Simultaneous Nonrigid Registration of Multiple Point Sets and Atlas Construction*

Fei Wang[1], Baba C. Vemuri[1], Anand Rangarajan[1],
Ilona M. Schmalfuss[2], and Stephan J. Eisenschenk[3]

[1] Department of Computer & Information Sciences & Engr.,
University of Florida, Gainesville
[2] Departments of Radiology, University of Florida, Gainesville
[3] Department of Neurology, University of Florida, Gainesville

Abstract. Estimating a meaningful average or mean shape from a set of shapes represented by unlabeled point-sets is a challenging problem since, usually this involves solving for point correspondence under a non-rigid motion setting. In this paper, we propose a novel and robust algorithm that is capable of simultaneously computing the mean shape from multiple unlabeled point-sets (represented by finite mixtures) and registering them nonrigidly to this emerging mean shape. This algorithm avoids the correspondence problem by minimizing the Jensen-Shannon (JS) divergence between the point sets represented as finite mixtures. We derive the analytic gradient of the cost function namely, the JS-divergence, in order to efficiently achieve the optimal solution. The cost function is fully symmetric with no bias toward any of the given shapes to be registered and whose mean is being sought. Our algorithm can be especially useful for creating atlases of various shapes present in images as well as for simultaneously (rigidly or non-rigidly) registering 3D range data sets without having to establish any correspondence. We present experimental results on non-rigidly registering 2D as well as 3D real data (point sets).

1 Introduction

In recent years, there has been considerable interest in the application of statistical shape analysis to problems in medical image analysis, computer graphics and computer vision. Regardless of whether shapes are parameterized by points, lines, curves etc., the fundamental problem of estimating mean and covariance of shapes remains. We are particularly interested in the unlabeled point-set parameterization since statistical shape analysis of point-sets is very mature [1]. Means, covariances and probability distributions on shape manifolds can now be defined and estimated.

The primary technical challenge in using point-set representations of shapes is the correspondence problem. Typically correspondences can be estimated once the point-sets are properly aligned with appropriate spatial transformations. If

* This research was in part funded by the NIH grants, RO1 NS046812 & NS42075 and NSF grant NSF 0307712.

A. Leonardis, H. Bischof, and A. Pinz (Eds.): ECCV 2006, Part III, LNCS 3953, pp. 551–563, 2006.

the objects at hand are deformable, the adequate transformation would obviously be a non-rigid spatial mapping. Solving for nonrigid deformations between point-sets with unknown correspondence is a hard problem. In fact, many current methods only attempt to solve for affine transformation for the alignment. Furthermore, we also encounter the issue of the bias problem in atlas creation. Since we have more than two sample point-sets to be aligned for creating an atlas, a question that arises is: How do we align all the point-sets in a symmetric manner so that there is no bias toward any particular point-set?

To overcome these aforementioned problems, we present a novel approach to simultaneously register multiple point-sets and construct the atlas. The idea is to model each point set by a kernel probability distribution, then quantify the distance between these probability distributions using an information-theoretic measure. The distance is optimized over a space of coordinate transformations yielding the desired registrations. It is obvious that once all the point sets are deformed into the same shape, the distance measure between these distributions should be minimized since all the distribution are identical to each other. We impose regularization on each deformation field to prevent over-deforming of each point-sets (e.g. all the point-sets may deform into a single data point). Jensen-Shannon divergence, first introduced in [2], serves as a model divergence measure between multiple probability distributions. It has some very desirable properties, researchers have used it as a dissimilarity measure for image registration and retrieval applications [3, 4].

The rest of this paper is organized as follows. The remainder of section 1 gives a brief review of the literature, focusing on difference between these methods and ours. Section 2 contains a description of our formulation using JS-divergence for our simultaneous nonrigid registration and atlas construction model. Experimental results on 2D as well as 3D point-sets are presented in Section 3.

1.1 Previous Work

Extensive studies on the atlas construction for deformable shapes can be found in literature covering both theoretical and practical issues relating to computer vision and pattern recognition. According to the shape representation, they can be classified into two distinct categories. One is the methods dealing with shapes represented by feature point-sets, and everything else is in the other category including those shapes represented as curves and surfaces of the shape boundary, and these curves and surfaces may be either intrinsicly or extrinsicly parameterized (e.g. using point locations and spline coefficients).

The work presented in [5] is a representative method using an intrinsic curve parameterization to analyze deformable shapes. Shapes are represented as elements of infinite-dimensional spaces and their pairwise difference are quantified using the lengths of geodesics connecting them on these spaces, the intrinsic mean (Karcher mean) can be computed as a point on the manifold (of shapes) which minimize the sum of square geodesic distance between this unknown point to each individual shape, which lies on the manifold. However the curves are limited by closed curves, and it has not been extended to the 3D surface shapes.

For methods using intrinsic curve or surface representations [5, 6], further statistical analysis on these representations is much more difficult than analysis on the point representation, but the reward maybe higher due to the use of intrinsic higher order representation.

Among these methods using point-sets parameterization, the idea of using nonrigid spatial mapping functions, specifically thin-plate splines [7, 8, 9], to analyze deformable shape has been widely adopted. Bookstein's work in [7], successfully initiated the research efforts on the usage of thin-plate splines to model the deformation of shapes. This method is landmark-based, it avoids the correspondence problem since the placement of corresponding points is driven by the visual perception of experts, however it suffers from the the typical problem besetting landmark methods, e.g. inconsistency. Several significant articles on robust and non-rigid point set matching have been published by Rangaranjan and collaborators [8] using thin-plate splines. The main strength of their work is the ability to jointly determine the correspondences and non-rigid transformation between each point sets to the emerging mean shape using deterministic annealing and soft-assign. However, in their work, the stability of the registration result is not guaranteed in the case of data with outliers, and hence a good stopping criterion is required. Unlike their approach, we do not need to first solve a correspondence problem in order to subsequently solve a non-rigid registration problem.

The active shape model proposed in [10] utilized points to represent deformable shapes. Their work pioneered the efforts in building point distribution models to understand deformable shapes [10]. Objects are represented as carefully-defined landmark points and variation of shapes are modeled using a principal component analysis. These landmark points are acquired through a more or less manual landmarking process where an expert goes through all the samples to mark corresponding points on each sample. It is a rather tedious process and accuracy is limited. In recent work [11], the authors attempt to overcome this limitation by attempting to automatically solve for the correspondences in a nonrigid setting. The resulting algorithm is very similar to the earlier work in [6] and is restricted to curves.

There are several papers in the point-sets alignment literature which bear close relation to our research reported here. For instance, Tsin and Kanade [12] proposed a kernel correlation based point set registration approach where the cost function is proportional to the correlation of two kernel density estimates. It is similar to our work since we too model each of the point sets by a kernel density function and then quantify the (dis)similarity between them using an information-theoretic measure, followed by an optimization of a (dis)similarity function over a space of coordinate transformations yielding the desired transformation. The difference lies in the fact that JS-divergence used in our work is a lot more general than the information-theoretic measure used in [12], and can be easily extended to multiple point-sets. More recently, in [13], Glaunes et al. convert the point matching problem into an image matching problem by treating points as delta functions. Then they "lift" these delta functions and

diffeomorphically match them. The main problem for this technique is that they need a 3D spatial integral which must be numerically computed, while we do not need this due to the empirical computation of the JS-divergence. We will show it in the experimental results that our method, when applied to match point-sets, achieves very good performance in terms of both robustness and accuracy.

2 Methodology

In this section, we present the details of the proposed simultaneous atlas construction and non-rigid registration method. The basic idea is to model each point set by a probability distribution, then quantify the distance between these probability distributions using an information-theoretic measure. The distance measure is optimized over a space of coordinate transformations yielding the desired transformations. We will begin by presenting the finite mixtures used to model the probability distributions of the given point-sets.

2.1 Finite Mixture Models

Considering the point set as a collection of Dirac Delta functions, it is natural to think of a finite mixture model as representation of a point set. As the most frequently used mixture model, a Gaussian mixture [14] is defined as a convex combination of Gaussian component densities.

We use the following notation: The data point-sets are denoted by $\{X^p, p \in \{1, ..., N\}\}$. Each point-set X^p consists of points $\{x_i^p \in \mathcal{R}_D, i \in \{1, ..., n_p\}\ \}$. To model each point-set as a Gaussian mixture, we define a set of cluster centers, one for each point-set, to serve as the Gaussian mixture centers. Since the feature point-sets are usually highly structured, we can expect them to cluster well. Furthermore we can greatly improve the algorithm efficiency by using limited number of clusters. Note that we can choose the cluster centers to be the point-set itself if the size of point-sets are quite small. The cluster center point-sets are denoted by $\{V^p, p \in \{1, ..., N\}\}$. Each point-set V^p consists of points $\{v_i^p \in \mathcal{R}_D, i \in \{1, ..., K^p\}\ \}$. Note that there are K^p points in each V^p, and the number of clusters for each point-set may be different (in our implementation, the number of clusters were usually chosen to be proportional to the size of the point-sets). The cluster centers are estimated by using a clustering process over the original sample points x_i^p, and we only need to do this once before the process of joint atlas estimation and point-sets registration. The atlas points-set is denoted by Z. We begin by specifying the density function of each point set.

$$p(X^p|V^p, \alpha^p) = \prod_{i=1}^{n_p} \sum_{a=1}^{K^p} \alpha_a^p p(x_i^p|v_a^p) \tag{1}$$

In Equation (1), the occupancy probability which is different for each data point-set is denoted by α^p. $p(X^p|V^p, \alpha^p)$ is a mixture model containing the component densities $p(x_i^p|v_a^p)$, where

$$p(x_i^p)|v_a^p) = \frac{1}{(2\pi)^{\frac{D}{2}}\Sigma_a^{\frac{1}{2}}} \exp\Big(-\frac{1}{2}\big(x_i^p - v_a^p\big)^T \Sigma_a^{-1}\big(x_i^p - v_a^p\big)\Big) \tag{2}$$

Later, we set the occupancy probability to be uniform and make the covariance matrices Σ_a to be proportional to the identity matrix in order to simplify atlas estimation procedure.

Having specified the Gaussian mixtures of each point-set, we would like to compute a meaningful average/mean (shape) point-set Z, given all the sample sets and their associated distributions. Intuitively, if these point-sets are aligned correctly under appropriate nonrigid deformations, the resulting mixtures should be statistically similar to each other. Consequently, this raises the key question: how to measure the similarity/closeness between these distributions represented by Gaussian mixtures? We will answer this in the following paragraphs.

2.2 Jensen-Shannon Divergence for Learning the Atlas

Jensen-Shannon (JS) divergence, first introduced in [2], serves as a measure of cohesion between multiple probability distributions. It has been used by some researchers as a dissimilarity measure for image registration and retrieval applications [3, 4] with very good results. It has some very desirable properties, to name a few, 1) The square root of JS-divergence (in the case when its parameter is fixed to $\frac{1}{2}$) is a metric [15]; 2) JS-divergence relates to other information-theoretic functionals, such as the relative entropy or the Kullback divergence, and hence it shares their mathematical properties as well as their intuitive appeal; 3) The compared distributions using the JS-divergence can be weighted, which allows one to take into account the different sizes of the point set samples from which the probability distributions are computed; 4) The JS-divergence measure also allows us to have different numbers of cluster centers in each point-set. There is NO requirement that the cluster centers be in correspondence as is required by Chui et al [16]. Given n probability distributions $\mathbf{P_i}$, $i \in \{1, ..., n\}$, the JS-divergence of $\mathbf{P_i}$ is defined by

$$JS_\pi(\mathbf{P_1}, \mathbf{P_2}, ..., \mathbf{P_n}) = H(\sum \pi_i \mathbf{P_i}) - \sum \pi_i H(\mathbf{P_i}) \tag{3}$$

where $\pi = \{\pi_1, \pi_2, ..., \pi_n | \pi_i > 0, \sum \pi_i = 1\}$ are the weights of the probability distributions $\mathbf{P_i}$ and $H(P_i)$ is the Shannon entropy. The two terms on the right hand side of Equation (3) are the entropy of $\mathbf{P} := \sum \pi_i \mathbf{P_i}$ (the π- convex combination of the $\mathbf{P_i}$s) and the same convex combination of the respective entropies.

Assume that each point set X^p is related to Z via a function f^p, $\boldsymbol{\mu}^p$ is the set of the transformation parameters associated with each function f^p. To compute the mean shape from these point-sets and register them to the emerging mean shape, we need to recover these transformation parameters to construct the mean shape. This problem can modeled as an optimization problem with the objective function being the JS-divergence between the distributions of the deformed point-sets, represented as $\mathbf{P}_i = p(f^i(X^i))$, the atlas construction problem can now be formulated as,

$$\begin{aligned}&\min_{\boldsymbol{\mu}^i} JS_\beta(\mathbf{P}_1, \mathbf{P}_2, ..., \mathbf{P}_N) + \lambda \sum_{i=1}^{N} ||Lf^i||^2 \\ =&\min_{\boldsymbol{\mu}^i} H(\sum \beta_i \mathbf{P}_i) - \sum \beta_i H(\mathbf{P}_i) + \lambda \sum_{i=1}^{N} ||Lf^i||^2\end{aligned} \tag{4}$$

In (4), the weight parameter λ is a positive constant the operator L determines the kind of regularization imposed. For example, L could correspond to a thin-plate spline, a Gaussian radial basis function, etc. Each choice of L is in turn related to a kernel and a metric of the deformation from and to Z.

Following the approach in [8], we choose the thin-plate spline (TPS) to represent the non-rigid deformation. Given n control points $\mathbf{x}_1, \ldots, \mathbf{x}_n$ in $\mathbb{R}^d$, a general nonrigid mapping $f : \mathbb{R}^d \to \mathbb{R}^d$ represented by thin-plate spline can be written analytically as: $f(x) = \mathbf{WU}(\mathbf{x}) + \mathbf{Ax} + \mathbf{t}$ Here $\mathbf{Ax} + \mathbf{t}$ is the linear part of f. The nonlinear part is determined by a $d \times n$ matrix, $\mathbf{W}$. And $\mathbf{U}(\mathbf{x})$ is an $n \times 1$ vector consisting of n basis functions $U_i(\mathbf{x}) = U(\mathbf{x}, \mathbf{x}_i) = U(\|\mathbf{x} - \mathbf{x}_i\|)$ where $U(r)$ is the kernel function of thin-plate spline. For example, if the dimension is 2 ($d = 2$) and the regularization functional is defined on the second derivatives of f, we have $U(r) = 1/(8\pi)r^2 ln(r)$.

Therefore, the cost function for non-rigid registration can be formulated as an energy functional in a regularization framework, where the regularization term in equation 4 is governed by the bending energy of the thin-plate spline warping and can be explicitly given by $trace(\mathbf{WKW}^T)$ where $\mathbf{K} = (K_{ij})$, $K_{ij} = U(p_i, p_j)$ describes the internal structure of the control point sets. In our experiments, the clusters is used as control points. Other schemes to choose control points may also be considered. Note the linear part can be obtained by an initial affine registration, then an optimization can be performed to find the parameter $\mathbf{W}$.

Having introduced the cost function and the transformation model, now the task is to design an efficient way to estimate empirical JS-divergence from the Gaussian mixtures and derive the analytic gradient of the estimated divergence in order to achieve the optimal solution efficiently.

2.3 Estimating the Empirical JS

For simplicity, we choose $\beta_i = \frac{1}{N}, \forall i = \{1, 2, ..., N\}$. Let $Q_p^{x_i^j} := \sum_{a=1}^{K} \alpha_a^p p(f^j(x_i^j) | f^p(v_a^p))$ be a mixture model containing component densities $p(f^j(x_i^j)|f^p(v_a^p))$,

$$p(f^j(x_i^j)|f^p(v_a^p)) = \frac{1}{(2\pi)^{\frac{D}{2}} \Sigma_a^{\frac{1}{2}}} \exp\Big(-\frac{1}{2}\big(f^j(x_i^j) - f^p(v_a^p)\big)^T \Sigma_a^{-1} \big(f^j(x_i^j) - f^p(v_a^p)\big)\Big) \tag{5}$$

Where $\{\Sigma_a, a \in \{1, ..., K\}\}$ is the set of cluster covariance matrices. For the sake of simplicity and ease of implementation, we assume that the occupancy probabilities are uniform ($\alpha_a^p = \frac{1}{K}$) and the covariance matrices Σ_a are isotropic, diagonal, and identical [$(\Sigma_a = \sigma 2 I_D)$]. Having specified the density function of the data, we can then rewrite Equation (4) as follows,

$$\begin{aligned}JS_\beta(\mathbf{P}_1, \mathbf{P}_2, ..., \mathbf{P}_N) = \frac{1}{N}\Big\{&[H(\sum \frac{1}{N}\mathbf{P}_i) - \sum H(\mathbf{P}_1)] \\ &+ [H(\sum \frac{1}{N}\mathbf{P}_i) - \sum H(\mathbf{P}_2)] + \cdots + [H(\sum \frac{1}{N}\mathbf{P}_i) - \sum H(\mathbf{P}_N)]\Big\}\end{aligned} \tag{6}$$

For each term in the equation, we can estimate the entropy using the weak law of large numbers, which is given by,

$$\begin{aligned} H(\sum \frac{1}{N}\mathbf{P}_i) - H(\mathbf{P}_j)) &= -\frac{1}{n_i}\sum_{i=1}^{n_i} \log \frac{Q_1^{x_i^j} + Q_2^{x_i^j} + ... + Q_N^{x_i^j}}{N} + \frac{1}{n_i}\sum_{i=1}^{n_i} \log Q_j^{x_i^j} \\ &= \frac{1}{n_i}\sum_{i=1}^{n_i} \log \frac{NQ_j^{x_i^j}}{Q_1^{x_i^j} + Q_2^{x_i^j} + ... + Q_N^{x_i^j}} \end{aligned} \tag{7}$$

Combining these terms we have,

$$\begin{aligned} JS(\mathbf{P}_1, \mathbf{P}_2, ..., \mathbf{P}_N) = \Big\{ & \frac{1}{n_1}\sum_{i=1}^{n_1} \log \frac{NQ_1^{x_i^1}}{Q_1^{x_i^1} + Q_2^{x_i^1} + ... + Q_N^{x_i^1}} \\ & + \frac{1}{n_2}\sum_{i=1}^{n_2} \log \frac{NQ_2^{x_i^2}}{Q_1^{x_i^2} + Q_2^{x_i^2} + ... + Q_N^{x_i^2}} + \cdots + \frac{1}{n_N}\sum_{i=1}^{n_N} \log \frac{NQ_N^{x_i^N}}{Q_1^{x_i^N} + Q_2^{x_i^N} + ... + Q_N^{x_i^N}} \Big\} \end{aligned} \tag{8}$$

2.4 Optimizing the Cost Function

Computation of the gradient of the energy function is necessary in the minimization process when employing a gradient-based scheme. If this can be done in analytical form, it leads to an efficient optimization method. We now present the analytic form of the gradient of the JS-divergence (our cost function):

$$\nabla JS = [\frac{\partial JS}{\partial \boldsymbol{\mu}^1}, \frac{\partial JS}{\partial \boldsymbol{\mu}^2}, ..., \frac{\partial JS}{\partial \boldsymbol{\mu}^N}] \tag{9}$$

Each component of the gradient maybe found by differentiating Eqn (8) with respect to the transformation parameters. In order to compute this gradient, let's first calculate the derivative of $Q_p^{x_i^j}$ with respect to $\boldsymbol{\mu}^l$,

$$\frac{\partial Q_p^{x_i^j}}{\partial \boldsymbol{\mu}^l} = \begin{cases} \frac{1}{(2\pi)^{\frac{D}{2}}\sigma 3K}\sum_{a=1}^{K} -\exp\Big(-\frac{1}{2\sigma 2}|\mathbf{F}_{jp}|^2\Big)(\mathbf{F}_{jp}\cdot\frac{\partial f^j(x_i^j)}{\partial \boldsymbol{\mu}^l}) & \text{if } l = j \neq p \\ \frac{1}{(2\pi)^{\frac{D}{2}}\sigma 3K}\sum_{a=1}^{K} \exp\Big(-\frac{1}{2\sigma 2}|\mathbf{F}_{jp}|^2\Big)(\mathbf{F}_{jp}\cdot\frac{\partial f^p(v_a^p)}{\partial \boldsymbol{\mu}^l}) & \text{if } l = p \neq j \\ \frac{1}{(2\pi)^{\frac{D}{2}}\sigma 3K}\sum_{a=1}^{K} \exp\Big(-\frac{1}{2\sigma 2}|\mathbf{F}_{jp}|^2\Big)(\mathbf{F}_{jp}\cdot[\frac{\partial f^p(v_a^p)}{\partial \boldsymbol{\mu}^l} - \frac{\partial f^j(x_i^j)}{\partial \boldsymbol{\mu}^l}] & \text{if } l = p = j \end{cases} \tag{10}$$

where $\mathbf{F}_{jp} := f^j(x_i^j) - f^p(v_a^p)$. Based on this, it is straight forward to derive the gradient of the JS-divergence with respect to the transformation parameters $\boldsymbol{\mu}^l$, which is given by

$$\begin{aligned} \frac{\partial JS}{\partial \boldsymbol{\mu}^l} = \Big\{ & \frac{1}{n_1 N}\sum_{i=1}^{n_1}\Big(\log\frac{Q_1^{x_i^1} + Q_2^{x_i^1} + ... + Q_N^{x_i^1}}{N}\Big)\frac{\partial Q_l^{x_i^1}}{\partial \boldsymbol{\mu}^l} \\ & + \frac{1}{n_2 N}\sum_{i=1}^{n_2}\Big(\log\frac{Q_1^{x_i^2} + Q_2^{x_i^2} + ... + Q_N^{x_i^2}}{N}\Big)\frac{\partial Q_l^{x_i^2}}{\partial \boldsymbol{\mu}^l} + ... \\ & + \frac{1}{n_l N}\sum_{i=1}^{n_l}\Big(\log\frac{Q_1^{x_i^l} + Q_2^{x_i^l} + ... + Q_N^{x_i^l}}{N}\Big)\Big[\frac{\partial Q_1^{x_i^l}}{\partial \boldsymbol{\mu}^l} + ... + \frac{\partial Q_N^{x_i^l}}{\partial \boldsymbol{\mu}^l}\Big] - \frac{1}{n_l}\sum_{i=1}^{n_l}\Big(\log Q_l^{x_i^l}\Big)\frac{\partial Q_l^{x_i^l}}{\partial \boldsymbol{\mu}^l} \\ & + ... + \frac{1}{n_N N}\sum_{i=1}^{n_N}\Big(\log\frac{Q_1^{x_i^N} + Q_2^{x_i^N} + ... + Q_N^{x_i^N}}{N}\Big)\frac{\partial Q_l^{x_i^N}}{\partial \boldsymbol{\mu}^l}\Big\} \end{aligned} \tag{11}$$

Since the analytic gradients with respect to these transformation parameters has be explicitly derived in equation (12), we can use them in gradient-based numerical optimization techniques like the Quasi-Newton method and the nonlinear Conjugate-Gradient method to yield a fast solution.

Note that our algorithm can be applied to registration problems other than the atlas construction, e.g. we can apply it to align any two point-sets in 2D or 3D, in this case, there is a model point-set and a scene point-set (N=2). The only modification to the above procedure is to keep the scene point-set fixed and we try to recover the motion from the model point-set to the scene point-set such that the JS-divergence between these two distributions is minimized. We will present experimental results on point-set alignment between two given point-sets as well as atlas construction from multiple point-sets in the next section.

3 Experiment Results

We now present experimental results on the application of our algorithm to both synthetic and real data sets. First, to demonstrate the robustness and accuracy of our algorithm, we show the alignment results by applying the JS-divergence to the point-set matching problem. Then, we will present the atlas construction results in the second part of this section.

3.1 Alignment Results

First, to test the validity of our approach, we perform a set of exact rigid registration experiments on both synthetic and real data sets without noise and outliers. Some examples are shown in Figure 1. The top row shows the registration result for a 2D real range data set of a road (which was also used in Tsin and Kanade's experiments [12]). The figure depicts the real data and the registered (using rigid motion). Top left frame contains two unregistered point sets superposed on each other. Top right frame contains the same point sets after registration using our algorithm. A 3D helix example is presented in the second row (with the same arrangement as the top row). We also tested our method against the KC method [12] and the ICP methods, as expected, our method and

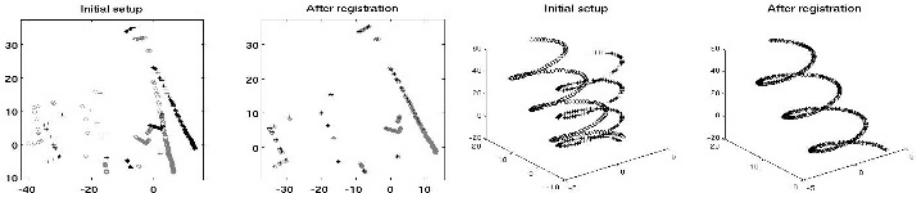

Fig. 1. Results of rigid registration in noiseless case. 'o' and '+' indicate the model and scene points respectively.

KC method exhibit a much wider convergence basin/range than the ICP and both achieve very high accuracy in the noiseless case.

Next, to see how our method behaves in the presence of noise and outliers, we designed the following procedure to generate a corrupted template point set from a model set. For a model set with n points, we control the degree of corruption by (1) discarding a subset of size $(1-\rho)n$ from the model point set, (2) applying a rigid transformation $(\mathbf{R},\mathbf{t})$ to the template, (3) perturbing the points of the template with noise (of strength ϵ), and (4) adding $(\tau-\rho)n$ spurious, uniformly distributed points to the template. Thus, after corruption, a template point set will have a total of τn points, of which only ρn correspond to points in the model set. Since ICP is known to be prone to outliers, we only compare our method with the more robust KC method in terms of the sensitivity of noise and outliers. The comparison is done via a set of 2D experiments.*At each of several noise levels and outlier strengths, we generate five models and six corrupted templates from each model for a total of 30 pairs at each noise and outlier strength setting.* For each pair, we use our algorithm and the KC method to estimate the known rigid

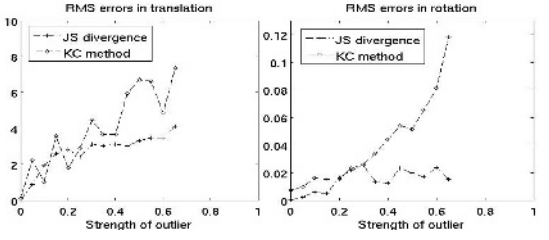

Fig. 2. Robustness to outliers in the presence of large noise. Errors in estimated rigid transform vs. proportion of outliers $((\tau-\rho)/(\rho))$ for both our method and KC method.

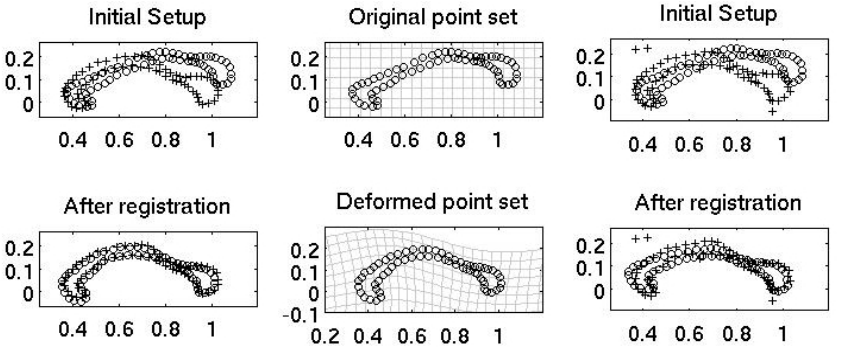

Fig. 3. Nonrigid registration of the corpus callosum data. Left column: two manually segmented corpus callosum slices before and after registration; Middle column: warping of the 2D grid using the recovered motion; Top right: same slices with one corrupted by noise and outliers, before and after registration.

transformation which was partially responsible for the corruption. Results show when the noise level is low, both KC and the presented method have strong resistance to outliers. However, we observe that when the noise level is high, our method exhibits stronger resistance to outliers than the KC method, as shown in Figure 2. We also applied our algorithm to nonrigidly register medical datasets (2D point-sets). Figure 3 depicts some results of our registration method applied to a set of 2D corpus callosum slices with feature points manually extracted by human experts. Registration result is shown in the left column with the warping of 2D grid under the recovered motion which is shown in the middle column. Our non-rigid alignment performs well in the presence of noise and outliers (Figure 3 right column). For the purpose of comparison, we also tested the TPS-RPM program provided in [8] on this data set, and found that TPS-RPM can correctly register the pair without outliers (Figure 3 top left) but failed to match the corrupted pair (Figure 3 top right).

3.2 Atlas Construction Results

In this section, we begin with a simple but demonstrative example of our algorithm for 2D atlas estimation. After this example, we describe a 3D implementations on real hippocampal data sets. The structure we are interested in this experiment is the corpus callosum as it appears in MR brain images. Constructing an atlas for the corpus callosum and subsequently analyzing the individual shape variation from "normal" anatomy has been regarded as potentially

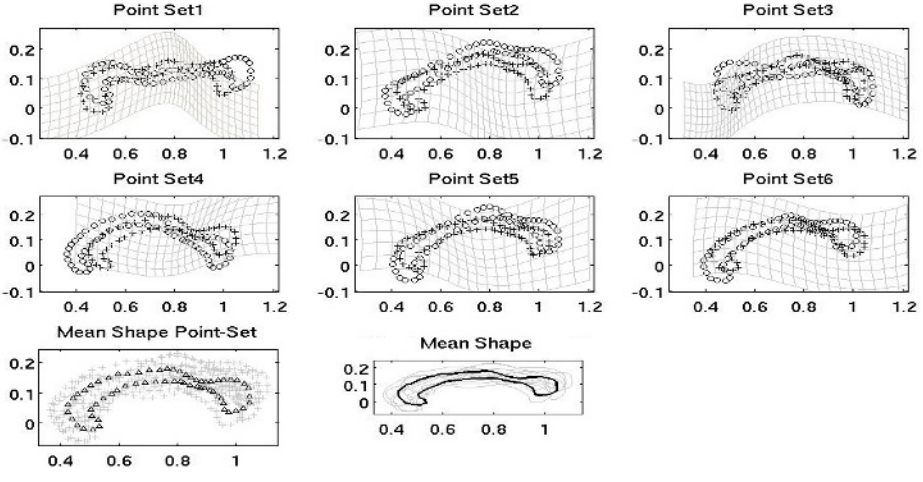

Fig. 4. Experiment results on 6 2D corpus collasum point sets. The first two rows shows the deformation of each point-set to the atlas, superimposed with initial point set (show in 'o') and deformed point-set (shown in '+'). Left image in the third row: The estimated atlas is shown superimposed over all the point-sets. Right: An atlas contour is traced and shown superimposed over all the original contours.

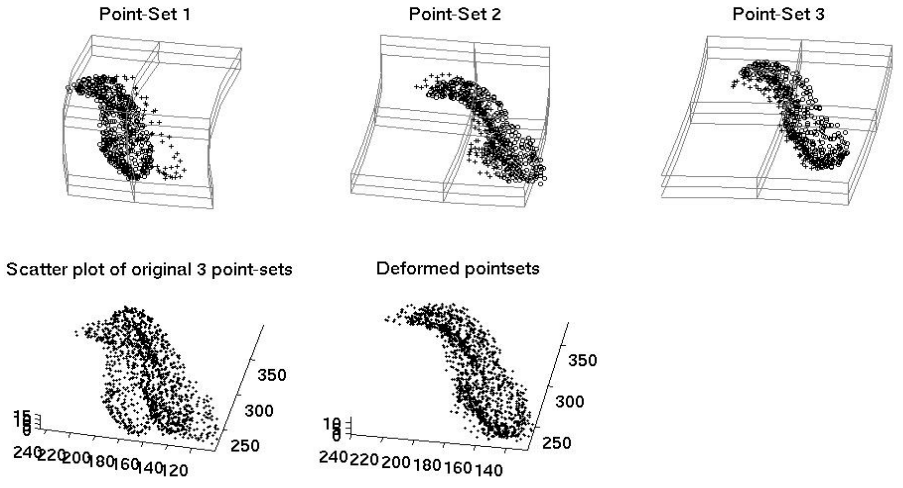

Fig. 5. Atlas construction from three 3D hipcampal point sets. The first row shows the deformation of each point-set to the atlas (represented as cluster centers), superimposed with initial point set (show in 'o') and deformed point-set (shown in '+'). Left image in the second row: Scatter plot of the original three hippocampal point-sets. Right: Scatter plot of all the warped point-sets.

valuable for the study of brain diseases such as agenesis of the corpus callosum(ACC), and fetal alcohol syndrome(FAS).

We manually extracted points on the outer contour of the corpus callosum from six normal subjects, (as shown Figure 4, indicated by "o"). The recovered deformation between each point-set and the mean shape are superimposed on the first two rows in Figure 4. The resulting atlas (mean point-set) is shown in third row of Figure 4, and is superimposed over all the point-sets. As we described earlier, all these results are computed simultaneously and automatically. This example clearly demonstrate that our joint matching and atlas construction algorithm can simultaneously align multiple shapes (modeled by sample point-sets) and compute a meaningful atlas/mean shape.

Next, we present results on 3D hippocampal point-sets. Three 3D point-sets were extracted from epilepsy patients with left anterior temporal lobe foci identified with EEG. An interactive segmentation tool was used to segment the hippocampus in the 3D anatomical brain MRI of the 3 subjects. The point-sets differ in shape, with the number of points $450, 421, 376$ in each point-set respectively. In the first row of Figure 5, the recovered nonrigid deformation between each hippocampal point-set to the atlas is shown along with a superimposition on all of the original data sets. In second row of the Figure 5, we also show the scatter plot of original point-sets along with all the point-sets after the non-rigid warping. An examination of the two scatter plots clearly shows the efficacy of our recovered non-rigid warping. Note that validation of what an atlas shape ought to be in the real data case is not feasible.

4 Conclusions

In this paper, we presented a novel and robust algorithm that utilize an information theoretic measure, namely Jensen-Shannon divergence, to simultaneously compute the mean shape from multiple unlabeled point-sets (represented by finite mixtures) and register them nonrigidly to this emerging mean shape. Atlas construction normally requires the task of non-rigid registration prior to forming the atlas. However, the unique feature of our work is that the atlas emerges as a byproduct of the non-rigid registration. Other advantages of using the JS-divergence over existing methods in literature for atlas construction and non-rigid registration is that, the JS-divergence is symmetric, is a metric and allows for use of unequal cardinality of the given point sets to be registered. The cost function optimization is achieved very efficiently by computing analytic gradients of the same and utilizing them in a quasi-Newton scheme. We compared our algorithm performance with competing methods on real and synthetic data sets and showed significantly improved performance in the context of robustness to noise and outliers in the data. Experiments were depicted with both 2D and 3D point sets from medical and non-medical domains. Our future work will focus on generalizing the non-rigid deformations to diffeomorphic mappings.

References

1. Small, C.: The Statistical theory of shape. Springer, New York (1996)
2. Lin, J.: Divergence measures based on the shannon entropy. IEEE Trans. Infor. Theory **37** (1991) 145–151
3. Hero, A., B. Ma, O.M., Gorman, J.: Applications of entropic spanning graphs. IEEE Trans. Signal Processing **19** (2002) 85–95
4. He, Y., Ben-Hamza, A., Krim, H.: A generalized divergence measure for robust image registration. IEEE Trans. Signal Processing **51** (2003) 1211–1220
5. Klassen, E., Srivastava, A., Mio, W., Joshi, S.H.: Analysis of planar shapes using geodesic paths on shape spaces. IEEE Trans. Pattern Anal. Mach. Intell. **26** (2003) 372–383
6. Tagare, H.: Shape-based nonrigid correspondence with application to heart motion analysis. IEEE Trans. Med. Imaging **18** (1999) 570–579
7. Bookstein, F.L.: Principal warps: Thin-plate splines and the decomposition of deformations. IEEE Trans. Pattern Anal. Mach. Intell. **11** (1989) 567–585
8. Chui, H., Rangarajan, A., Zhang, J., Leonard, C.M.: Unsupervised learning of an atlas from unlabeled point-sets. IEEE Trans. Pattern Anal. Mach. Intell. **26** (2004) 160–172
9. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. IEEE Trans. Pattern Anal. Mach. Intell. **24** (2002) 509–522
10. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models: their training and application. Comput. Vis. Image Underst. **61** (1995) 38–59
11. Hill, A., Taylor, C.J., Brett, A.D.: A framework for automatic landmark identification using a new method of nonrigid correspondence. IEEE Trans. Pattern Anal. Mach. Intell. **22** (2000) 241–251
12. Tsin, Y., Kanade, T.: A correlation-based approach to robust point set registration. In: ECCV2004(3). (2004) 558–569

13. Glaunes, J., Trouvé, A., Younes, L.: Diffeomorphic matching of distributions: A new approach for unlabelled point-sets and sub-manifolds matching. In: CVPR2004 (2). (2004) 712–718
14. McLachlan, G., Basford, K.: Mixture Model:Inference and Applications to Clustering. Marcel Dekker, New York (1988)
15. Endres, D.M., Schindelin, J.E.: A new metric for probability distributions. IEEE Trans. Inf. Theory **49** (2003) 1858–60
16. Chui, H., Rangarajan, A.: A new point matching algorithm for non-rigid registration. Computer Vision and Image Understanding (CVIU) **89** (2003) 114–141