Robust Point Set Matching Under Variational Bayesian Framework

Han-Bing Qu, Ji-Chao Li, Jia-Qiang Wang  
Key Laboratory of Pattern Recognition  
Beijing Academy of Science and Technology, China  
{quhanbing,wangjiaqiangbj}@gmail.com,ljch@163.com

Lin Xiang, Hai-Jun Tao  
College of Information Engineering  
China Jiliang University, Hangzhou, China  
{xianlin,hjtao}@cjlu.edu.cn

Abstract—In this paper, we formulate the probabilistic point set matching problem under variational Bayesian framework and propose an iterative algorithm in which the posterior distributions for random variables are updated in sequent until a local optimum is reached. The variational Bayesian point set matching algorithm explicitly accounts for matching uncertainty with respect to parameters and variables and is thus less prone to local optima in comparison with other matching algorithms based on point estimation methods. Furthermore, an extra Gaussian mixtures model is used for the estimation of outlier distribution and the proposed framework demonstrates favorable robust performance in terms of spurious outliers comparing with other matching algorithms.

I. INTRODUCTION

Point set matching is an important and well-studied problem with many applications in computer vision and other fields. Example applications includes medical image registration, the alignment of a set of biometric images and face alignment. There is a long tradition of algorithms for finding an optimal alignment between two sets of points [1], [2], [3]. There are mainly two classes of point matching/registration methods with (1) known correspondence,(2) unknown correspondence. Although the prior finding of correspondence between two point sets is attractive for solving complicated matching problems[4], [5], [6], these algorithms is not very helpful in the case of second class, such as scanned data registration[2], protein analysis[7], the alignment of sulcal points sets [8], matching of two fingerprints [9], etc. Because there are no image information can be used for the establishment of feature-based corresponding pairs and only the location information is available. Our method belongs to the second class which can be roughly divided into three categories: Iterative closest point (ICP) and its numerous extensions [10], soft assignment methods [8], [2] and probabilistic methods [11], [12], [13].

In this work, we focus on probabilistic methods for the modeling of point set matching problems, especially Gaussian mixtures model(GMM). In general, GMM-like methods assume that points from the scene set are normally distributed around points belonging to the model set [14]. Therefore the point-to-point assignment problem is re-interpreted as that of estimating the parameters of a mixture and finding the mapping between two point sets[11]. Mixture point matching(MPM) method is firstly proposed for feature points registration framework by use of GMM explicitly, which includes a unified probabilistic treatment of noise, outliers, transformation and regularization parameters [11]. It is a simplified version of GMM for the reason that an isotropic(spherical) covariance is common to all the components. Coherent point drift(CPD) is another well-known EM-like points matching method which assumes an isotropic(spherical) covariance for all the components as well [13]. Without the use of EM-like updates, the KL-divergence can be minimized explicitly between two GMMs of common isotropic covariance with respect to model and scene, which leads to a non-linear optimization problem under non-convex rigidity constraints [12]. The main shortcomings of isotropic covariance assumption is that anisotropic noise in the data is not properly handled and it dose not use the full Gaussian mixtures model. To overcome this drawbacks, Horaud et al. [14] propose an expectation conditional maximization for point registration(ECMPR), where the covariances of mixture components are anisotropic. However, this assumption of anisotropic leads to a non-linear optimization step in the EM update process. It should be mentioned that these aforementioned methods merely make point estimations of parameters of mixtures and mapping functions by use of maximum likelihood estimation(ML), maximum a posteriori estimation(MAP), non-linear optimization techniques. Therefore it is inconvenient to make Bayesian inference on parameters and variables during the approximation process. As far as outliers are concerned, most of existing matching methods append just one extra normal or uniform component to estimate the spurious points. It is apparent that more complicate statistical methods should be studied to model the outliers aiming to alleviate the influence of multi-cluster of outliers and improve the robustness of matching.

In this paper, we propose a new variational inference algorithm for robust point set matching problem. The contributions of the present work is mainly from the following aspects: (1) Variational inference algorithm is proposed for the reasoning of uncertainty of parameters of mixtures and mapping functions under the full Bayesian framework. Unlike ML and MAP, variational Bayesian framework explicitly accounts for matching uncertainty and is thus less prone to local optima comparing to point estimation methods. (2) An extra Gaussian mixtures model is established for the modeling of spurious outlier points, which provide our method favorable robust performance for detecting and rejecting outliers. (3) Probabilistic anisotropic covariance is established for the modeling of point set matching problem and this anisotropic formulation provide our method the flexible ability of mixtures model for different applications. Finally, experiments are made on synthetic and real data and the proposed approach demonstrates comparable performance in terms of both robustness and accuracy.
II. PROBABILISTIC POINT SET MATCHING

A. Problem Formulation

The point set matching problem can be formulated as: Given a shape model $M$ (in the form of a vector of landmarks) and a collection of points in a scene $S$, the goal of point set matching is to find the correspondence between the landmark and scene points under rigid and non-rigid transformation, also allowing outliers. The probabilistic model for point set matching problem can be expressed as follows [15]:

$$p(S|M) = \int p(S,T|M)p(T)dt,$$  \hspace{1cm} (1)

where $M = \{ m_k \}_{k=1}^K$ is the model point set containing $K$ feature components and $S = \{ s_n \}_{n=1}^N$ is the scene point set containing $N$ data points. The $m_k$ and $s_n$ are feature vectors of dimension $D$ with respect to model and scene.

According to Equation 1, the goal of Bayesian point set matching is to find a posterior distribution over transformation $T$ that maximizes the marginal probability $p(S|M)$ under certain objective function. The transformation $T$ can be divided into two additive parts: linear transformation for affine matching and nonlinear transformation for nonrigid matching, which can be expressed as follows:

$$T(M) = AM + B\Phi(M),$$  \hspace{1cm} (2)

where $A$ is affine parameters, $B$ and $\Phi(M)$ are weight and basis matrix for nonrigid matching, respectively. The bias vector is absorbed into $A$ for simplicity. After the linearized configuration of nonrigid matching function, the automatic relevance determination (ARD) [16] can be conveniently utilized to infer the posteriors of transformation $A$ and $B$.

B. Bayesian Point Set Matching with GMM

As the numbers of model and scene set points may be different, latent random variable $X = \{ x_k \}_{k=1}^K$ is explicitly introduced into the matching model to established the relationship between $M$ and $S$, such that $x_k = A\tilde{m}_k + B\big(\Phi(m_k)\big) + \epsilon_k$ where $\tilde{m}_k = [1;'m_k]$ and $\epsilon_k$ is a $D$-dimensional zero-mean Gaussian noise with diagonal precision $\Psi$. The conditional distribution of $X$ given $A$, $B$, $\Psi$ and $M$ can be written as follow:

$$p(X|\Theta,M) = \prod_{k=1}^K N(x_k|A\tilde{m}_k + B\big(\Phi(m_k)\big),\Psi^{-1}),$$  \hspace{1cm} (3)

where $\Theta = \{ A, B, \Psi \}$. By means of the transformation, $K$ mixture components are obtained in the observational space so that the scene data points can be seen as being generated from a Gaussian mixtures.

In addition, an extra Gaussian mixtures is added in our framework to account for the spurious outliers in the scene set. The number of components for outlier mixtures can be deemed known or learned from the matching process. With respect to each scene point $s_n$, there is a corresponding indicator latent random variable $Z = \{ z_n, n = 1, \ldots, N \}$ in mixtures, where $z_n$ is a 1-of-$K+K_o$ binary vector with elements $\{ z_{nk}, k = 1, \ldots, K + K_o \}$, where $K_o$ is the number of components for outlier mixtures. Therefore, we have the following distribution:

$$p(Z|\Omega) = \prod_{n=1}^N \prod_{k=1}^{K+K_o} N(z_n|y_k,\Lambda_k^{-1})z_{nk},$$  \hspace{1cm} (4)

where $\Omega = \{ Z, X, Y, A, Y \}$, $A = \{ A_k \}_{k=1}^K$ are precision for each Gaussian component with respect to latent variable $X$, $Y = \{ y_k \}_{k=1}^{K+K_o}$ and $Y = \{ Y_k \}_{k=1}^{K+K_o}$ are the mean and precision of Gaussian mixtures for outlier points, respectively. An independent Gaussian-Wishart prior is introduced to govern the mean $\nu$ and precision $\Psi$ of mixture component with prior parameters $\{ \nu^0, \Omega^0 \}$. The indicator variable $Z$ has a distribution with mixing coefficients $\pi$:

$$p(Z|\pi) = \prod_{n=1}^N \prod_{k=1}^{K+K_o} \pi_k^{z_{nk}}.$$  \hspace{1cm} (5)

The prior of mixing proportions $\pi$ is Dirichlet distribution with parameter $\alpha^0$.

We choose hierarchical priors over affine transformation and weight matrix to perform ARD over the entries of $A$ and $B$. Each columns of $A$ and $B$ have Gaussian priors with zero mean and different precision parameter $\nu$ and $\eta$ which is drawn from Gamma distributions with parameters $\{ a^0, b^0 \}$ and $\{ c^0, d^0 \}$ respectively. The prior of noise precision $\Psi$ is assumed to be Wishart distribution with degree of freedom $\lambda$ and precision matrix $\Omega^0$. Similar to $\Psi$, the precision $\Lambda_k$ of each component in GMM for $X$ assumed on prior Wishart distributions with parameters of degree of freedom $\gamma^0_k$ and precision matrix $W^0_k$.

III. VARIATIONAL POSTERIOR DISTRIBUTION

The log-marginal likelihood $\log p(S|M)$ is the objective function for variational Bayesian optimization for point set matching. One can obtain a lower bound $F$ for the log marginal probability, which is called the negative free energy [17]. We can factorize the posterior approximate distribution $q(\cdot)$ thus the prior and posterior will preserve the same probability distribution form under the conjugate-exponential structure. The variational posterior distributions over latent variables and transformation parameters can be written as follows, where $\langle \cdot \rangle$ is the expectation operator:

(1) Latent variable $X$:

$$q(X) = \prod_{k=1}^K q(x_k) = \prod_{k=1}^K N(x_k|\mu_{x_k}, \Sigma_{x_k}),$$  \hspace{1cm} (6)

where

$$\hat{\Sigma}_{x_k}^{-1} = \sum_{n=1}^N (z_{nk}A_k) + \langle \Psi \rangle,$$

$$\mu_{x_k} = \Sigma_{x_k}^{-1} \left[ \sum_{n=1}^N (z_{nk}A_k s_n) + \langle \Psi (A\tilde{m}_k + B\big(\Phi(m_k)) \rangle \right].$$

The mean and covariance of latent variable $X$ are the combination of model and scene information. The covariance of
$x_k$ is different with each other because of the relative position of transformed model point with respect to all of the scene points.

(2) Affine transformation $A$:

Following the idea of [18], the posterior over transformation is factorized over the rows of $A$ and the precision parameter $\nu$ is placed over the columns of $A$. Let $\Xi_q$ and $\Gamma_q$ denote the approximate posterior mean and covariance over the $q$th row of $A$, and these variational posterior parameters are given by

$$\Gamma_q^{-1} = \text{diag}(\nu) + \left(\Psi_{qq}\right) \sum_{k=1}^{K} \hat{m}_k \hat{m}_k^T,$$

$$\Xi_q = \Gamma_q \left[\left(\Psi_{qq}\right) \sum_{k=1}^{K} \left(\langle x_{k,q} \rangle - \langle A \Phi(m_k) \rangle \right) m_k^T \right].$$

The posterior distribution over $A$ has block diagonal covariance structure, which is a $(D \times (D+1) \times (D+1))$ matrix.

(3) Weight matrix $B$:

Similar to the affine transformation $A$, the posterior over weight $B$ is also factorized over the rows and the precision parameter $\eta$ is placed over the columns of $B$. Therefore, variational posterior mean $\hat{\Theta}_q$ and covariance $\hat{\Omega}_q$ over the $q$th row of $B$ are expressed as following:

$$\hat{\Omega}_q^{-1} = \text{diag}(\eta) + \left(\Psi_{qq}\right) \sum_{k=1}^{K} \hat{m}_k \hat{m}_k^T,$$

$$\hat{\Theta}_q = \hat{\Omega}_q \left[\left(\Psi_{qq}\right) \sum_{k=1}^{K} \left(\langle x_{k,q} \rangle - \langle A \Phi(m_k) \rangle \right) m_k^T \right].$$

Covariance $\hat{\Omega}_q$ has a diagonal structure and is a $(D \times K \times K)$ matrix.

(4) Noise precision matrix $\Psi$:

The variational posterior of the noise precision $\Psi$ is also a Wishart distribution. The updates of degree of freedom and precision matrix are

$$\lambda = \lambda^0 + K,$$

$$U^{-1} = \text{diag}\left[U_2^{0,-1} + \sum_{k=1}^{K} \left(\langle x_k - A \hat{m}_k - B \Phi(m_k) \rangle \right) \right].$$

The diag operator sets off-diagonal terms to zero, which means the noise is assumed to be no correlation with respect to axis.

(5) Precision matrix $\Lambda$ of Gaussian mixtures:

The posterior of the component precision $\Lambda$ is a Wishart distribution as well. The updates of degree of freedom and precision matrix for each component are

$$\gamma_k = \gamma_k^0 + \sum_{n=1}^{N} \langle z_{nk} \rangle,$$

$$W_k^{-1} = W_k^{0,-1} + \sum_{n=1}^{N} \langle z_{nk} (s_n - x_k) (s_n - x_k)^T \rangle.$$

(6) Indicator variables $Z$:

The approximate distribution of $Z$ contains two parts: $N \times K$ and $N \times K_o$ submatrix for the responsibilities of $X$ and outlier GMM, respectively. Let $\bar{r}_{nk} = \langle z_{nk} \rangle$, we have the following equation for $1 \leq k \leq K$:

$$r_{nk} = \frac{1}{\bar{r}_{nk}} \exp \left\{ (\log \pi_k) + \frac{1}{2} (\log |A_k|) \right\} - \frac{1}{2} \left\{ (s_n - x_k)^T A_k (s_n - x_k) \right\},$$

where $Z_{nk}$ is a normalization constant for each data point.

The variational approximate posteriors of GMM for outliers can be found in the textbook [17]. These update equations are coupled and therefore must be solved iteratively. The learning steps for this method may be conveniently implemented according to Algorithm 1, which is illustrated in a pseudo-code form.

**Algorithm 1 Variational Bayesian for Point Set Matching (VBPSM)**

1. **Input:**

   $M = \{m_k\}_{k=1}^{K}, S = \{s_n\}_{n=1}^{N}, \in F.$

2. **Initialization:**

   - parameters for priors: $a^0, b^0, c^0, d^0, \alpha^0, U^0, \lambda^0, W^0, \gamma^0$.
   - parameters for outlier model: $y^0, L^0, t^0, \beta^0$.
   - transformations: $A^0$ and $B^0$.

3. **Reconstruct:** $X = \{x_k\}_{k=1}^{K}$.

   repeat
   5. Infer the indicate variable $Z$ using Equation 12.
   6. Infer the posterior $\eta, L_t, \beta$ for outliers.
   7. Infer the affine parameters $A$ using Equation 8.
   8. Infer the weight $B$ using Equation 9.
   9. Infer the latent variable $X$ using Equation 7.
   10. Update parameters $a, b, c, d, \alpha, U, \lambda, W, \gamma$.

   until $|F(t + 1) - F(t)| < \varepsilon_F$.

**IV. THE PRIORS**

Initial setup of prior parameters has great influence on inferring posteriors in variational inference. For example, if prior $W^0$ is set to be small value, then the posterior precision will be small(big covariance) as well. The larger uncertainty may help GBP to capture the main tendency of scene and avoid being trapped in local minima, where Figure 1 demonstrates how the different prior $W^0$s exert influence on matching results at rotation of $\theta = 40^\circ$. However, one can just obtain a coarse point-to-point correspondence because of lower posterior precision $||W||$. On the contrary, a fine point-to-point matching is at the expense of being apt to be trapped into local minima. This phenomenon is similar to the tuning of annealing temperature $T$ in MPM [11]. In this work, we propose an empirical coarse-to-fine strategy for our method, that is, firstly taking advantage of small prior precision to grasp the main tendency of scene, and then acquiring accurate matching result by use of larger prior precision. $W^0$ is generally chosen from $1 \times 1$ to $20 \times 1$, which are accurate enough for all these experiments.

As far as the other prior parameters is concerned, we set the values of these parameters as follows: (1) $a^0$, $b^0$, $c^0$ and $d^0$ are set to be 0.001, respectively, which means noninformative priors on $A$ and $B$. (2) $\alpha^0$ and $\gamma^0$ are all set to be $D + 1$. (3) $\beta^0$ refers to the prior information about the
numbers of data in scene set belonging to model and outliers, respectively. In general, the elements of $\alpha^0$ are set to be 1. (4) The hyperparameters with respect to outliers are set as following: $y^0$ is normally generated with zero-mean and unit covariance, $L^0$ is set to be $1^D$, $t^0$ is set to $D + 1$ and $\beta^0 = 1$. Notice that, the prior knowledge about outliers is explicitly represented in prior parameters and they are updated during the approximation process. Therefore, one do not need to set a strong assumption on outlier ratio just as the CPD and GMM-L2 have done.

V. EXPERIMENTAL RESULTS

In this section, we present some results obtained from the application of our method to both synthetic and real data sets including 2D shapes, 3D surfaces and different real applications. Quantitative comparisons with competing point set matching methods are also presented. Our algorithm is implemented in Matlab v7.12 and tested on a Intel Core2 CPU 2.67GHz with 4G RAM. The complexity of a single VBPSM iteration is $O(N \times (K + K_o))$ and it takes seconds for matching hundred-point problems in general.

A. 2D synthetic data

The five 2D data sets are obtained from these papers [8], [13], [2], which is shown in Figure 2. In the first experiment, our method is performed on a 2D point set matching problem where multiple clusters of outliers are imposed on scene in order to illustrate how Gaussian mixtures estimate the spurious outliers as well as the mapping function, which is shown in Figure 4. There are three clusters of outliers added in the scene set, which are (1) normally distributed outliers, (2) uniformly distributed outliers, and (3) points from a straight line. The true target fish data is also transformed by an affine matrix from model. As the number of outlier clusters is not known in advance, we use a $K_o=4$ to represent the prior information about the clusters. The approximate evolution is illustrated in Figure 6. As the log-marginal likelihood is approximated by negative free energy, the structure centers (model points) are transformed to match the target fish points and these spurious outliers are estimated by Gaussian mixtures as well.

In the second experiment, two quantitative comparisons are conducted on four affine point set matching algorithms. The scene data sets are generated by moderate affine transformations from original model point sets: (1) translating model with a displacement $[T_x \ T_y]=[0.5 0.50]$, (2) scaling the point set with $S_x=1.20$, $S_y=1.15$ with respect to $x$ and $y$ axes, (3) rotating the point set with angle 30, and (4) finally, shear the shape with $sh_x=0.10$ and $sh_y=0.15$. Since the true affine parameters are known in advance, the Euclidean norm between the true transformation and the estimated affine parameters is used as the metric of matching accuracy and 20 runs are conducted on each experiment. To test the ability of these algorithms to occlusion, we delete increasing numbers of points from the head of model points and points from the tail part of scene, respectively. The missing points is counted from 1 to 21 on fish-2 data set and the result is shown in Figure 4(a). We also test the rotation convergence performance of VBPSM on five clean 2D data sets. Each data set is rotated from $-\pi$ to $\pi$ and four algorithms are performed for aligning the original point set towards the rotated one. The range of convergence angles are reported in Figure 4(b) with respect to these four methods. These comparisons show that the VBPSM algorithm has comparable performance to the other three approaches in terms of occlusion and rotation.
on two data sets. The matching results are shown in Figure 5. In the fish-1 data set, the target points is merged with outliers therefore the balance between two mixtures shrink the transformed model points at the middle of two distributions. For this reason, VBPSM cannot reach a good matching result for the points at the top of fin and tail.

Fig. 5. **Top Row:** Initialization with model (blue ‘+’) and scene(red ‘o’). **Bottom Row:** Matching results from VBPSM, where blue ‘+’ and ellipse represent the center and the covariance (blue ellipse) with respect to each components, respectively.

B. Quantitative Comparison

In the quantitative comparison, the scene data are generated from random transformation of fish-2 model set. (1) translation $T_x$ and $T_y$ are uniformly sampled from [-1.0 1.0], (2) scaling $S_x$ and $S_y$ are uniformly sampled from [0.2 1.2], (3) rotate angle $\theta$ is uniformly sampled from [-20 -20], and (4) shape shearing $s_{bx}$ and $s_{by}$ are uniformly sampled from [-0.1 0.1]. The scene points are corrupted by spurious outliers which are generated from gaussian distribution with different standard deviation. The outliers to data ratios are $w=0.4, 0.8, 1.2$ and 1.6, respectively. The Euclidean norm between the true transformation and the estimated affine parameters is used as the metric of matching accuracy and 20 runs are conducted on each noise standard deviation. As one can see from the results in Figure 6, the performance VBPSM is very consistent in comparison with other three methods. Though CPD achieves the best performance at the lowest outliers level, it performs poor in other two experiments just as KC and GMM-L2. In addition, while KC, CPD and GMM-L2 lose sight of the case of multi-cluster outliers, VBPSM can also handle with this kind of outliers. For the limit of pages, we left detailed discussion in future work.

C. Real data

VBPSM is also applied to noised 3D face alignment experiment and the results are demonstrated in Figure 7. There are $N=392$ points in model. In the noise setup, we first delete $64(16\%)$ data points around the left face in model. After that, three clusters of uniformly and normally distributed outliers are added to the corrupt model, where each outlier cluster contains $0.3*N$ outliers. Finally moderate 3D rotation and translation are applied to the data set to construct scene points, as is shown in Figure 7(a). Figure 7(b) demonstrates that VBPSM achieves good performance on 3D data in the presence of outliers and missing points. There are three mixture components preset for outliers and they estimate the outliers in scene in a good manner.

Fig. 6. Performance comparison of KC, CPD, GMM-L2 and VBPSM with respect to outliers. (a) to (d) are comparison with different outliers to data ratio. VBPSM outperforms other three method at $w=0.8, 1.2$ and 1.6.

In the second real data example, nonrigid VBPSM is used to align two retinal images on the top row of Figure 8. The moving between eyeballs and the corresponding minutiae is clearly seen from the figure. There is serious bias between two point sets and almost half of the minutiae are outliers with respect to two images. In Figure 8(c), the feature points extracted from the left image are successfully registered to the feature points extracted from the right image with 23 matches out of 40 putative points. The influence of outliers is greatly alleviated by the mixture model for outliers. Since the goal of the point set registration in the retinal matching problem to establish the pairing of minutiae, it is sufficient to bring the corresponding minutiae according to response matrix and distance threshold instead of perfect alignment.

Fig. 7. Performance of VBPSM on 3D face alignment, (a) initial setup, (b) matching results.

Finally, we perform image alignment of two astronaut photos of earth by our method. The images are downloaded from the "The Gateway to Astronaut Photography of Earth" program of Image Science and Analysis Laboratory, NASA-Johnson Space Center. Two photographs, ISS006-E-51977 and ISS006-E-51981, are camera captures of San Diego bay from the international space station. There are 179 matches are found from two images by SIFT feature detector, which are superimposed on these two images as shown in the top
row of Figure 9. It should be to note that there are outliers in these putative match pairs so that five extra mixtures are used to model these incorrect candidates. The affine transformation parameters are estimated from two feature point sets and are applied for the alignment of two images. We also demonstrate the estimated outlier clusters obtained from our method in Figure 9(c). The mixtures model provides a flexible estimation tool for these scattered spurious data points, and contributes a lot for robust and accurate alignment of LANSAT images.

VI. CONCLUSION

In this work, we have introduced a variational Bayesian probabilistic framework for affine and nonrigid point set matching. We consider the matching of two point sets as probabilistic inference and the transformation and parameters are all assumed to be random variables. In addition, Gaussian mixtures is used to estimate the uncertainty of outliers during the matching process as well. Various experiments have been performed on synthetic and real data sets in comparison with other probabilistic methods. In these experiments, our method demonstrates robust and accurate performance with respect to outliers and missing data and is comparable to other state-of-art point set matching algorithms.

ACKNOWLEDGMENT

This work is supported by BIAST Youth Backbone Training Plan (2011-20) and Beijing Science and Technology Plan (Z131100004113001).

REFERENCES