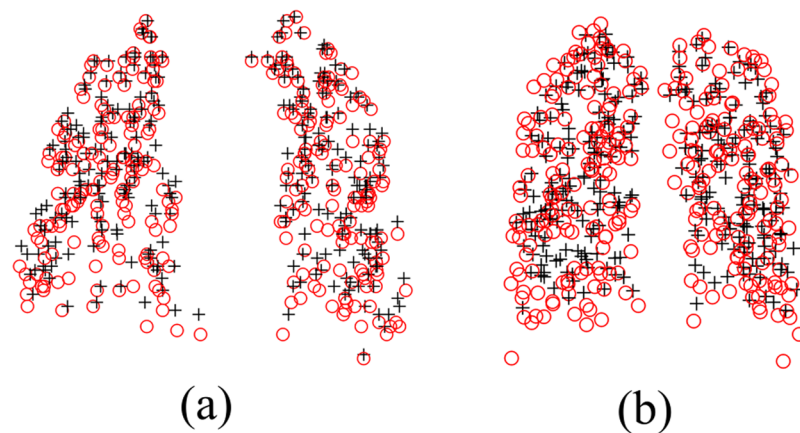


**Figure 4.** Performance of DSMM on 2D Chinese character shapes with deformation, noise, outlier and occlusion. We align the template point set (black “+”) to the target set (red “o”). For each sample, the figures in the top row show superimposed points of template point set and the target data before registration, and the figures in bottom show the registration results.

shown in Fig. 6. Each sample has a pair of 3D lung point set, one is identified from the maximum inhalation phase image and the other is identified from the maximum exhalation phase image. Each 3D lung point set respectively has 300 points, which are selected by experts to make the two point sets correspond to each other. It is a herculean task for non-rigid registration algorithms to match cloud-like point sets accurately due to lack of topological structures or geometry structures in such data. Table 2 demonstrates the mean 3D Euclidean distance between correspondences for each point set before registration. Figures 7 and 8 respectively show the performance of



**Figure 5.** Comparison of DSMM with PR-GLS, SMM, GMM-L2, CPD, RPM-RBF and RPM-TPS on Chinese characters with deformation, noise outliers and occlusion. The y axis of error bar indicates the statistical registration error, where a small value reveals a good performance.



**Figure 6.** Examples of 3D cloud-like point sets extracted from thoracic 4D CT images. Each subject contains a pair of 3D lung point sets with 300 points both in template set and target set respectively. The target point sets are extracted from images of the maximum inhalation phase and the template point set are extracted from images of the maximum exhalation phase. We denote the target points with red “o” and the template ones black “+”. (a) An example of the 4D CT lung dataset used in the quantitative evaluation. (b) An example of the COPD dataset.

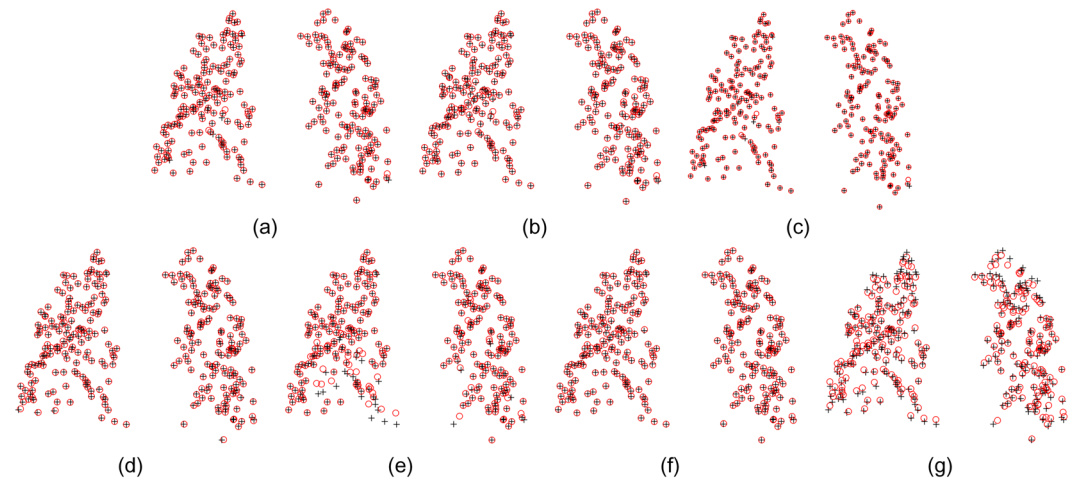
DSMM and other non-rigid point set algorithms on samples of datasets from 4D CT and COPD. From the statistical measure shown in Figs 9, 10 and Table 3, we can intuitively see the performance of DSMM is better than other mixture-model-based algorithms on real 3D cloud-like point sets.

In order to evaluate the performance of our method on the various distortion, we then perform the third quantitative evaluation on 4D CT point sets identified from thoracic 4D CT images<sup>56</sup>. Each 4D CT point set consists of six expiratory phases (T00, T10, T20, T30, T40 and T50) and there are 75 points (a subset of the point set containing 300 points) in each sample. The T00 point sets are identified from the maximum inhalation phase images, and the T50 point sets are identified from the maximum exhalation phase images. The T10, T20, T30, and T40 point sets are respectively extracted from the expiratory phase images between the maximum exhalation phase and the maximum exhalation phase. As shown in Fig. 10, the red “o” denotes the point in T00 image, the black + denotes the point in T10~T50 images. We show transformation vectors between correspondences in Fig. 10. Table 4. demonstrates the performance of DSMM on T00 and T50 of each subject.

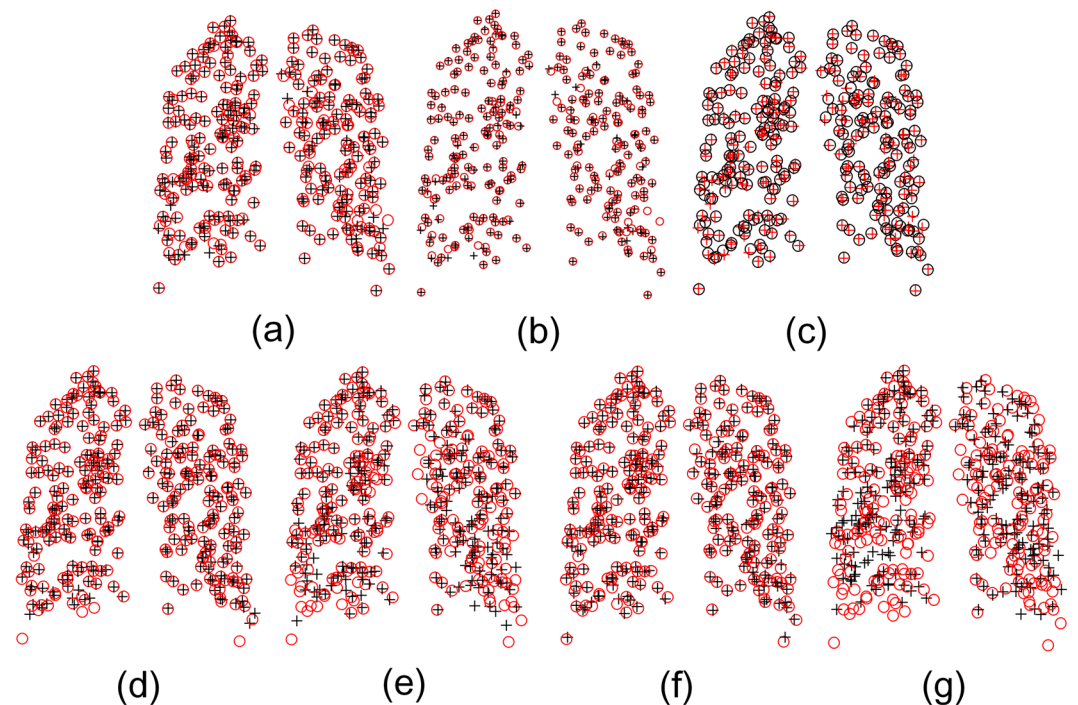
We further test the ability of our algorithm to handle outliers and missing correspondences in the subsequent evaluation on the point sets from 4D CT images. In order to break up the one-to-one correspondence between the given point sets and add missing correspondences, we randomly delete the increasing number of points both in the target point sets and template points sets, as shown in the top row of Fig. 11. In the first subfigure, we do not delete any point, while in other subfigures, we respectively remove 15, 30, 45, 60, and 75 points both in the target set and the template set, which means only 270, 240, 210, 180, and 150 correspondences existing in figure (b)~(f). In order to explicitly reveal the outliers, we use red “∇” for denoting the outliers in the target set, whose correspondences having been removed in the template sets, and use black “Δ” for denoting the outliers in the template set. Subsequently, we test DSMM and other algorithms on these pairs of incomplete samples. Figure 11 shows the performance of our method on these incomplete data. For clarity, we only show the correspondences in the result subfigures, which clearly shows that only few points diverge from the ground truth even though 75 points are removed in the data sets. In the evident from Fig. 11, our method shows its excellent performance in the presence

	1	2	3	4	5	6	7	8	9	10
4D CT	4.01 ± 2.91	4.65 ± 4.09	6.73 ± 4.21	9.42 ± 4.81	7.10 ± 5.14	11.10 ± 6.98	11.59 ± 7.87	15.26 ± 9.11	7.82 ± 3.99	7.63 ± 6.54
COPD	25.90 ± 11.57	21.77 ± 6.46	12.29 ± 6.39	30.90 ± 13.49	30.90 ± 14.05	28.32 ± 9.20	21.66 ± 7.66	25.57 ± 13.61	14.84 ± 10.01	22.48 ± 10.64

**Table 2.** Mean 3D Euclidean magnitude distance and standard deviation (unit: mm) for all subjects of 3D point sets before registration.

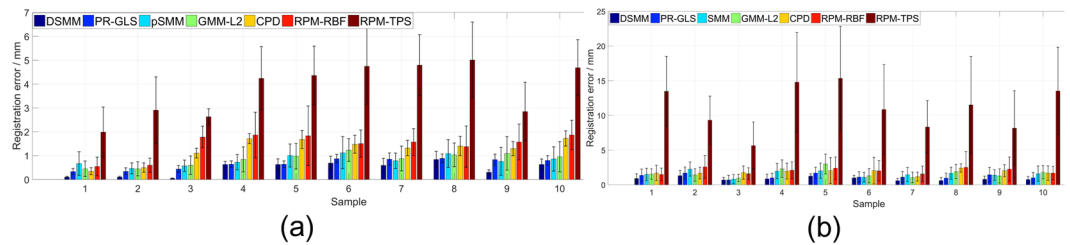


**Figure 7.** Performance of non-rigid point set registration algorithms on the point set extracted from 4D CT, which is shown in Fig. 6(a). (a) DSMM, (b) PR-GLS, (c) pSMM, (d) GMM-L2, (e) CPD, (f) RBF-RPM, (g) TPS-RPM.

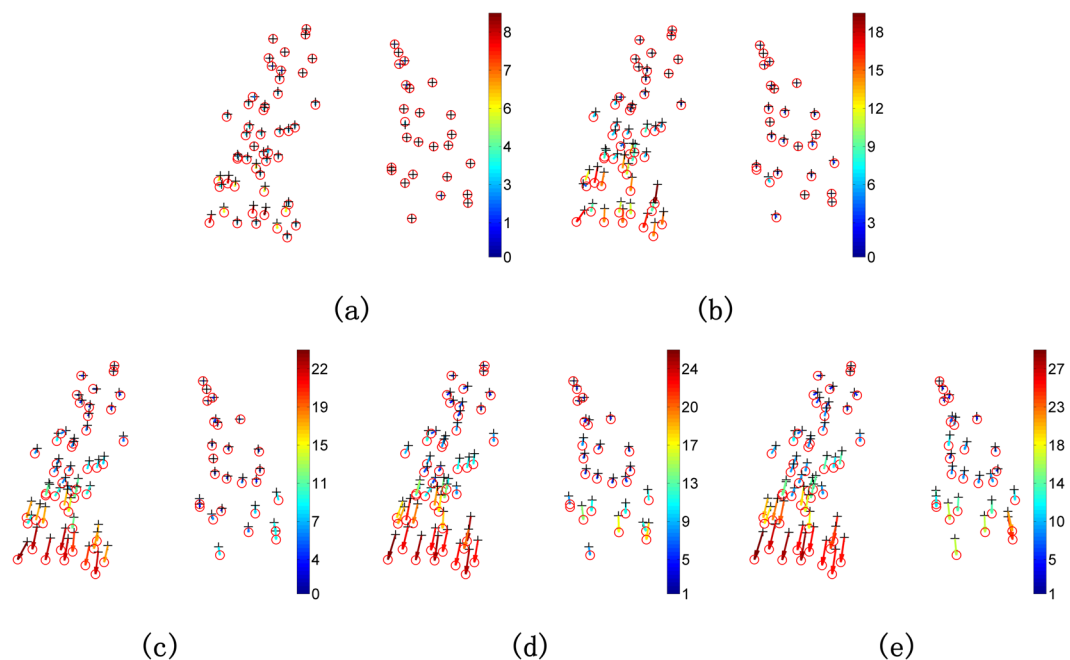


**Figure 8.** Performance of non-rigid point set registration algorithms on the point set extracted from COPD, which is shown in Fig. 6(b). (a–f) show the final registration results of DSMM, PR-GLS, pSMM, GMM-L2, CPD, RBF-RPM, and TPS-RPM respectively.

of significant amounts of missing correspondences and outliers due to the local spatial representations and the prior probability modeling of each component in the mixture model. Figure 12(a–e) respectively show the mean



**Figure 9.** Quantitative evaluations of non-rigid point set registration algorithms on twenty 3D lung point sets, in which 10 samples are extracted from 4D CT and the other 10 samples are extracted from COPD. Mean 3D Euclidean magnitude distance between correspondences is used as a statistical measure for quantitative evaluations. The mean distance of each sample before registration is demonstrated in Table 3. (a) Performance of the non-rigid point set registration algorithms on 3D point sets extracted from thoracic 4D CT images. (b) Performance of each algorithm on the dataset extracted from COPD images. DSMM outperforms other competing algorithms both on the datasets from thoracic 4D CT images and COPD images.



**Figure 10.** Performance of DSMM on a 4D point set from 4D CT for thoracic motion modeling. 75 points are identified from each expiratory phase image (T00, T10, T20, T30, T40 and T50), where T00 are from the maximum inhalation phase, and T50 are from the exhalation phase image. We also denote the points from the maximum inhalation phase with red “o” and points from other phase with black “+”. (a) Performance of DSMM on a point set between T00 and T10, (b) T00 and T20, (c) T00 and T30, (d) T00 and T40, (e) T00 and T50.

3D Euclidean magnitude distance between correspondences for different algorithms on the incomplete data sets, which indicates the statistical accuracy and robustness of our method.

Finally, we conduct the last quantitative experiment for matching 3D surface-like “wolf” shapes. Each point set typically contains about 5000 points, and there is absence between template point and target points. We show only 1600 points in the top row of Fig. 13 for clarity. In order to evaluate the robustness of DSMM on occlusion and outliers, we remove about 25 percentage of total number of points for representing occlusion, and add about 25 percentage of total number of points for representing outliers, which are respectively shown in the middle and right columns. The figures in top row of Fig. 13 show the superimposed points before registration, and the bottom figures show the matching results of DSMM. Figure 14 shows quantitative comparisons of DSMM and other competitive methods on wolf data, where DSMM performs the best results on ideal data and degeneration data.

## Discussion

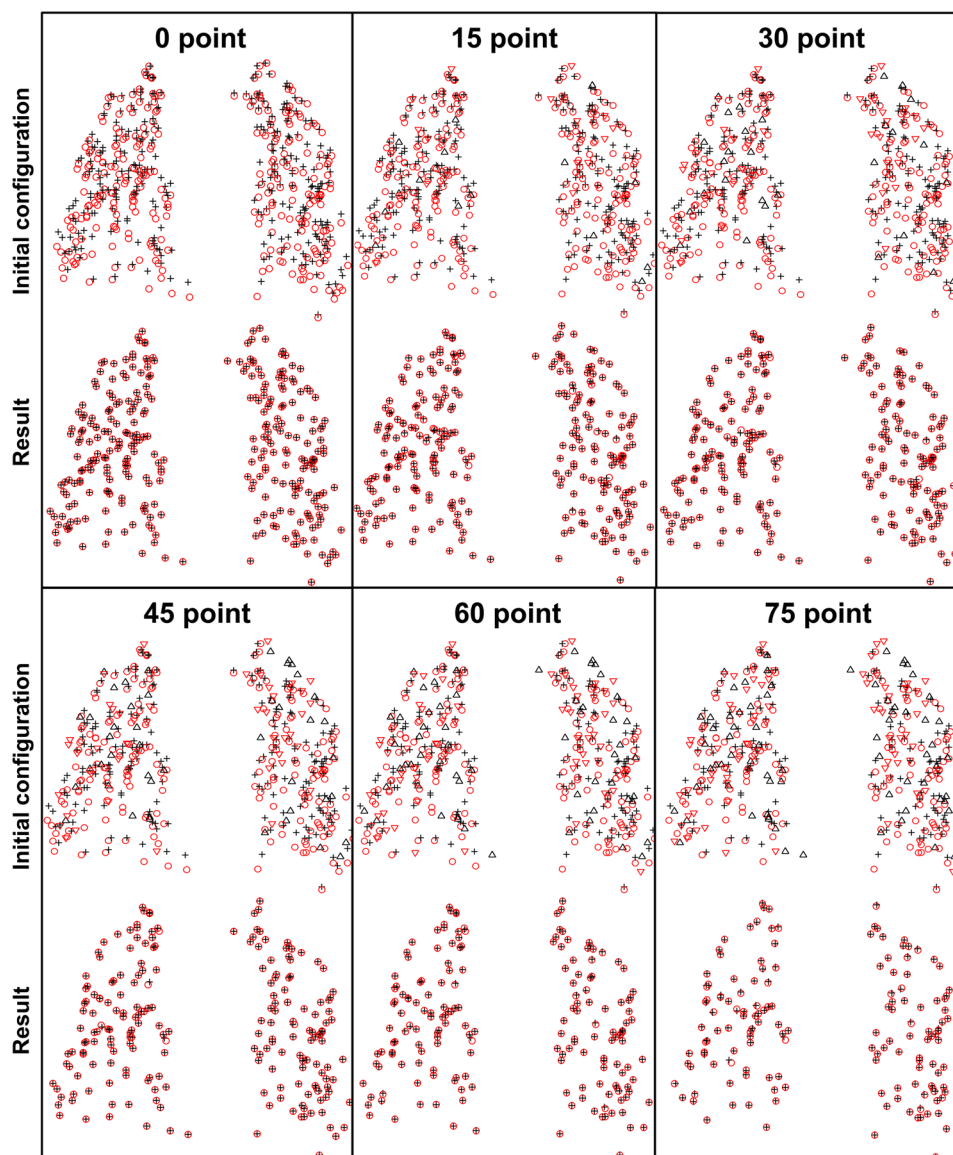
Point set registration is a key problem in various applications. We focus on the model of point set registration which is a core point that has been received sustaining attentions in the recent years. In this work, we introduce a SMM-based non-rigid point set registration approach, named DSMM, which models the prior probabilities by using Dirichlet distribution and Dirichlet law. The main motion of our method is that we want

	DSMM	PR-GLS	pSMM	GMM-L2	CPD	RBF-RPM	TPS-RPM	before warp
4D CT	<b>0.455 ± 0.169</b>	0.685 ± 0.173	0.792 ± 0.444	0.849 ± 0.490	1.205 ± 0.285	1.449 ± 0.671	3.620 ± 1.219	8.52 ± 5.56
COPD	<b>0.917 ± 0.4766</b>	1.326 ± 0.7966	1.576 ± 0.9214	1.660 ± 0.9630	1.852 ± 1.0993	2.017 ± 1.3862	11.099 ± 5.5439	23.46 ± 10.31

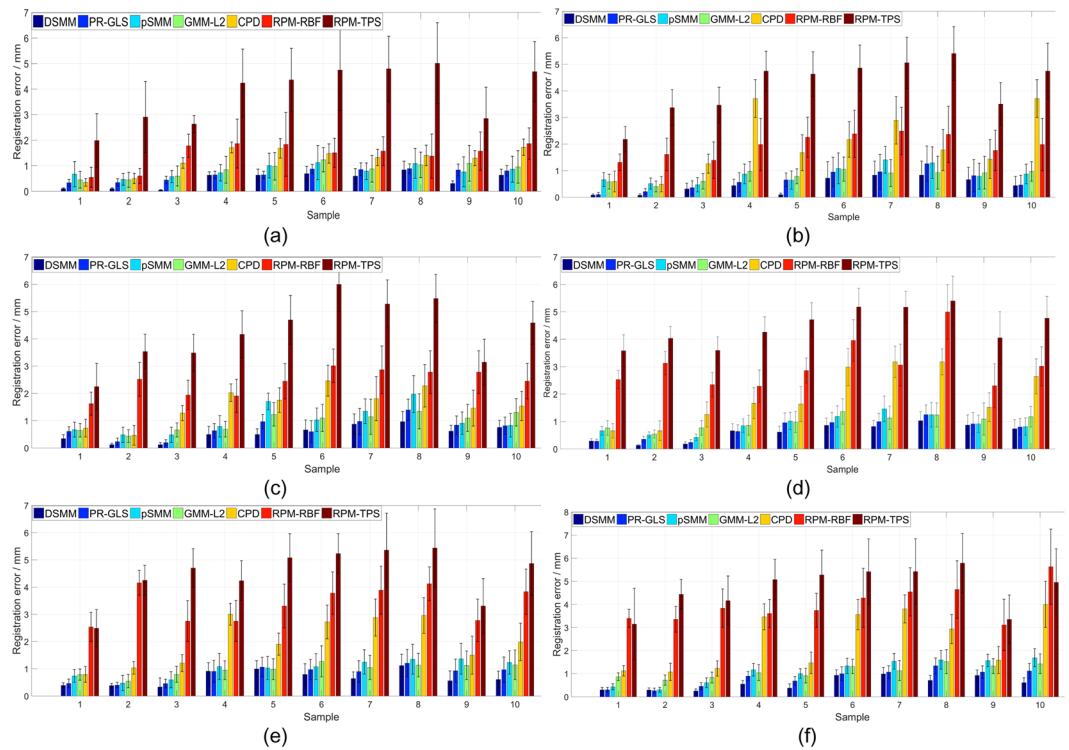
**Table 3.** Mean 3D Euclidean magnitude distance (unit: mm) for all 3D point sets after registration.

	1	2	3	4	5	6	7	8	9	10
before	3.89 ± 2.78	4.34 ± 3.90	6.94 ± 4.05	9.83 ± 4.86	7.48 ± 5.51	10.89 ± 6.97	11.03 ± 7.43	15.00 ± 9.01	7.92 ± 3.98	7.30 ± 6.35
after	0.05 ± 0.03	0.04 ± 0.03	0.03 ± 0.02	0.04 ± 0.02	0.09 ± 0.04	0.28 ± 0.06	0.05 ± 0.02	0.36 ± 0.13	0.03 ± 0.02	0.04 ± 0.03

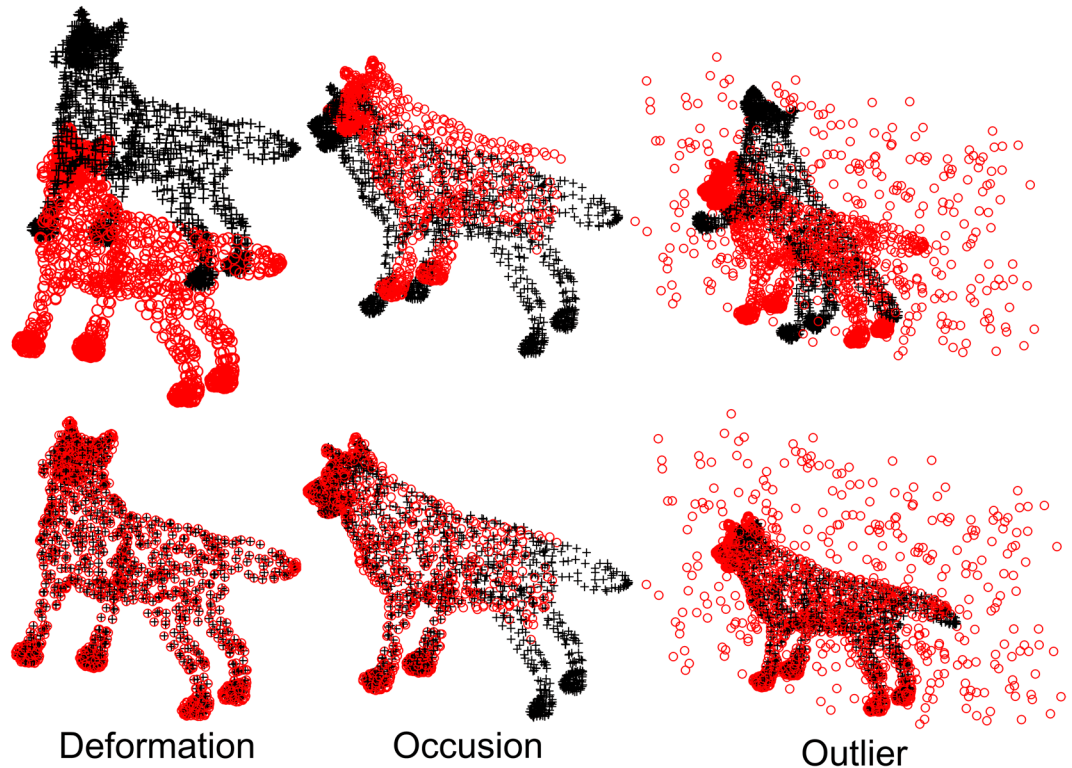
**Table 4.** Mean 3D Euclidean magnitude distance (unit: mm) between correspondence of T00 and T50 by using DSMM.



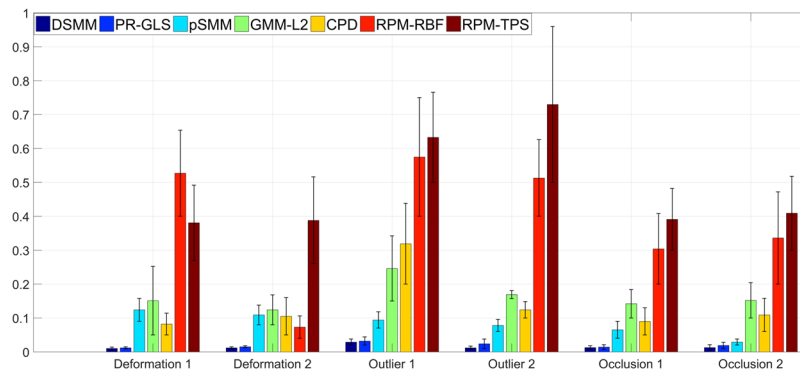
**Figure 11.** Performance of DSMM on the incomplete 3D lung data pairs identified from 4D CT images. We randomly remove increasing number of points both in the target point set and the template point set, which breaks up the one-to-one correspondence between the given data. For clarity, the red “∇” denote the missing correspondences, whose correspondences are removed in the template set, and the black “Δ” denotes the missing correspondences in the template set. Our method shows its stable performance on the missing correspondences and occlusion. The top image in the each subfigure shows the initial configuration of incomplete data, and the bottom ones show the result of our method.



**Figure 12.** Quantitative evaluations of the non-rigid point set registration algorithms on the incomplete 3D lung data pairs of 4D CT images. (a) complete data, (b) 15 points removed, (c) 30 points removed, (d) 45 points removed, (e) 60 points removed, (f) 75 points removed.



**Figure 13.** Preformation of DSMM on 3D wolf shapes with deformation, occlusion and outliers. We align the template points (black “+”) to the target points (red “o”) For each sample, the figures in the top row show template points and the target data before registration, and the figures in bottom show the registration results.



**Figure 14.** Performance of DSMM and other methods on wolf data.

to use a Bayesian framework to estimate the prior probabilities since the existing methods estimate them via a least-square method, which is a well-known method lack of robustness. Fortunately, Dirichlet distribution and its mixture models are fully Bayesian framework, which could automatically determine the model complexity (in terms of the total number of necessary mixture components) based on the data, not depend on any prior knowledge. Concretely speaking, we firstly consider the non-rigid point set registration as a probability density estimation, where one point set is represented as Student's-t mixture model centroids, the other one is represented as data set. The main advantage of multivariate Student's-t distribution is that it is heavily tailed than the Gaussian distribution, hence it is more robust against degradations than GMM. Secondly, we explicitly exploit Dirichlet distribution and Dirichlet law to incorporate the local spatial representation in the given point sets. We later assign various prior probability values of prior distribution depending on the input point sets, instead of the same value to all points, leading DSMM be more accurate than other existing algorithms. Thirdly, we formulate the SMM as an infinite scaled GMM integral form in order to obtain closed-form solutions. Subsequently, we iteratively fit the SMM centroids to the data set by using EM framework and estimate the posterior probabilities of centroids, which provides correspondence probabilities between the target point set and the template set. Finally, we calculate all registration parameters and transformation via the EM framework. We perform qualitative and quantitative evaluations for DSMM on various shapes. These evaluations intuitively indicate the favorable performance of DSMM.

## Conclusion

We have tested DSMM on various shape (2D contour-like, 3D cloud-like and 3D surface-like) point sets, and compared it with pSMM, PR-GLS, GMM-L2, CPD, RBF-RPM, and TPS-RPM. The results demonstrate that DSMM is robust against significant amount of missing correspondences and outliers, and is more accurate and robust than the other existing non-rigid point set registration methods.

## References

1. Besl, J. P. & McKay, H. D. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **14**, 239–256 (1992).
2. Ying, S., Peng, J., Du, S. & Qiao, H. A. Scale stretch method based on ICP for 3D data registration. *IEEE Transactions on Automation Science and Engineering*. **6**, 559–565 (2009).
3. Granger, S. & Pennec, X. Multi-scale EM-ICP: A fast and robust approach for surface registration. *7th European Conference on Computer Vision*. **2353**, 418–432 (2002).
4. Stewart, C. V., Tsai, C. L. & Roysam, B. The dual-bootstrap iterative closest point algorithm with application to retinal image registration. *IEEE Transactions on Medical Imaging*. **22**, 1379–1394 (2003).
5. Tsai, C. L., Li, C. Y. & Yang, G. The edge-driven dual-bootstrap iterative closest point algorithm for registration of multimodal fluorescein angiogram sequence. *IEEE Transactions on Medical Imaging*. **29**, 636–649 (2010).
6. Liu, Y. H. Improving ICP with easy implementation for free-form surface matching. *Pattern Recognition*. **37**, 211–226 (2004).
7. Gold, S., Rangarajan, A. & Lu, C. P. New algorithms for 2D and 3D point matching: pose estimation and correspondence. *Pattern Recognition*. **31**, 1019–1031 (1998).
8. Chui, H. L. & Rangarajan, A. A new algorithm for non-rigid point matching. *IEEE Conference on Computer Vision and Pattern Recognition*. **2**, 44–51 (2000).
9. Chui, H. L. & Rangarajan, A. A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding*. **89**, 114–141 (2003).
10. Tsai, Y. & Kanade, T. A Correlation-Based Approach to Robust Point Set Registration. *8th European Conference on Computer Vision*. **3023**, 558–569 (2004).
11. Chui, H. L. & Rangarajan, A. A feature registration framework using mixture models. *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*. 190–197 (2000).
12. Jian, B. & Vemuri, B. C. A robust algorithm for point set registration using mixture of Gaussians. *10th IEEE International Conference on Computer Vision*. 1246–1251 (2005).
13. Revow, M., Williams, C. K. I. & Hinton, G. E. Using generative models for handwritten digit recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **18**, 592–606 (1996).
14. Myronenko, A. & Song, X. Carreira-Perpinan M. Non-rigid point set registration: Coherent Point Drift. *Advances in neural information processing systems*. **19**, 1009–1016 (2006).
15. Myronenko, A. & Song, X. Point set registration: coherent point drift. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **32**, 2262–2275 (2010).
16. Yuille, A. L. & Grzywacz, N. M. The motion coherence theory. *2nd International Conference on Computer Vision*. 344–353 (1988).

17. Yuille, A. L. & Grzywacz, N. M. A mathematical analysis of the motion coherence theory. *International Journal of Computer Vision*. **3**, 155–175 (1989).
18. Ma, J. Y., Zhao, J., Tian, J. W., Yuille, A. L. & Tu, Z. W. Robust point matching via vector field consensus. *IEEE Transactions on Image Processing*. **23**, 1706–1721 (2014).
19. Jian, B. & Vemuri, B. C. Robust point set registration using gaussian mixture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **33**, 1633–1645 (2011).
20. Tustison, N. J., Awate, S. P. & Song, G. Point set registration using Havrda-Charvat-Tsallis entropy measures. *IEEE Transactions on Medical Imaging*. **30**, 451–460 (2011).
21. Rangarajan, A., Chui, H. & Duncan, J. S. Rigid point feature registration using mutual information. *Medical Image Analysis*. **3**, 425–440 (1999).
22. Taron, M., Paragios, N. & Jolly, M. P. Registration with uncertainties and statistical modeling of shapes with variable metric kernels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **31**, 99–113 (2009).
23. Tu, Z., Zheng, S. & Alan, Y. Shape matching and registration by data-driven EM. *Computer Vision and Image Understanding*. **109**, 290–304 (2008).
24. Belongie, S., Malik, J. & Puzicha, J. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **24**, 509–522 (2002).
25. Wang, F., Vemuri, B. & Syeda-Mahmood, T. Generalized L2-divergence and its application to shape alignment. *21st International Conference on Information Processing in Medical Imaging*. 227–238 (2009).
26. Wang, F., Vemuri, B. C. & Rangarajan, A. Simultaneous nonrigid registration of multiple point sets and atlas construction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **30**, 2011–2022 (2008).
27. Svensen, M. & Bishop, C. M. Robust bayesian mixture modelling. *Neuro computing*. **64**, 235–252 (2005).
28. Shoham, S. Robust clustering by deterministic agglomeration EM of mixtures of multivariate t-distributions. *Pattern Recognition*. **35**, 1127–1142 (2002).
29. Chatzis, S. P., Kosmopoulos, D. I. & Varvarigou, T. A. Robust sequential data modeling using an outlier tolerant hidden markov model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **31**, 1657–1669 (2009).
30. Gerogiannis, D., Nikou, C. & Likas, A. Robust image registration using mixtures of t-distributions. *11th IEEE International Conference on Computer Vision*. 2330–2337 (2007).
31. Gerogiannis, D., Nikou, C. & Likas, A. The mixtures of Student's t-distributions as a robust framework for rigid registration. *Image and Vision Computing*. **27**, 1285–1294 (2009).
32. Zhou, Z., Dai, Y., Zhou, Z. & Chen, S. Robust non-rigid point set registration using Student's-t mixture model. *PLoS ONE*. **9**, e91381, <https://doi.org/10.1371/journal.pone.0091381> (2014).
33. Zhou, Z. *et al.* Direct point-based registration for precise non-rigid surface matching using Student's-t mixture model. *Biomedical Signal Processing and Control*. **33**, 10–18 (2017).
34. Peel, D. & McLachlan, G. J. Robust mixture modelling using the t distribution. *Statistics and Computing*. **10**, 339–348 (2000).
35. Sfikas, G., Nikou, C. & Galatsanos, N. Edge preserving spatially varying mixtures for image segmentation. *IEEE Conference on Computer Vision and Pattern Recognition*. 1–7 (2008).
36. Liu, C. H. & Rubin, D. B. ML estimation of the t distribution using EM and its extensions, ECM and ECM. *Statistica Sinica*. **5**, 19–39 (1995).
37. Liu, C. H. ML Estimation of the Multivariate t Distribution and the EM Algorithm. *Journal of Multivariate Analysis*. **62**, 296–312 (1997).
38. Hawkins, S. D. M. & McLachlan, G. J. High-breakdown linear discriminant analysis. *Journal of the American Statistical Association*. **92**, 136–143 (1997).
39. Chatzis, S. & Varvarigou, T. Robust fuzzy clustering using mixtures of Student's-t distributions. *Pattern Recognition Letters*. **29**, 1901–1905 (2008).
40. Wang, Z. M., Song, Q. & Soh, Y. C. Robust curve clustering based multivariate t-distribution model. *IEEE transactions on neural networks*. **21**, 1976–1984 (2010).
41. Sfikas, G., Nikou, C., Galatsanos, N. & Heinrich, C. MR Brain tissue classification using an edge-preserving spatially variant bayesian mixture model. *11th International Conference on Medical Image Computing and Computer-Assisted Intervention*. **5241**, 43–50 (2008).
42. Thanh, M. N., Wu, Q. M. & Jonathan Robust Student's-t mixture model with spatial constraints and its application in medical image segmentation. *IEEE Transactions on Medical Imaging*. **31**, 103–116 (2012).
43. Sfikas, G., Nikou, C., Galatsanos, N. & Heinrich, C. Spatially varying mixtures incorporating line processes for image segmentation. *Journal of Mathematical Imaging and Vision*. **36**, 91–110 (2010).
44. Sfikas, G., Nikou, C. & Galatsanos, N. Robust image segmentation with mixtures of Student's t-distributions. *IEEE International Conference on Image Processing*. **1**, 273–276 (2007).
45. Ma, J. Y. *et al.* Robust L2E estimation of transformation for non-rigid registration. *IEEE Transactions on Signal Processing*. **63**, 1115–1129 (2015).
46. Ma, J. Y., Zhao, J. & Yuille, A. Non-rigid point set Registration by preserving Global and Local Structures. *IEEE transactions on image processing*. **25**, 53–64 (2016).
47. Bouguila, N. & Ziou, D. A hybrid SEM algorithm for high-dimensional unsupervised learning using a finite generalized Dirichlet mixture. *IEEE Transactions on Image Processing*. **15**, 2657–2668 (2006).
48. Bouguila, N., Ziou, D. & Vaillancourt, J. Unsupervised learning of a finite mixture model based on the Dirichlet distribution and its application. *IEEE Transactions on Image Processing*. **13**, 1533–1543 (2004).
49. Van, L. K., Maes, F. & Vandermeulen, D. Automated model-based tissue classification of MR images of the brain. *IEEE Transactions on Medical Imaging*. **18**, 897–908 (1999).
50. Sang, Q., Zhang, J. Z. & Yu, Z. Robust non-rigid point registration based on feature-dependant finite mixture model. *Pattern Recognition Letters*. **34**, 1557–1565 (2013).
51. Ma, Z. Y., Rana, P. K., Taghia, J., Flierl, M. & Leijon, A. Bayesian estimation of Dirichlet mixture model with variational inference. *Pattern Recognition*. **47**, 3143–3157 (2014).
52. Nikou, C., Likas, A. C. & Galatsanos, N. P. A bayesian framework for image segmentation with spatially varying mixtures. *IEEE Transactions on Image Processing*. **19**, 2278–2289 (2010).
53. Madsen, R. E., Kauchak, D. & Elkan, C. Modeling word burstiness using the Dirichlet distribution. *Proceedings of the 22nd international conference on Machine learning*. 545–552 (2005).
54. Edward, C., Richard, C., Josue, M., Maithili, S. & Thomas, G. Four-dimensional deformable image registration using trajectory modeling. *Physics in Medicine and Biology*. **55**, 305–327 (2010).
55. Castillo, R. *et al.* A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. *Physics in Medicine and Biology*. **54**, 1849–1870 (2009).
56. Castillo, R. *et al.* A reference dataset for deformable image registration spatial accuracy evaluation using the COPDgene study archive. *Physics in Medicine and Biology* **58**(9), 2861–2877 (2013).



## Acknowledgements

This study was funded by National Key Research and Development Plan (2017YFB1103602, 2017YFC0114304), NSFC grants (61501452), Jiangsu Society Development Program (BE2017675, BE2017664), Jiangsu Key Technology RD Program (BE2016010-3, BE2016010-4), Zhejiang Key Technology RD Program (2018C03024), Suzhou Industry Technological Innovation Projects (SYG201606), Foundation of Key Laboratory in Science and Technology Development Project of Suzhou (SYG201606, SZS201609, SYG201706, SYG201707), SND Medical Plan Project (2016Z010), and the Public Welfare Research Project of Zhejiang province (LGF18H160035).

## Author Contributions

Writing the article: Zhiyong Zhou and Yakang Dai; Revision of the article: Jiansong Ji, Jianfei Tu; Data analysis and statistical expertise: Chen Geng, Jisu Hu; Data collection: Baotong Tong; Obtaining funding: Zhiyong Zhou, Yakang Dai.

## Additional Information

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018