Simultaneous Nonrigid Registration of Multiple Point-Sets and Atlas Construction

Fei Wang, Baba C. Vemuri, Anand Rangarajan and Stephan J. Eisenschenk

***Abstract*— Group-wise registration of a set of shapes represented by unlabeled point-sets is a challenging problem since, usually this involves solving for point correspondence in a non-rigid motion setting. In this paper, we propose a novel and robust algorithm that is capable of simultaneously computing the mean shape—represented by a probability density function—from multiple unlabeled point-sets (represented by finite mixture models) and registering them non-rigidly 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 of Gaussian densities. We motivate the use of the JS divergence by pointing out its close relationship to hypothesis testing. Essentially, minimizing the JS divergence is asymptotically equivalent to maximizing the likelihood ratio formed from a probability density of the pooled point-sets and the product of the probability densities of the individual point-sets. 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. A by product of the registration process is a probabilistic atlas which is defined as the convex combination of the probability densities of the input point sets being aligned. 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 (in vision and graphics applications) without having to establish any correspondence. We present experimental results on non-rigidly registering 2D as well as 3D real and synthetic data (point sets).**

I. Introduction

In recent years, there has been considerable interest in the application of statistical shape analysis to problems in medical image analysis 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 representation since statistical analysis of point-set representation of shapes is very mature [1]. Means, covariances and probability distributions on manifolds constructed from point sets can now be defined and estimated [1].

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 the objects under consideration 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 alignment [2]. Furthermore, we also encounter the issue of the bias problem in registering two or more data sets, this is a significant issue in the Atlas construction problem. Atlas construction here entails creation of a representative of a population of point data sets each of which represents a 3D/2D shape. Before creating such a representative, one needs to register the data sets. 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? Once the registration is achieved, the representative, atlas, is generally taken to be some sort of an average of the aligned point sets.

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 probability density function, then quantify the distance between these probability densities 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 density functions are transformed through appropriate changes in their parameters, the distance measure between these densities would be minimized since all these densities should be similar to each other. We impose regularization on each deformation field to prevent large transformations on each density representing the point-sets. Jensen-Shannon divergence, first introduced in [3], serves as a model divergence measure between multiple probability densities, researchers have used it as a dissimilarity measure for image registration and retrieval applications in the past (Hero et al. [4], He et al. [5] and Chiang et al. [6]), *but never for registration of two or more point sets*. It has some very desirable properties. To name a few, i) the square root of JS-divergence (in the case when its (convex combination) parameter is fixed to $\frac{1}{2}$) is a metric [7], ii) the JS-divergence relates to other information-theoretical functionals, such as the relative entropy or the Kullback-Leibler (KL) divergence, and hence it shares their mathematical properties as well as their intuitive interpretability, and iii) the compared densities can be weighted, which allows us to take into account the different size of the point samples from which the probability densities are computed. Some of these advantages will be explicitly brought to limelight in subsequent sections during the course of description of our method.

The rest of this paper is organized as follows. The remainder

F. Wang is with IBM Almaden Research Center, 650 Harry Road, San Jose, CA 95120. B. C. Vemuri and A. Rangarajan are with Department of CISE, University of Florida, Gainesville, FL 32611. S. J. Eisenschenk is with Department of Neurology, University of Florida, Gainesville, FL 32611.

This research was in part funded by the NIH grant RO1 NS046812 and NSF grant NSF 0307712.

of section (I) gives a brief review of the literature, focusing on differences between existing methods and ours. Section II contains a description of the formulation using JS-divergence for our simultaneous nonrigid registration and atlas construction model as well as a derivation of the associated gradient of the energy function. Section III contains a description of the computational techniques employed in our algorithm for simultaneous non-rigid registration and atlas construction. Experimental results on 2D as well as 3D point-sets are presented in Section IV. Finally, we draw conclusions in section V.

A. 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. Based on the shape representation used, 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 intrinsically or extrinsically parameterized (e.g. using point locations and spline coefficients).

The work presented in Klassen et al. [8] 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 minimizes the sum of squared geodesic distances between this unknown point to each individual shape, which lies on the manifold. However, the curves are limited to be closed curves, and it has not been extended to the 3D surface shapes. For methods using intrinsic curve or surface representations [8], [9], [10], further statistical analysis on them is much more difficult than analysis on the point representation, but the reward maybe higher due to the use of the intrinsic higher order representation.

Among the methods using point-sets parameterization, the idea of using nonrigid spatial mapping functions, specifically thin-plate splines [11], [12], [13], to analyze deformable shape has been widely adopted. Bookstein's work in [11], 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 Rangarajan et al. using thin-plate splines [12], [14], [15]. In their recent work [12], they extend their work to the construction of an mean shape from a set of unlabeled shapes which are represented by unlabeled point-sets. 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, and the generated mean shape is entirely symmetric with no bias toward any of the original shapes. Garcin et al. [16] attempt to solve the unlabeled point-set averaging problem using a similar idea as in [12], except that it's under the diffeomorphism setting, and consequently the estimated distance between point-sets are geodesic distances. However, in both of these works, the annealing procedure results in a slow algorithm. Unlike their approaches, 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 [17] utilized points to represent deformable shapes. Their work pioneered the efforts in building point distribution models to understand deformable shapes [17], [18]. 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 [19], they attempt to overcome this limitation by trying to automatically solve for the correspondences in a nonrigid setting. The resulting algorithm is very similar to the earlier work in [10] and is restricted to curves. The work in [2] also uses 2D points to learn shape statistics, which is quite similar to the active shape model method except that more attention has been paid to the sample point-sets generation process from the shape. Unlike our method, the transformation between curves is limited to rigid mapping, and the registration process is not symmetric.

There are several articles on point-sets alignment in recent literature which bear close relation to our research reported here. For instance, Tsin et al. [20] proposed a kernel correlation based point-set registration approach where the cost function is proportional to the correlation of two kernel density estimates. In [21], Jian et al. introduced a novel and robust algorithm for rigidly and non-rigidly registering pairs of data sets using the L2 distance between mixtures of Gaussians representing the point-set data. They derived a closed form expression for the L2 distance between the mixtures of Gaussians and used it in their registration algorithm. Therefore. their algorithm is very fast in comparison to existing methods on point-set registration and the results shown are quantitatively satisfactory. However, they do not actually fit a mixture density to each point-set choosing instead to allow each point in each set to be a cluster center. Consequently, their method is actually more similar to the image matching method of [22] discussed below but with the advantage of not having to evaluate the cost function involving spatial integrals numerically since, a closed form expression is derived for the same. Their method however has not been extended to the problem of unbiased matching of multiple point-sets being addressed in this paper. Perhaps closest to our work is the approach of Wang et al. [23] where an the relative entropy measure (Kullback-Leibler distance) is used to find a similarity transformation between two point-sets. The Kullback-Leibler distance, as we know, is a special case of our proposed JS-

divergence for two random variables [3], and the approach in [23] only tackles the pairwise rigid matching problem. These methods are 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 measures used in [20], [21], [23], and can easily cope with multiple point-sets. Recently, in [22], Glaunes et al. represent points as delta functions and match them using the dual norm in a reproducing kernel Hilbert space. The main problem for this technique is that they need a 3D spatial integral which must be numerically computed. In contrast, we compute the JS-divergence using an empirical framework where the computations converge in the limit to the true values. We will show that our method when applied to match point-sets, achieves very good performance in terms of both robustness and accuracy.

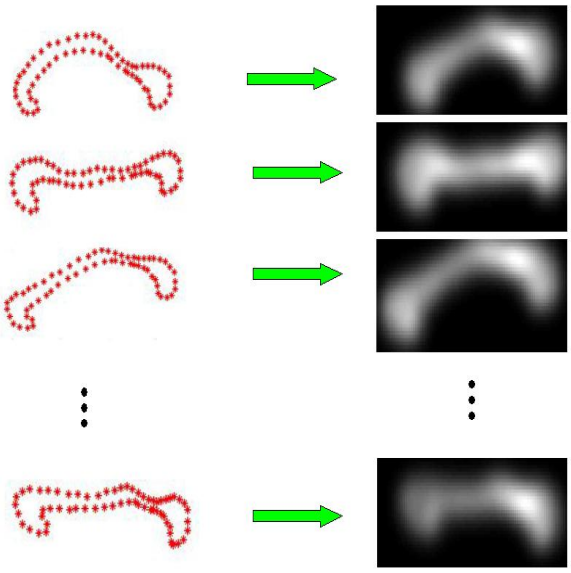

Fig. 1. Illustration of corpus callosum shapes (point-sets) represented as density functions.

II. Mathematical Formulation

In this section, we present the mathematical formulation of our simultaneous non-rigid registration and atlas construction method. Note that normally, non-rigid registration precedes atlas construction since the later requires the data to be registered. However, in our work, the atlas emerges as a by product of the non-rigid registration and hence is not achieved in the aforementioned, traditional sequential order. The basic idea is to model each point-set by a probability density function, then quantify the distance between these probability densities using an information-theoretic measure. Figure 1 illustrates this idea, wherein the right column of the figure depicts the density functions corresponding to the point-sets drawn from a cortical substructure in the human brain called the corpus callosum shown in the left column. The dissimilarity measure between these density functions is then optimized over a space of coordinate transformations yielding the desired transformations. We will begin by presenting the finite mixture of Gaussian densities used to model the probability densities of the given point-sets.

A. The Finite Mixture Models

Given the fact that nonrigid point matching is fraught with the problems of noise, outliers and deformations with unknown parameterizations, it is natural to use probability distributions to model each point-set. We also think that it is natural to use a finite Gaussian mixture model as the representation of a point-set. As the most frequently used mixture model, a Gaussian mixture [24] is defined as a convex combination of Gaussian component densities.

We use the following notation: The data point-sets are denoted by $\{X^c, c \in \{1, ..., N\}\}$. Each point-set X^c consists of points $\{x_i^c \in \mathbb{R}^d, i \in \{1, ..., n_c\}\}$. 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 $\{V^c, p \in \{1, ..., N\}\}$. Each cluster point-set V^c consists of points $\{v_a^c \in \mathbb{R}^d, a \in \{1, ..., K^c\}\}$. Note that there are K^c points in each V^c, 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^c, and we only need to do this once before the process of joint atlas estimation and point-sets registration. The atlas point-set is denoted by Z. We begin by specifying the density function of each point-set.

$$\mathbf{P}_c(x) = \sum_{a=1}^{K^c} \alpha_a^c p(x|v_a^c) \tag{1}$$

In Eqn. (1), the occupancy probability which is different for each data point-set is denoted by α^c. The component densities $p(x|v_a^c)$ is

$$\begin{aligned} p(x|v_a^c) &= G(x - v_a^c, \Sigma_a) \\ &= \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma_a|^{\frac{1}{2}}} \exp\Big(-\frac{1}{2}(x - v_a^c)^T \Sigma_a^{-1} (x - v_a^c)\Big) \end{aligned} \tag{2}$$

where $G(x - v_a^p, \Sigma_a)$ is the Gaussian Kernel in d-dimensional space, and $|\cdot|$ denotes the determinant. Probability of the point-set X^c coming from this mixture is then

$$\Pr(X^c|V^c, \alpha^c) = \prod_{i=1}^{n_c} \mathbf{P}_p(x_i^c) = \prod_{i=1}^{n_c} \sum_{a=1}^{K^c} \alpha_a^c p(x_i^c|v_a^c) \tag{3}$$

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 the 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 densities represented by Gaussian mixtures? We answer this in the following section.

B. Jensen-Shannon Divergence for Learning the Atlas

The Jensen-Shannon (JS) divergence, first introduced in [3], serves as a measure of cohesion between multiple probability densities. It has been used by some researchers as a dissimilarity measure for image registration and retrieval applications [4], [5]. This dissimilarity measure has some very desirable properties; to name a few, i) The square root of JS-divergence (in the case when its parameter is fixed to $\frac{1}{2}$) is a metric [7], ii) the JS-divergence relates to other information-theoretic functionals, such as the relative entropy or the Kullback-Liebler (KL) divergence, and hence it shares their mathematical properties as well as their intuitive appeal, iii) the probability densities compared 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 densities are computed, iv) 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 and Rangarajan [14]. Given n probability distributions $\mathbf{P}_i$, $i \in \{1, ..., n\}$, the JS-divergence of $\mathbf{P}_i$ is defined by

$$JS_\pi(\mathbf{P}_1, \mathbf{P}_2, ..., \mathbf{P}_n) = H(\sum \pi_i \mathbf{P}_i) - \sum \pi_i H(\mathbf{P_i}) \qquad (4)$$

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 $H(\mathbf{P}_i)$ is the Shannon entropy. The two terms on the right hand side of Eqn. (4) 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^c is related to Z via a function f^c and $\boldsymbol{\mu}^c$ is the set of the transformation parameters associated with each function f^c. The densities of the deformed point-sets can be written as $\mathbf{P}_c = \mathbf{P}_c(x|V^c, \boldsymbol{\mu}^c) = \sum_{a=1}^{K^c} \alpha_a^c p(x|f^c(v_a^c))$. To compute the mean shape density a.k.a. the probabilistic atlas, from these point-sets and register them to the emerging mean shape density, we need to recover the transformation parameters $\boldsymbol{\mu}^c$. This problem can be modeled as an optimization problem with the objective function being the JS-divergence between the densities of the deformed point-sets, represented as $\mathbf{P}_c = \mathbf{P}_c(x|V^c, \boldsymbol{\mu}^c)$, the probabilistic atlas construction problem can now be formulated as,

$$\begin{aligned} &\min_{\boldsymbol{\mu}^c} JS_\pi(\mathbf{P}_1, \mathbf{P}_2, ..., \mathbf{P}_N) + \lambda \sum_{c=1}^{N} ||Lf^c||^2 \\ &= \min_{\boldsymbol{\mu}^c} H(\sum_c \pi_c \mathbf{P}_c) - \sum_c \pi_c H(\mathbf{P}_c) + \lambda \sum_{c=1}^{N} ||Lf^c||^2 \end{aligned} \qquad (5)$$

In Eqn. (5), the weight parameter λ is a positive constant and 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 [14], 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 $U(r) = 1/(8\pi) r^2 ln(r)$.

Thus, the cost function for non-rigid registration can be formulated as an energy functional in a regularization framework, where the regularization term in Eqn. (5) 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 are 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 $\mathbf{W}$.

Next we will present some properties of the Jensen-Shannon divergence in the context of groupwise point-sets registration.

C. JS Divergence in a Hypothesis Testing Framework

In this section, we show that the Jensen-Shannon divergence can be interpreted in the statistical framework of hypothesis testing. We first give an intuitive presentation followed by a more formal one. Assume for the moment that we have only two point-sets $X^{(1)}$ and $X^{(2)}$ that need to be registered. We construct the following hypothesis test. For any given nonrigid transformation, consider the following two hypotheses for the pooled point-set $X = X^{(1)} \cup X^{(2)}$. The null hypothesis is that the samples $X^{(1)}$ and $X^{(2)}$ are independent but drawn from two different distributions $\mathbf{P}_1$ and $\mathbf{P}_2$ respectively. The alternative hypothesis is that the samples $X = X^{(1)} \cup X^{(2)}$ are independent and drawn from a pooled distribution $\mathbf{P}$. The likelihood ratio for this hypothesis test is

$$\Lambda = \frac{\prod_{k=1}^{n_1+n_2} \mathbf{P}(X_k)}{\prod_{k_i=1}^{n_1} \mathbf{P}_1(X_{k_1}^{(1)}) \prod_{k_2=1}^{n_2} \mathbf{P}_2(X_{k_2}^{(2)})} \qquad (6)$$

where n_1 is the number of instances from point set $X^{(1)}$ and n_2 is the number of instances from $X^{(2)}$. It should be understood that the distribution $\mathbf{P}$ is a maximum likelihood estimate over the n_1+n_2 samples drawn from $X = X^{(1)} \cup X^{(2)}$ and that the distributions $\mathbf{P}_1$ and $\mathbf{P}_2$ are maximum likelihood estimates over the n_1 samples drawn from $X^{(1)}$ and the n_2 samples drawn from $X^{(2)}$ respectively. Furthermore, the same set of samples are used in the numerator and denominator of (6) with the main difference being that the identity of the point-set $X^{(1)}$ or $X^{(2)}$ is erased when considering samples from the pooled point-set $X = X^{(1)} \cup X^{(2)}$. We then maximize the likelihood ratio Λ over the set of nonrigid transformations. Maximizing the likelihood ratio corresponds to favoring the likelihood of a single pooled distribution $\mathbf{P}$ over the product of the likelihoods of separate distributions $\mathbf{P}_1$ and $\mathbf{P}_2$ evaluated at the *same* set of samples.

It can be shown that the likelihood ratio asymptotically converges to the Jensen-Shannon (JS) divergence when the distribution $\mathbf{P}$ above is modeled as a mixture $\pi_1\mathbf{P}_1 + \pi_2\mathbf{P}_2$ with $\pi_1 = \frac{n_1}{n_1+n_2}$ and $\pi_2 = \frac{n_2}{n_1+n_2}$. This approach has the advantage that we do not need a separate model for the overlay. More formally (and moving over to the case of multiple point-sets), we construct a likelihood ratio between i.i.d. samples drawn from a mixture ($\sum_a \pi_a \mathbf{P}_a$) with $\pi_a = \frac{n_q}{\sum_b n_b}$ and i.i.d. samples drawn from a heterogeneous collection of densities ($\mathbf{P}_1, \mathbf{P}_2, ..., \mathbf{P}_N$) with the samples being indexed by the specific member densities in the family from which they are drawn. Assume that n_1 samples are drawn from $\mathbf{P_1}$, n_2 from $\mathbf{P_2}$ etc. Let the total number of pooled samples be defined as $M \stackrel{\text{def}}{=} \sum_{a=1}^{N} n_a$. The likelihood ratio then is

$$\Lambda = \frac{\prod_{k=1}^{M} \sum_{a=1}^{N} \pi_a \mathbf{P}_a(x_k)}{\prod_{a=1}^{N} \Pi_{k_a=1}^{n_a} \mathbf{P}_a(x_{k_a}^a)} \tag{7}$$

where x_k consists of points $\{x_{k_a}^a, k_a \in \{1, ..., n_a\} a \in \{1, ..., N\}\}$, which is the pooled data of all the samples. In contrast to the typical statistical test relative to a threshold, we seek the maximum of the likelihood ratio in Eqn. (7). The following theorem shows the relationship between Jensen-Shannon divergence and the above likelihood ratio.

Theorem 1: Given N probability distributions $\mathbf{P}_a, a \in \{1, ..., N\}$, maximizing the hypothesis ratio in Eqn. (7) is equivalent to minimizing the Jensen-Shannon divergence between the N probability distributions $\mathbf{P}_a$, $a \in \{1, ..., N\}$.

Proof: Taking the negative logarithm of the likelihood ratio, we have,

$$\begin{aligned} & -\log \Lambda \\ &= -\sum_{k=1}^{M} \log \sum_{a=1}^{N} \pi_a \mathbf{P}_a(x_k) + \sum_{a=1}^{N} \log \prod_{k_a=1}^{n_a} \mathbf{P}_a(x_{k_a}^a) \\ &= -\sum_{k=1}^{M} \log \sum_{a=1}^{N} \pi_a \mathbf{P}_a(x_k) + \sum_{a=1}^{N} \sum_{k_a=1}^{n_a} \log \mathbf{P}_a(x_{k_a}^a) \end{aligned} \tag{8}$$

We now apply the law of large numbers after assuming that the individual point-set counts $\{n_a\}$ are large enough. We get

$$\begin{aligned} -\log \Lambda &= MH(\sum_{a=1}^{N} \pi_a \mathbf{P}_a) - \sum_{a=1}^{N} n_a H(\mathbf{P}_a) \\ &= M[H(\sum_{a=1}^{N} \pi_a \mathbf{P}_a) - \sum_{a=1}^{N} \frac{n_a}{M} H(\mathbf{P}_a)] \\ &= M[JS_\pi(\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_N)]. \end{aligned} \tag{9}$$

■

The interpretation of minimizing the JS divergence as a type of hypothesis testing has intuitive appeal for us. Maximizing the likelihood ratio above means that we favor a maximum likelihood explanation of fitting a mixture to the pooled data rather than separately fitting mixtures to the individual point-sets. Please note that in our groupwise registration approach, the warping is not between a source and a fixed target. Instead, the warping is performed on the parameters of the original mixtures such that the likelihood ratio is maximized.

D. JS Divergence is unbiased

Typically we are required to construct an atlas from very large number of point-sets, and this process will usually take a long time since the computational complexity grows polynomially with the number of point-sets (N) that we want to register. We now introduce a hierarchical registration technique that significantly reduces the computational complexity.

Assume that we are given N point-sets, from which we are required to construct the probabilistic atlas. We divide the N point-sets into m subsets (generally $m \ll N$), therefore, we can construct m probabilistic atlases from these subsets using our algorithms, and all the point-sets inside each of the subsets are registered. Then, we can either construct a single atlas from these m atlases, or we can further divide the m atlas point-sets into even smaller subsets, and follow the same process until a single atlas is constructed. The remaining question is whether the atlas constructed this way is biased or not? The following theorem will help to give us the answer with the exclusion of the thin-plate spline part of the cost function.

Theorem 2: Given N probability densities $\mathbf{P}_a, a \in \{1, ..., N\}$, each with a weight π_a in the Jensen-Shannon divergence. Let us divide these set of N densities into m subsets, such that i^{th} subset contains n_i densities $\mathbf{P}_a, a \in \{k_1^{(i)}, k_2^{(i)}, ..., k_{n_i}^{(i)}\}$, and $\sum_i n_i = N$. Assume $\mathbf{S}_i$ is the convex combination of all the densities, the i^{th} subset, with the weights $\frac{\pi_{k^{(i)}}}{\beta_i}$, where $\beta_i = \sum_j \pi_{k_j^{(i)}}$, i.e. $\mathbf{S}_i = \sum_{j=1}^{n_i} \pi_{k_j^{(i)}} \mathbf{P}_{k_j^{(i)}} / \beta_i$. The JS divergence of the $\mathbf{P}_a$s and JS divergence of the $\mathbf{S}_i$s are then related by:

$$\begin{aligned} & JS_\pi(\mathbf{P}_1, \mathbf{P}_2, \cdots, \mathbf{P}_N) - JS_\beta(\mathbf{S}_1, \mathbf{S}_2, \cdots, \mathbf{S}_m) \\ &= \sum_{i=1}^{m} \beta_i JS_{\frac{\pi_{k^{(i)}}}{\beta_i}}(\mathbf{P}_{k_1^{(i)}}, \mathbf{P}_{k_2^{(i)}}, \cdots, \mathbf{P}_{k_{n_i}^{(i)}}) \end{aligned} \tag{10}$$

Proof: We can rewrite $JS_\beta(\mathbf{S}_1, \mathbf{S}_2, \cdots, \mathbf{S}_m)$ as

$$\begin{aligned}
&JS_\beta(\mathbf{S}_1, \mathbf{S}_2, \cdots, \mathbf{S}_m) \\
&= H(\sum_{i=1}^{m} \beta_i \mathbf{S}_i) - \sum_{i=1}^{m} \beta_i H(\mathbf{S}_i) \\
&= H(\sum_{i=1}^{m} \beta_i \sum_{j=1}^{n_i} \frac{\pi_{k_j^{(i)}} \mathbf{P}_{k_j^{(i)}}}{\beta_i}) - \sum_{i=1}^{m} \beta_i H(\mathbf{S}_i) \\
&= H(\sum_{i=1}^{N} \pi_i \mathbf{P}_i) - \sum_{i=1}^{m} \beta_i H(\mathbf{S}_i)
\end{aligned}$$

Therefore, the left-hand side of Eqn. (10) can be rewritten as,

$$\begin{aligned}
&JS_\pi(\mathbf{P}_1, \mathbf{P}_2, \cdots, \mathbf{P}_N) - JS_\beta(\mathbf{S}_1, \mathbf{S}_2, \cdots, \mathbf{S}_m) \\
&= \sum_{i=1}^{m} \beta_i H(\mathbf{S}_i) - \sum_{i=1}^{N} \pi_i H(\mathbf{P}_i)
\end{aligned}$$

Meanwhile, right-hand side of Eqn. (10) can be rewritten as,

$$\begin{aligned}
&\sum_{i=1}^{m} \beta_i JS_{\frac{\pi_{k^{(i)}}}{\beta_i}}(\mathbf{P}_{k_1^{(i)}}, \mathbf{P}_{k_2^{(i)}}, \cdots, \mathbf{P}_{k_{n_i}^{(i)}}) \\
&= \sum_{i=1}^{m} \beta_i \Big[H(\sum_{j=1}^{n_i} \frac{\pi_{k_j^{(i)}}}{\beta_i} \mathbf{P}_{k_j^{(i)}}) - \sum_{j=1}^{n_i} \frac{\pi_{k_j^{(i)}}}{\beta_i} H(\mathbf{P}_{k_j^{(i)}}) \Big] \\
&= \sum_{i=1}^{m} \Big[\beta_i H(\mathbf{S}_i) - \sum_{j=1}^{n_i} \pi_{k_j^{(i)}} H(\mathbf{P}_{k_j^{(i)}}) \Big] \\
&= \sum_{i=1}^{m} \beta_i H(\mathbf{S}_i) - \sum_{i=1}^{N} \pi_i H(\mathbf{P}_i)
\end{aligned}$$

which is exactly the same as the left-hand side of Eqn. (10). ■

In our registration algorithm, all the point-sets are represented as probability densities, and the atlas constructed can be considered as convex combination of these densities. Therefore, we can treat $\mathbf{P}_a$s and $\mathbf{S}_i$s as the densities corresponding to the point-sets and the constructed atlases from the subsets respectively. Therefore, from Theorem 2, we know relationship in Eqn. (10) holds between the JS divergence of the $\mathbf{P}_a$s and $\mathbf{S}_i$s. Notice that the righthand side of Eqn. (10) is the JS divergences of the densities in all the subsets, which are minimized in each step of the hierarchical technique we proposed here. Intuitively, if these point-sets are aligned properly, the corresponding density functions should be statistically similar. Therefore the JS divergences of all the subsets should be zero or very close to zero, which means the right hand side of Eqn. (10) is zero. Consequently, the JS divergence of the $\mathbf{P}_a$s and JS divergence of the $\mathbf{S}_i$s are equal to each other. Therefore, minimizing JS divergence of all the resulting atlas point-sets is equivalent to minimizing the JS divergence of the original point-sets.

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

III. Estimating Empirical JS

In Eqn. (5), $\mathbf{P}_i = \mathbf{P}_i(x|V^i, \boldsymbol{\mu}^i) = \sum_{a=1}^{K_i} \alpha_a^i p(x|f^i(v_a^i)) = \frac{1}{K_i} \sum_{a=1}^{K_i} G(x - f^i(v_a^i), \Sigma_a) = \frac{1}{K_i} \sum_{a=1}^{K_i} G(x - f^i(v_a^i), \sigma^2 \mathbf{I})$, where we assume that the occupancy probabilities are uniform ($\alpha_a^i = \frac{1}{K_i}$) and the covariance matrices Σ_i are isotropic, diagonal, and identical $\Sigma_a = \sigma^2 I$. For simplicity, we denote deformed cluster centers as $u_a^i := f^i(v_a^i)$. We can then generate q_i random samples $s_1^{(i)}, s_2^{(i)}, ..., s_{q_i}^{(i)}$ from the mixture $\mathbf{P}_i$. $Q = \sum_i q_i$ is the total number of random samples from all N densities functions $\mathbf{P}_i, \forall i = \{1, 2, ..., N\}$, $\{s_1, s_2, ..., s_Q\} \equiv \{s_1^{(1)}, ..., s_j^{(i)}, ..., s_{q_N}^{(N)}\}$ are the pooled random samples. We have the estimation of the Shannon entropy for $\mathbf{P}_i$ using the weak law of large numbers,

$$\begin{aligned}
H(\mathbf{P}_i) &= -\frac{1}{q_i} \sum_{j=1}^{q_i} \log \mathbf{P}_i(s_j^{(i)}) \\
&= -\frac{1}{q_i} \sum_{j=1}^{q_i} \log \Big[\frac{1}{K_i} \sum_{a=1}^{K_i} G(s_j^{(i)} - f^i(v_a^i), \sigma^2 \mathbf{I}) \Big] \\
&= -\frac{1}{q_i} \sum_{j=1}^{q_i} \log \Big[\frac{1}{K_i} \sum_{a=1}^{K_i} G(s_j^{(i)} - u_a^i, \sigma^2 \mathbf{I}) \Big]
\end{aligned} \tag{11}$$

For the convex combination $\sum \pi_i \mathbf{P}_i$, if we choose $\pi_i = \frac{K_i}{M}$, where $M = \sum_i K_i$ is the total number of the cluster centers in the N point-sets that we want to register, we have the following,

$$\begin{aligned}
\sum_{i=1}^{N} \pi_i \mathbf{P}_i &= \sum_{i=1}^{N} \pi_i \sum_{a=1}^{K_i} \frac{1}{K_i} G(x - u_a^i), \sigma^2 \mathbf{I}) \\
&= \frac{1}{M} \sum_{i=1}^{N} \sum_{a=1}^{K_i} G(x - u_a^i), \sigma^2 \mathbf{I}) \\
&= \frac{1}{M} \sum_{j=1}^{M} G(x - u_j, \sigma^2 \mathbf{I})
\end{aligned} \tag{12}$$

where $\{u_1, u_2, ..., u_M\} \equiv \{u_1^1, ..., u_j^i, ..., u_{K_N}^N\}$ are the pooled cluster centers. Therefore the linear combination of the GMMs can be expressed as a single Gaussian Mixture centered on the deformed cluster centers. Consequently, we have the Shannon entropy estimation of the $\sum \pi_i \mathbf{P}_i$,

$$\begin{aligned}
H(\sum_{i=1}^{N} \pi_i \mathbf{P}_i) &= H(\frac{1}{M} \sum_{j=1}^{M} G(x - u_j, \sigma^2 \mathbf{I})) \\
&= -\frac{1}{Q} \sum_{j=1}^{Q} \log \Big[\frac{1}{M} \sum_{a=1}^{M} G(s_j - u_a, \sigma^2 \mathbf{I}) \Big]
\end{aligned} \tag{13}$$

Combine the two terms in Eqn. (11,13) together, we have,

$$\begin{aligned}
&JS_\pi(\mathbf{P}_1, \mathbf{P}_2, ..., \mathbf{P}_N) \\
&= H(\sum \pi_i \mathbf{P}_i) - \sum \pi_i H(\mathbf{P}_i) \\
&= -\frac{1}{Q} \sum_{j=1}^{Q} \log \Big[\frac{1}{M} \sum_{a=1}^{M} G(s_j - u_a, \sigma^2 \mathbf{I}) \Big] \\
&\quad + \sum_{i=1}^{N} \frac{K_i}{q_i M} \sum_{j=1}^{q_i} \log \Big[\frac{1}{K_i} \sum_{a=1}^{K_i} G(s_j^{(i)} - u_a^i, \sigma^2 \mathbf{I}) \Big]
\end{aligned} \tag{14}$$

A. Cost function optimization

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):

$$\nabla JS = [\frac{\partial JS}{\partial \boldsymbol{\mu}^1}, \frac{\partial JS}{\partial \boldsymbol{\mu}^2}, ..., \frac{\partial JS}{\partial \boldsymbol{\mu}^N}] \tag{15}$$

Each component of the gradient may be found by differentiating Eqn. (14) with respect to the transformation parameters. In order to compute this gradient, let's first calculate the derivative of $G(s_j^{(i)} - u_a^i, \sigma^2\mathbf{I})$ with respect to $\boldsymbol{\mu}^i$,

$$\begin{aligned}
&\frac{\partial G(s_j^{(i)} - u_a^i, \sigma^2\mathbf{I})}{\partial \boldsymbol{\mu}^i} \\
&= \frac{1}{\sigma^2} G(s_j^{(i)} - u_a^i, \sigma^2\mathbf{I})(s_j^{(i)} - u_a^i)\frac{\partial u_a^i}{\partial \boldsymbol{\mu}^i} \\
&= \frac{1}{\sigma^2} G(s_j^{(i)} - u_a^i, \sigma^2\mathbf{I})(s_j^{(i)} - u_a^i)\frac{\partial f^i(v_a^i, \boldsymbol{\mu}^i)}{\partial \boldsymbol{\mu}^i}
\end{aligned} \tag{16}$$

Based on this, it is straight forward to derive the gradient of the JS-divergence in Eqn. (14) with respect to the transformation parameters $\boldsymbol{\mu}^i$, which is given by

$$\begin{aligned}
&\frac{\partial JS}{\partial \boldsymbol{\mu}^i} \\
&= -\frac{1}{Q}\sum_{j=1}^{Q} \frac{1}{\sum_{a=1}^{M} G(s_j - u_a, \sigma^2\mathbf{I})} \sum_{a=1}^{M} \frac{\partial G(s_j - u_a, \sigma^2\mathbf{I})}{\partial \boldsymbol{\mu}^i} \\
&+ \frac{K_i}{q_i M}\sum_{j=1}^{q_i} \frac{1}{\sum_{a=1}^{K_i} G(s_j^{(i)} - u_a^i, \sigma^2\mathbf{I})} \sum_{a=1}^{K_i} \frac{\partial G(s_j^{(i)} - u_a^i, \sigma^2\mathbf{I})}{\partial \boldsymbol{\mu}^i} \\
&= -\frac{1}{Q}\sum_{j=1}^{Q} \frac{1}{\sum_{a=1}^{M} G(s_j - u_a, \sigma^2\mathbf{I})} \sum_{a=1}^{K_i} \frac{\partial G(s_j - u_a^i, \sigma^2\mathbf{I})}{\partial \boldsymbol{\mu}^i} \\
&+ \frac{K_i}{q_i M}\sum_{j=1}^{q_i} \frac{1}{\sum_{a=1}^{K_i} G(s_j^{(i)} - u_a^i, \sigma^2\mathbf{I})} \sum_{a=1}^{K_i} \frac{\partial G(s_j^{(i)} - u_a^i, \sigma^2\mathbf{I})}{\partial \boldsymbol{\mu}^i}
\end{aligned} \tag{17}$$

B. Algorithm Summary

Our simultaneous atlas construction and registration algorithm can be summarized as follows:

Given N point-sets $\{X^c, c \in \{1, ..., N\}\}$.

1) Estimate the cluster centers $\{V^c, c \in \{1, ..., N\}\}$ for each point X^c. In our implementation, we utilize deterministic annealing (DA) procedure with its proven benefit of robustness in clustering [25].

2) Set initial transformation parameters $\boldsymbol{\mu}^c$ to zero and optimize the cost function in Eqn. (5) with respect to the transformation parameter $\boldsymbol{\mu}^c$. Since the analytic gradients with respect to these transformation parameters have to be explicitly derived in Eqn. (17), 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. The samples $\{s_j^i\}$ from the mixture $\mathbf{P}_i$ are re-drawn every couple of iterations. We currently have two implementations of our registration algorithm using the Matlab Optimization toolbox: one with gradients explicitly computed and one without. Experiments show that results on datasets with large non-rigid deformations show the version with analytic gradients converges faster than the one without.

3) The successful registration process ensures that the deformed point-sets are close to each other. Therefore, we can apply one of the recovered deformation to the corresponding point-sets to recover the mean shape.

Note that our transformation model can be any type of geometric transformations, e.g. rigid, affine, polynomial or other types of nonrigid transformations. Therefore 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.

For a typical atlas construction problem, an affine registration of the multiple point-sets precedes the nonrigid registration to bring the point-sets relatively closer to each other, which will speed up the nonrigid registration process significantly. 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.

IV. 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.

A. 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 2. 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 [20]). The figure depicts the real data and the registered (using rigid motion). The top left frame contains two unregistered point-sets superposed on each other. The 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 [20] and the ICP methods. As expected, our method and the KC method exhibit a much wider convergence basin/range than the ICP and both achieve very high accuracy in the noiseless case.

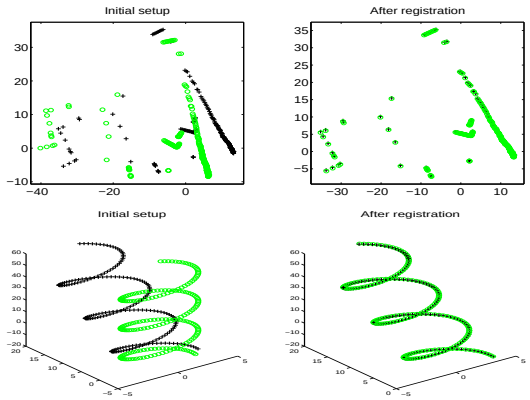

Fig. 2. Results of rigid registration in noiseless case. 'o' and '+' indicate the model and scene points respectively.

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 (**R**,**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 transformation which was partially responsible for the corruption. Results show that when the noise level is low, both KC and our 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 3.

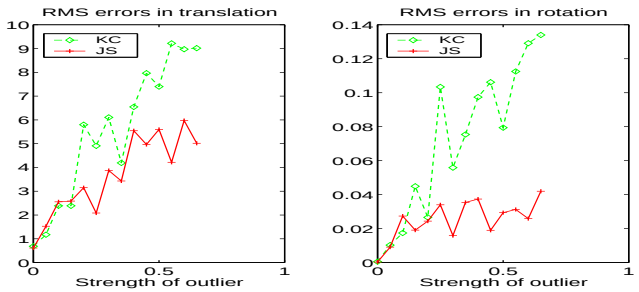

Fig. 3. Robustness to outliers in the presence of large noise. Errors in estimated rigid transform vs. proportion of outliers $((\tau-\rho)/(\rho))$ for both our method and KC method.

We also applied our algorithm to nonrigidly register medical datasets (2D point-sets). Figure 4 depicts some results of our registration method applied to two 2D corpus callosum slices with feature points manually extracted by human experts. Top left of Figure 4 contains these two unregistered point-sets superposed on each other ('o' and '+' indicate the model and scene points respectively), the registration result is shown in the lower left column. The warping of the 2D grid under the recovered motion is shown in the middle column. Our non-rigid alignment performs well in the presence of noise and outliers (Figure 4 right column). For the purpose of comparison, we also tested the TPS-RPM program provided in [14] on this data set, and found that TPS-RPM can correctly register the pair without outliers (Figure 4 top left) but failed to match the corrupted pair (Figure 4 top right).

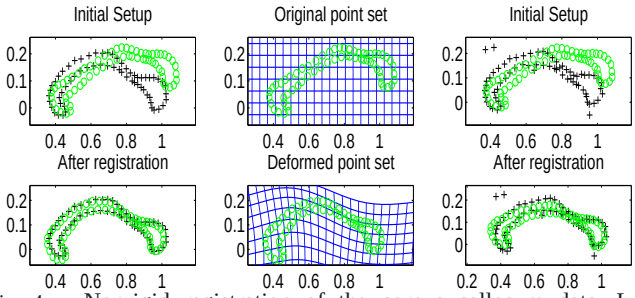

Fig. 4. Nonrigid registration of the corpus callosum data. Left column: two manually segmented corpus callosum slices before and after registration, 'o' and '+' indicate the model and scene points respectively; Middle column: warping of the 2D grid using the recovered motion; Top right: same slices with one corrupted by noise and outliers, before and after registration.

B. 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 implementation 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 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 seven normal subjects, (as shown Figure 5, indicated by "o"). The recovered deformation between each point-set and the mean shape are superimposed on the first two rows in Figure 5. The resulting atlas (mean point-set) is shown in third row of Figure 5, and is superimposed over all the point-sets. As we described earlier, all these results are computed simultaneously and automatically. This example clearly demonstrates 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.

Fig. 6 illustrates the effect of the regularization parameter λ of the thin plate spline in Eqn. (5). As the regularization parameter varies from 0.0001 to 0.005, we can see that the resulting atlas is relatively stable.

The next figure (Fig. 7) shows the stability of our algorithm by adding an eighth point-set, which is shown on the left of Fig. 7, to the original seven point-sets shown in Fig. 5. We

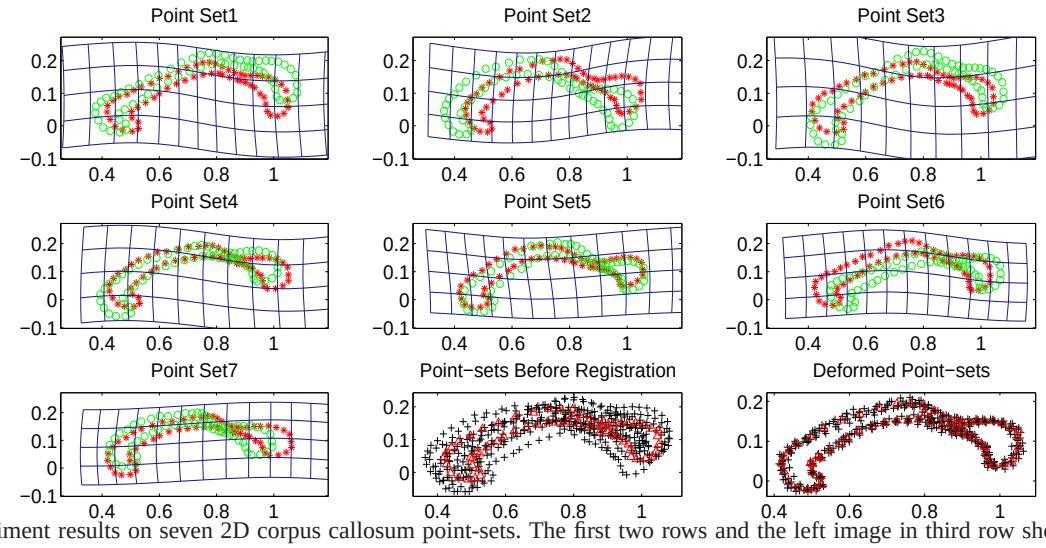

Fig. 5. Experiment results on seven 2D corpus callosum point-sets. The first two rows and the left image in third row show the deformation of each point-set to the atlas, superimposed with initial point-set (show in 'o') and deformed point-set (shown in '*'). Middle image in the third row: The estimated atlas is shown superimposed over all the point-sets. Right: The estimated atlas is shown superimposed over all the deformed point-sets.

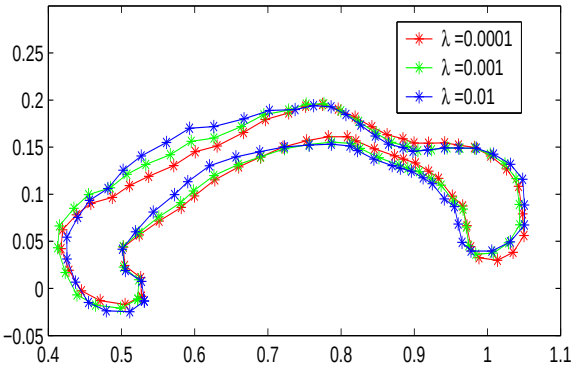

Fig. 6. Plots of the 2D atlas results with different regularization parameter of the thin plate spline.

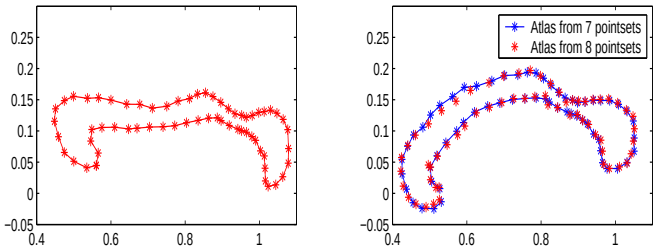

Fig. 7. Illustration of the effect of adding a point-set that is not largely varying different from the original set. On the left is the original seven point-sets augmented with a point set, on the right is the resulting atlas compared with the atlas constructed from the original seven point-sets in Fig. 5.

then constructed an new atlas from these eight point-sets. From the right plot of Fig. 7, it is evident that our algorithm yields an atlas not much different from the atlas constructed from the original seven point-sets in Fig. 5, which confirms that the constructed atlas using our algorithm is stable with the incorporation of more point-sets that are not largely varying from the original set.

Next, we present results on 3D hippocampal point-sets. Ten 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 10 subjects. The point-sets differ in shape, with the number of points varies from $412 - 805$. Nine of the 10 hippocampal point-sets are shown in Figure 8. In Figure 9, 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. We also show the scatter plot of original point-sets along with all the point-sets after the non-rigid warping in Figure 10 (a) and (b) respectively. 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.

V. Conclusions

In this paper, we presented a novel and robust algorithm using an information theoretic measure namely, the Jensen-Shannon divergence, to simultaneously compute a probabilistic mean (atlas) shape from multiple unlabeled point-sets (each represented by finite mixtures) and register them non-rigidly to this emerging mean (atlas) 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 a probabilistic 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 are that, the JS-divergence is symmetric, its square-root is a metric and allows for use of unequal cardinality of the given point sets to be registered. We also showed that the JS divergence has a number of desirable properties for use in group-wise point-sets registration e.g., i) It can be interpreted in a hypothesis testing framework and ii) it is unbiased i.e., our group-wise registration approach is not biased toward any particular point-set. However, the spatial

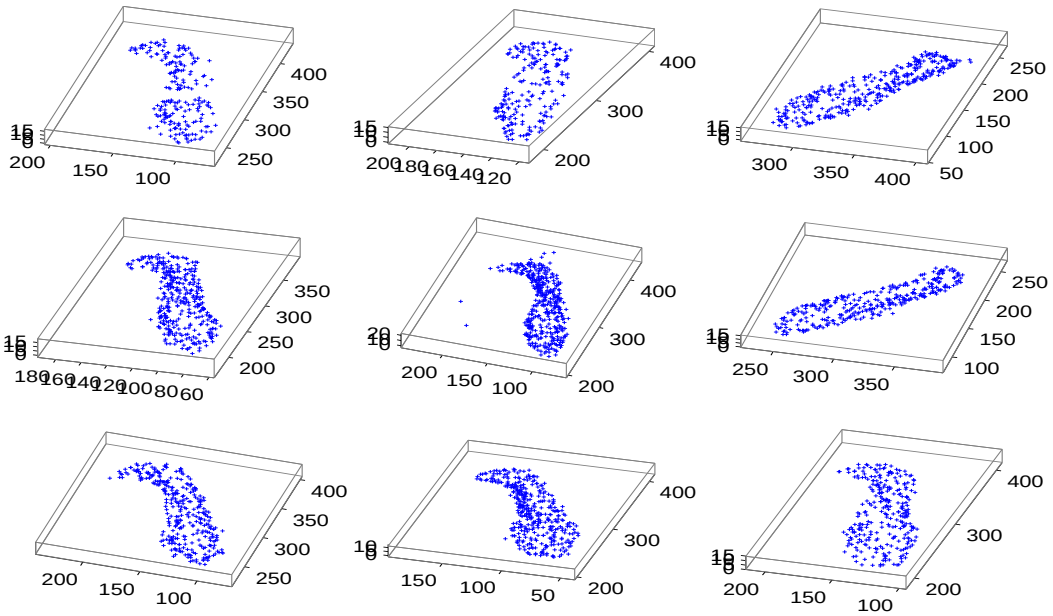

Fig. 8. 3D Hippocampal point-sets. Nine (of the 10) hippocampal point-sets are shown. Note that all the point-sets were subsampled for the purposes of display.

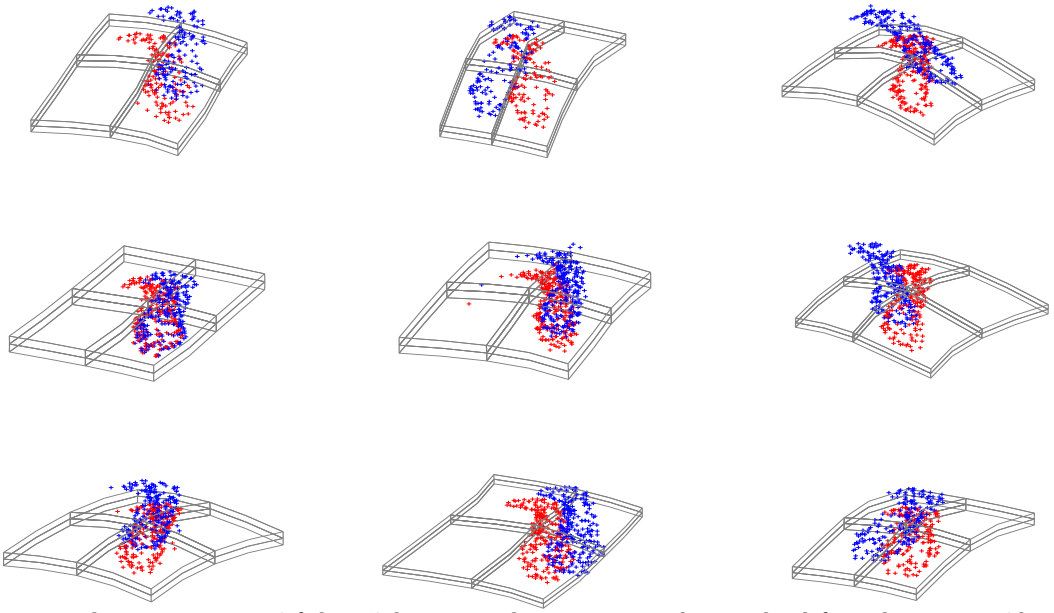

Fig. 9. 3D Hippocampal point-sets. Nine (of the 10) hippocampal point-sets are shown. The deformed point-sets (shown in red '+') is shown superimposed on the data (shown in blue '+') along with the underlying space deformation. All the point-sets were subsampled for the purposes of display.

regularization term in the cost function used for registration is not invariant in its form under a change of variables and this constitutes a type of bias—very different from the possible bias toward a particular point-set. We plan to examine this issue in future work.

The constructed atlas is a probabilistic atlas which is defined as the convex combination of the probability densities/distributions of the input point sets being aligned. The cost function optimization is achieved very efficiently by computing the analytic gradient of the same and utilizing it 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] C. Small, *The Statistical theory of shape*. New York: Springer, 1996.

[2] N. Duta, A. K. Jain, and M.-P. Dubuisson-Jolly, "Automatic construction of 2D shape models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 5, pp. 433–446, 2001.

[3] J. Lin, "Divergence measures based on the Shannon entropy," *IEEE Trans. Infor. Theory*, vol. 37, pp. 145–151, 1991.

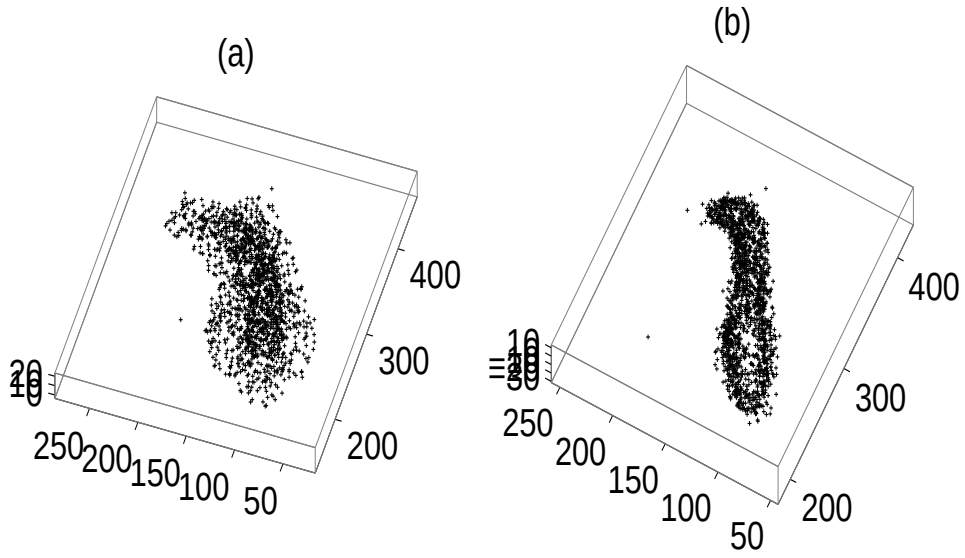

Fig. 10. Ten 3D Hippocampal point-sets. (a) Scatter plot of 10 3D hippocampal point-sets. (b) Scatter plot of the 10 deformed point-sets. Note that the point-sets were subsampled for the purposes of display.

[4] A. Hero, O. M. B. Ma, and J. Gorman, "Applications of entropic spanning graphs," *IEEE Trans. Signal Processing*, vol. 19, pp. 85–95, 2002.

[5] Y. He, A. Ben-Hamza, and H. Krim, "A generalized divergence measure for robust image registration," *IEEE Trans. Signal Processing*, vol. 51, pp. 1211–1220, 2003.

[6] M.-C. Chiang, R. A. Dutton, K. M. Hayashi, O. L. Lopez, H. J. Aizenstein, A. W. Toga, J. T. Becker, and P. M. Thompson, "3D pattern of brain atrophy in HIV/AIDS visualized using tensor-based morphometry," *NeuroImage*, vol. 34, no. 1, pp. 44–60, January 2007.

[7] D. M. Endres and J. E. Schindelin, "A new metric for probability distributions," *IEEE Trans. Info. Theory*, vol. 49, pp. 1858–60, 2003.

[8] E. Klassen, A. Srivastava, W. Mio, and S. H. Joshi, "Analysis of planar shapes using geodesic paths on shape spaces." *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 3, pp. 372–383, 2003.

[9] T. B. Sebastian, P. N. Klein, B. B. Kimia, and J. J. Crisco, "Constructing 2D curve atlases," in *IEEE Workshop MMBIA*, Washington, DC, USA, 2000, pp. 70–77.

[10] H. Tagare, "Shape-based nonrigid correspondence with application to heart motion analysis." *IEEE Trans. Med. Imaging*, vol. 18, no. 7, pp. 570–579, 1999.

[11] F. L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations." *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 6, pp. 567–585, 1989.

[12] H. Chui, A. Rangarajan, J. Zhang, and C. M. Leonard, "Unsupervised learning of an atlas from unlabeled point-sets." *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 2, pp. 160–172, 2004.

[13] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 509–522, 2002.

[14] H. Chui and A. Rangarajan, "A new point matching algorithm for non-rigid registration," *Computer Vision and Image Understanding (CVIU)*, vol. 89, pp. 114–141, 2003.

[15] H. Guo, A. Rangarajan, S. Joshi, and L. Younes, "Non-rigid registration of shapes via diffeomorphic point matching." in *ISBI*, 2004, pp. 924–927.

[16] L. Garcin and L. Younes, "Geodesic matching with free extremities," *Journal of Mathematical Imaging and Vision*, vol. 25, pp. 329–340(12), October 2006.

[17] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models: their training and application," *Comput. Vis. Image Underst.*, vol. 61, no. 1, pp. 38–59, 1995.

[18] Y. Wang and L. H. Staib, "Boundary finding with prior shape and smoothness models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 7, pp. 738–743, 2000.

[19] A. Hill, C. J. Taylor, and A. D. Brett, "A framework for automatic landmark identification using a new method of nonrigid correspondence." *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 3, pp. 241–251, 2000.

[20] Y. Tsin and T. Kanade, "A correlation-based approach to robust point set registration." in *ECCV2004(3)*, 2004, pp. 558–569.

[21] B. Jian and B. Vemuri, "A robust algorithm for point set registration using mixture of gaussians," in *ICCV2005*, 2005, pp. 1246–1251.

[22] J. Glaunes, A. Trouvé, and L. Younes, "Diffeomorphic matching of distributions: A new approach for unlabelled point-sets and sub-manifolds matching." in *CVPR2004 (2)*, 2004, pp. 712–718.

[23] Y. Wang, K. Woods, and M. McClain, "Information-theoretic matching of two point sets," *IEEE Transactions on Image Processing*, vol. 11, no. 8, pp. 868–872, August 2002.

[24] G. McLachlan and K. Basford, *Mixture Model:Inference and Applications to Clustering*. New York: Marcel Dekker, 1988.

[25] A. L. Yuille, P. Stolorz, and J. Utans, "Statistical physics, mixtures of distributions, and the EM algorithm," *Neural Comput.*, vol. 6, no. 2, pp. 334–340, 1994.