



Published in final edited form as:

*IEEE Trans Pattern Anal Mach Intell.* 2008 November ; 30(11): 2011–2022. doi:10.1109/TPAMI.2007.70829.

## Simultaneous Nonrigid Registration of Multiple Point Sets and Atlas Construction

**Fei Wang [Member, IEEE],**

IBM Almaden Research Center, G1-003, 650 Harry Road, San Jose, CA 95120

**Baba C. Vemuri [Fellow, IEEE],**

Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL 32611

**Anand Rangarajan [Member, IEEE], and**

Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL 32611

**Stephan J. Eisenschenk**

Department of Neurology, University of Florida, L3-100 McKnight Brain Institute, Newell Drive, Gainesville, FL 32610

Fei Wang: wangfe@us.ibm.com; Baba C. Vemuri: vemuri@cise.ufl.edu; Anand Rangarajan: anand@cise.ufl.edu; Stephan J. Eisenschenk: stephan.eisenschenk@neurology.ufl.edu

### Abstract

Groupwise 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 nonrigid 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 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 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 and for simultaneously (rigidly or nonrigidly) registering 3D range data sets (in vision and graphics applications), without having to establish any correspondence. We present experimental results on nonrigidly registering 2D and 3D real and synthetic data (point sets).

## Index Terms

Groupwise point set registration; finite-mixture model; Jensen-Shannon divergence; hypothesis testing; thin-plate spline

---

## 1 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 the mean and covariance of shapes remains. We are particularly interested in the unlabeled point set representation since the 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 nonrigid 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 the 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 of 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 average of the aligned point sets.

To overcome the 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 and then quantify the distance between these probability densities by 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 of the density functions are transformed through appropriate changes in their parameters, the distance measure between these densities would be minimized since all of 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. The Jensen-Shannon divergence, first introduced in [3], serves as a model divergence measure between multiple probability densities and researchers have used it as a dissimilarity measure for image registration and retrieval applications in the past (see [4], [5], [6]), *but never for the registration of two or more point sets*. It has some very desirable properties. To name a few, 1) the square root of the JS divergence (in the case when its (convex combination) parameter

is fixed to  $\frac{1}{2}$ ) is a metric [7], 2) the JS divergence relates to other information-theoretic functionals such as the relative entropy or the Kullback-Leibler (KL) divergence and, hence, it shares their mathematical properties and their intuitive interpretability, and 3) the compared densities can be weighted, which allows us to take into account the different sizes of the point samples from which the probability densities are computed. Some of these

advantages will be explicitly brought to light in subsequent sections during the course of description of our method.

The rest of this paper is organized as follows: The remainder of Section 1 gives a brief review of the literature, focusing on the differences between existing methods and ours. Section 2 contains a description of the formulation using the JS divergence for our simultaneous nonrigid registration and atlas construction model and also a derivation of the associated gradient of the energy function. Section 3 contains a description of the computational techniques employed in our algorithm for simultaneous nonrigid registration and atlas construction. Experimental results on 2D and 3D point sets are presented in Section 4. Finally, we draw conclusions in Section 5.

## 1.1 Previous Work

Extensive studies on the atlas construction for deformable shapes can be found in the 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 (for example, using point locations and spline coefficients).

The work presented in [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 differences are quantified using the lengths of geodesics connecting them on these spaces. The intrinsic mean (the Karcher mean) can be computed as a point on the manifold (of shapes) that 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 being 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 may be higher due to the use of the intrinsic higher order representation.

Among the methods using point set parameterization, the idea of using nonrigid spatial mapping functions, specifically thin-plate splines (TPSs) [11], [12], [13], to analyze deformable shape has been widely adopted. Bookstein [11] successfully initiated the research efforts on the usage of TPSs 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 typical problem besetting landmark methods, e.g., inconsistency. Several significant articles on robust and nonrigid point set matching have been published by Rangarajan et al., using TPSs [12], [14], [15]. In their recent work [12], they extend their work to the construction of a 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 nonrigid transformation between each point set to the emerging mean shape using deterministic annealing (DA) and soft assign, and the generated mean shape is entirely symmetric, with no bias toward any of the original shapes. Garcin and Younes [16] attempt solving the unlabeled point set averaging problem by using a similar idea as in [12], except that it is under the diffeomorphism setting and, consequently, the estimated distances between point sets are geodesic distances. However, in both 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 nonrigid registration problem.

The active shape model proposed in [17] utilized points to represent deformable shapes. This work pioneered the efforts to build point distribution models to understand deformable shapes [17], [18]. Objects are represented as carefully defined landmark points and the variation of shapes is modeled using the principal component analysis. These landmark points are acquired through a more or less manual landmarking process, where an expert goes through all of the samples to mark corresponding points on each sample. It is a rather tedious process and accuracy is limited. In recent work [19], an attempt is made to overcome this limitation by trying to automatically solve for the correspondences in a nonrigid setting. The resulting algorithm is very similar to 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 set 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 set alignment in recent literature which bear a close relation to our research reported here. For instance, Tsing and Kanade [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 and Vemuri introduced a novel and robust algorithm for rigidly and nonrigidly 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 in [22], as discussed in the following, 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, the approach of Wang et al. [23] is closest to our work, where a relative entropy measure (the KL distance) is used to find a similarity transformation between two point sets. The KL 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 also model each of the point sets by a kernel density function and then quantify the (dis)similarity between them by 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 the JS divergence used in our work is a lot more general than the information-theoretic measures used in [20], [21], [23] and it can easily cope with multiple point sets. Recently, in [22], Glaunes et al. have represented points as delta functions and matched them using the dual norm in a reproducing kernel Hilbert space. The main problem with this technique is that it needs a 3D spatial integral, which must be numerically computed. In contrast, we compute the JS divergence by 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.

## 2 MATHEMATICAL FORMULATION

In this section, we present the mathematical formulation of our simultaneous nonrigid registration and atlas construction method. Note that, normally, nonrigid registration precedes atlas construction since the latter requires the data to be registered. However, in our work, the atlas emerges as a by-product of the nonrigid 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 and then quantify the distance between these probability densities by using an information-theoretic measure. Fig. 1 illustrates this idea, wherein the right column depicts the density functions corresponding to the point sets drawn from a cortical substructure in the human brain called the corpus callosum, which is 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.

## 2.1 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  $\{X^c, c \in \{1, \dots, N\}\}$ . Each point set  $X^c$  consists of points  $\{x_i^c \in \mathbb{R}^d, i \in \{1, \dots, n_c\}\}$ . 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 a limited number of clusters. Note that we can choose the cluster centers to be the point set itself if the size of point sets is quite small. The cluster center point sets are denoted by  $\{V^c, p \in \{1, \dots, N\}\}$ . Each cluster point set  $V^c$  consists of points

$\{v_a^c \in \mathbb{R}^d, a \in \{1, \dots, K_c\}\}$ . 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 is 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 set registration. The atlas point set is denoted by  $Z$ . We begin by specifying the density function of each point set:

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

In (1), the occupancy probability, which is different for each data point set, is denoted by  $\alpha^c$ .

The component density  $p(x|v_a^c)$  is

$$\begin{aligned} p(x|v_a^c) &= G(x - v_a^c, \Sigma_a) \\ &= \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma_a|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x - v_a^c)^T \Sigma_a^{-1} (x - v_a^c)\right), \end{aligned} \quad (2)$$

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

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

Later, we will set the occupancy probability to be uniform and make the covariance matrices  $\Sigma_a$  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 of 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 can we measure the similarity/closeness between these densities represented by Gaussian mixtures? We answer this in the following section.

## 2.2 The Jensen-Shannon Divergence for Learning the Atlas

The 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:

1. The square root of the JS divergence (in the case when its parameter is fixed to  $\frac{1}{2}$  is a metric [7].
2. The JS divergence relates to other information-theoretic functionals, such as the relative entropy or the KL divergence, and, hence, it shares their mathematical properties and their intuitive appeal.
3. 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.
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 required by Chui and Rangarajan [14]. Given  $n$  probability distributions  $\mathbf{P}_i$ ,  $i \in \{1, \dots, n\}$ , the JS divergence of  $\mathbf{P}_i$  is defined by

$$JS_{\pi}(\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_n) = H\left(\sum \pi_i \mathbf{P}_i\right) - \sum \pi_i H(\mathbf{P}_i), \quad (4)$$

where  $\pi = \{\pi_1, \pi_2, \dots, \pi_n | \pi_i > 0, \sum \pi_i = 1\}$  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 (4) are the entropy of  $\mathbf{P} = \sum \pi_i \mathbf{P}_i$  (the  $\pi$ -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  $\mu^c$  is the set of the transformation parameters associated with each function  $f^c$ . The densities of the deformed point sets can be written as  $\mathbf{P}_c = \mathbf{P}_c(x|V^c, \mu^c) = \sum_{a=1}^{K^c} \alpha_a^c p(x|f^c(v_a^c))$ . To compute the mean shape density, that is, the probabilistic atlas, from these point sets and register them to the emerging mean shape density, we need to recover the transformation parameters  $\mu^c$ . 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

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

In (5), the weight parameter  $\lambda$  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] and [21], we choose the thin-plate spline (TPS) to represent the nonrigid deformation. Given  $n$  control points  $\mathbf{x}_1, \dots, \mathbf{x}_n$  in  $\mathbb{R}^d$ , the TPS mapping  $f: \mathbb{R}^d \rightarrow \mathbb{R}^d$  has the form  $f(x) = \mathbf{t} + \mathbf{A}\mathbf{x} + \mathbf{W}\mathbf{U}(\mathbf{x})$ . Here,  $\mathbf{t} + \mathbf{A}\mathbf{x}$  is the linear part of TPS, whereas  $\mathbf{W}\mathbf{U}(\mathbf{x})$  is the nonlinear part, which is determined by a  $d \times n$  matrix  $\mathbf{W}$ ,  $\mathbf{U}(\mathbf{x})$  is an  $n \times 1$  vector consisting of  $n$  basis functions  $U_i(\mathbf{x}) = U(\mathbf{x}, \mathbf{x}_i)$ , where  $U(r)$  is the kernel function of TPS. For example, if the dimension is 2 ( $d = 2$ ), we have  $U(r) = 1/(8\pi)r^2 \ln(r)$ . Therefore, the regularization term in (5) is governed by the bending energy of the TPS warping and can be explicitly as  $\text{trace}(\mathbf{W}\mathbf{K}\mathbf{W}^T)$ , where  $\mathbf{K} = (K_{ij})$  and  $K_{ij} = U(\mathbf{x}_i, \mathbf{x}_j)$ . In our experiments, the clusters are used as control points. Other schemes for choosing control points may also be considered. Note the linear part of the TPS can be obtained by an initial affine registration, then numerical optimization techniques can be applied to find the nonrigid parameter  $\mathbf{W}$ .

Next, we will present some properties of the JS divergence in the context of groupwise point sets registration.

### 2.3 Jensen-Shannon Divergence in a Hypothesis Testing Framework

In this section, we show that the JS 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 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, that is,  $\mathbf{P}_1$  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

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

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 the 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 is used in the numerator and denominator of (6), with the main difference 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  $\Lambda$  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 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 independent and identically distributed (i.i.d.)

samples drawn from a mixture  $(\sum_a \pi_a \mathbf{P}_a)$ , with  $\pi_a = \frac{n_a}{\sum_b n_b}$ , and i.i.d. samples drawn from a heterogeneous collection of densities  $(\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_N)$ , 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

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

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

**Theorem 1.** *Given  $N$  probability distributions  $\mathbf{P}_a$ ,  $a \in \{1, \dots, N\}$ , maximizing the hypothesis ratio in (7) is equivalent to minimizing the JS divergence between the  $N$  probability distributions  $\mathbf{P}_a$ ,  $a \in \{1, \dots, N\}$ .*

**Proof.** Taking the negative logarithm of the likelihood ratio, we have

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

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

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

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. 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.

## 2.4 Jensen-Shannon Divergence Is Unbiased

Typically, we are required to construct an atlas from a 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 by using our algorithms and all of the point sets inside each of the subsets are registered. Then, we can either construct a single atlas from these  $m$  atlases or 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 us give the answer, with the exclusion of the TPS part of the cost function.

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

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

**Proof.** We can rewrite  $JS_{\beta}(\mathbf{S}_1, \mathbf{S}_2, \dots, \mathbf{S}_m)$  as

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

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

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

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

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

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

In our registration algorithm, all of the point sets are represented as probability densities and the atlas constructed can be considered a 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 that the relationship in (10) holds between the JS divergence of the  $\mathbf{P}_a$ s and  $\mathbf{S}_i$ s. Notice that the right-hand side of (10) is the JS divergences of the densities in all the subsets, which are minimized in each step of the hierarchical technique that 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 that the right-hand side of (10) is zero. Consequently, the JS divergence of the  $\mathbf{P}_a$ s and the JS divergence of the  $\mathbf{S}_i$ s are equal to each other. Therefore, minimizing the JS divergence of all of 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 of estimating 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.

### 3 ESTIMATING THE EMPIRICAL JENSEN-SHANNON

In (5),

$$\begin{aligned}\mathbf{P}_i &= \mathbf{P}_i(x|V^i, \mu^i) = \sum_{a=1}^{K_i} \alpha_a^i p(x|f^i(v_a^i)) \\ &= \frac{1}{K_i} \sum_{a=1}^{K_i} G(x - f^i(v_a^i), \Sigma_a) = \frac{1}{K_i} \sum_{a=1}^{K_i} G(x - f^i(v_a^i), \sigma^2 \mathbf{I}),\end{aligned}$$

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

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

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

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

Combining the two terms in (11) and (13), we have

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

### 3.1 Cost Function Optimization

The 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):

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

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

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

Based on this, it is straightforward to derive the gradient of the JS divergence in (14) with respect to the transformation parameters  $\mu^i$ , which is given by

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

### 3.2 Algorithm Summary

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

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

1. Estimate the cluster centers  $\{V^c, c \in \{1, \dots, N\}\}$  for each point  $X^c$ . In our implementation, we utilize the DA procedure, with its proven benefit of robustness in clustering [25].
2. Set the initial transformation parameters  $\mu^c$  to zero and optimize the cost function in (5) with respect to the transformation parameter  $\mu^c$ . Since the analytic gradients with respect to these transformation parameters have to be explicitly derived in (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 redrawn 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 data sets with large nonrigid deformations show that 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 deformations to the corresponding point sets to recover the mean shape.

Note that our transformation model can be any type of geometric transformations, for example, rigid, affine, polynomial, or other type of nonrigid transformations. Therefore, our algorithm can be applied to registration problems other than the atlas construction. For example, 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 recovering 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 and an atlas construction from multiple point sets in the next section.

## 4 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.

### 4.1 Alignment Results

We first perform a set of exact rigid registration experiments on noiseless data sets to test the validity of our approach and the alignment results are shown in Fig. 2. The top row shows the registration result for a 2D real range data set of a road [20]. The figure depicts the real data and the registered (using rigid motion). The top left frame contains two unregistered point sets superimposed 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]; both our method and the KC method achieve very high accuracy in the noiseless case.

Following a similar experimental design in [21], we used the following procedure to test how our method behaves in the presence of noise and outliers. For a model point set with  $n$  points, we first discard a subset of size  $(1 - \eta)n$  from the point set and then we apply a synthetic rigid transformation to the template. Finally, we add  $(\tau - \eta)\eta$  spurious, uniformly distributed points to the point set, which have a total of  $\tau n$  points, of which only  $\eta n$  comes from the uncorrupted model point set. We compare our method with the relatively more robust KC method [20] (compared with ICP) and the comparison is done via a set of 2D experiments. *At each of the outlier strengths, we generate five models and six corrupted point sets from each model for a total of 30 pairs at each 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 high, our method exhibits stronger resistance to outliers than the KC method, as shown in Fig. 3.

We also applied our algorithm to nonrigidly register medical data sets (2D point sets). Fig. 4 depicts some results of our registration method applied to two 2D corpus callosum slices, with feature points manually extracted by human experts. The top left of Fig. 4 contains these two unregistered point sets superimposed on each other (“o” and “+” indicate the model and scene points, respectively); the registration result is shown in the lower left column of Fig. 4. The warping of the 2D grid under the recovered motion is shown in the middle column of Figs. 4. Our nonrigid alignment performs well in the presence of noise and outliers (Figs. 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 (Fig. 4 top left) but failed to match the corrupted pair (Fig. 4 top right).

## 4.2 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 implementation on real hippocampal data sets. The structure that we are interested in 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 in Fig. 5, indicated by “o”). The recovered deformation between each point set and the mean shape are superimposed on the first two rows in Fig. 5. The resulting atlas (mean point set) is shown in the third row of Fig. 5 and is superimposed over all of the point sets. As we have described earlier, all of 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  $\lambda$  of the TPS in (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 the eighth point set, which is shown on the left in Fig. 7, to the original seven point sets shown in Fig. 5. We then constructed a 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 varying from 412 to 805. Nine of the 10 hippocampal point sets are shown in Fig. 8. In Fig. 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 scatterplot of original point sets, along with all the point sets after the nonrigid warping in Figs. 10a and 10b, respectively. An examination of the two scatterplots clearly shows the efficacy of our recovered nonrigid warping. Note that validation of what an atlas shape ought to be in the real data case is not feasible.

## 5 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 nonrigidly to this emerging mean (atlas) shape. Atlas construction normally requires the task of nonrigid registration prior to forming the atlas. However, the unique feature of our work is that a probabilistic atlas emerges as a by-product of the nonrigid registration. Other advantages of using the JS divergence over existing methods in the literature for atlas construction and nonrigid registration are that the JS divergence is symmetric, its square root is a metric, and it allows the 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 groupwise point set registration. For example, 1) it can be interpreted in a hypothesis testing framework and 2) it is unbiased, that is, our groupwise registration approach is not biased toward any particular point set. However, the spatial 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 which is very different from the possible bias toward a particular point set. We plan to examine this issue in our 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 nonmedical domains. Our future work will focus on generalizing the nonrigid deformations to diffeomorphic mappings.

## Acknowledgments

This research was funded in part by US National Institutes of Health grant RO1 NS046812 and US National Science Foundation grant NSF 0307712.

## REFERENCES

1. Small, C. *The Statistical Theory of Shape*. Springer; 1996.
2. Duta N, Jain AK, Dubuisson-Jolly M-P. Automatic Construction of 2D Shape Models. *IEEE Trans. Pattern Analysis and Machine Intelligence* 2001 May;vol. 23(no. 5):433–446.
3. Lin J. Divergence Measures Based on the Shannon Entropy. *IEEE Trans. Information Theory* 1991;vol. 37:145–151.
4. Hero A, Ma OMB, Gorman J. Applications of Entropic Spanning Graphs. *IEEE Trans. Signal Processing* 2002;vol. 19:85–95.
5. He Y, Ben-Hamza A, Krim H. A Generalized Divergence Measure for Robust Image Registration. *IEEE Trans. Signal Processing* 2003;vol. 51:1211–1220.
6. Chiang M-C, Dutton RA, Hayashi KM, Lopez OL, Aizenstein HJ, Toga AW, Becker JT, Thompson PM. 3D Pattern of Brain Atrophy in HIV/AIDS Visualized Using Tensor-Based Morphometry. *NeuroImage* 2007 Jan;vol. 34(no. 1):44–60. [PubMed: 17035049]
7. Endres DM, Schindelin JE. A New Metric for Probability Distributions. *IEEE Trans. Information Theory* 2003;vol. 49:1858–1860.
8. Klassen E, Srivastava A, Mio W, Joshi SH. Analysis of Planar Shapes Using Geodesic Paths on Shape Spaces. *IEEE Trans. Pattern Analysis and Machine Intelligence* 2003 Mar;vol. 26(no. 3):372–383.
9. Sebastian TB, Klein PN, Kimia BB, Crisco JJ. Constructing 2D Curve Atlases. *Proc. IEEE Workshop Math. Methods in Biomedical Image Analysis* 2000:70–77.
10. Tagare H. Shape-Based Nonrigid Correspondence with Application to Heart Motion Analysis. *IEEE Trans. Medical Imaging* 1999;vol. 18(no. 7):570–579.
11. Bookstein FL. Principal Warps: Thin-Plate Splines and the Decomposition of Deformations. *IEEE Trans. Pattern Analysis and Machine Intelligence* 1989 June;vol. 11(no. 6):567–585.
12. Chui H, Rangarajan A, Zhang J, Leonard CM. Unsupervised Learning of an Atlas from Unlabeled Point Sets. *IEEE Trans. Pattern Analysis and Machine Intelligence* 2004 Feb;vol. 26(no. 2):160–172.
13. Belongie S, Malik J, Puzicha J. Shape Matching and Object Recognition Using Shape Contexts. *IEEE Trans. Pattern Analysis and Machine Intelligence* 2002 Apr;vol. 24(no. 4):509–522.
14. Chui H, Rangarajan A. A New Point Matching Algorithm for Non-Rigid Registration. *Computer Vision and Image Understanding* 2003;vol. 89:114–141.
15. Guo, H.; Rangarajan, A.; Joshi, S.; Younes, L. Nonrigid Registration of Shapes via Diffeomorphic Point Matching. *Proc. First IEEE Int'l Symp. Biomedical Imaging*; 2004. p. 924-927.
16. Garcin L, Younes L. Geodesic Matching with Free Extremities. *J. Math. Imaging and Vision* 2006 Oct;vol. 25(no. 12):329–340.
17. Cootes TF, Taylor CJ, Cooper DH, Graham J. Active Shape Models: Their Training and Application. *Computer Vision and Image Understanding* 1995;vol. 61(no. 1):38–59.
18. Wang Y, Staib LH. Boundary Finding with Prior Shape and Smoothness Models. *IEEE Trans. Pattern Analysis and Machine Intelligence* 2000 July;vol. 22(no. 7):738–743.
19. Hill A, Taylor CJ, Brett AD. A Framework for Automatic Landmark Identification Using a New Method of Nonrigid Correspondence. *IEEE Trans. Pattern Analysis and Machine Intelligence* 2000 Mar;vol. 22(no. 3):241–251.
20. Tsin, Y.; Kanade, T. A Correlation-Based Approach to Robust Point Set Registration. *Proc. Eighth European Conf. Computer Vision*; 2004. p. 558-569.
21. Jian, B.; Vemuri, B. A Robust Algorithm for Point Set Registration Using Mixture of Gaussians. *Proc. 10th IEEE Int'l Conf. Computer Vision*; 2005. p. 1246-1251.
22. Glaunes, J.; Trounev, A.; Younes, L. Diffeomorphic Matching of Distributions: A New Approach for Unlabeled Point-Sets and Sub-Manifolds Matching. *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*; 2004. p. 712-718.
23. Wang Y, Woods K, McClain M. Information-Theoretic Matching of Two Point Sets. *IEEE Trans. Image Processing* 2002 Aug;vol. 11(no. 8):868–872.

24. McLachlan, G.; Basford, K. *Mixture Model: Inference and Applications to Clustering*. Marcel Dekker; 1988.
25. Yuille AL, Stolorz P, Utans J. Statistical Physics, Mixtures of Distributions, and the EM Algorithm. *Neural Computation* 1994;vol. 6(no. 2):334–340.

## Biographies



**Fei Wang** received the BSc degree in electrical engineering from the University of Science and Technology of China in 2001 and the MSc and PhD degrees in computer science from the University of Florida in 2003 and 2006, respectively. He is currently a research staff member with the Health Informatics Department at the IBM Almaden Research Center, working on the multimodal mining project for healthcare decision support. His work covers multiple aspects of medical image analysis, computer vision, machine learning, computer graphics, and shape modeling. He is a reviewer of many major international conference proceedings and journals. He is a member of the IEEE and the Sigma Xi Honor Society.



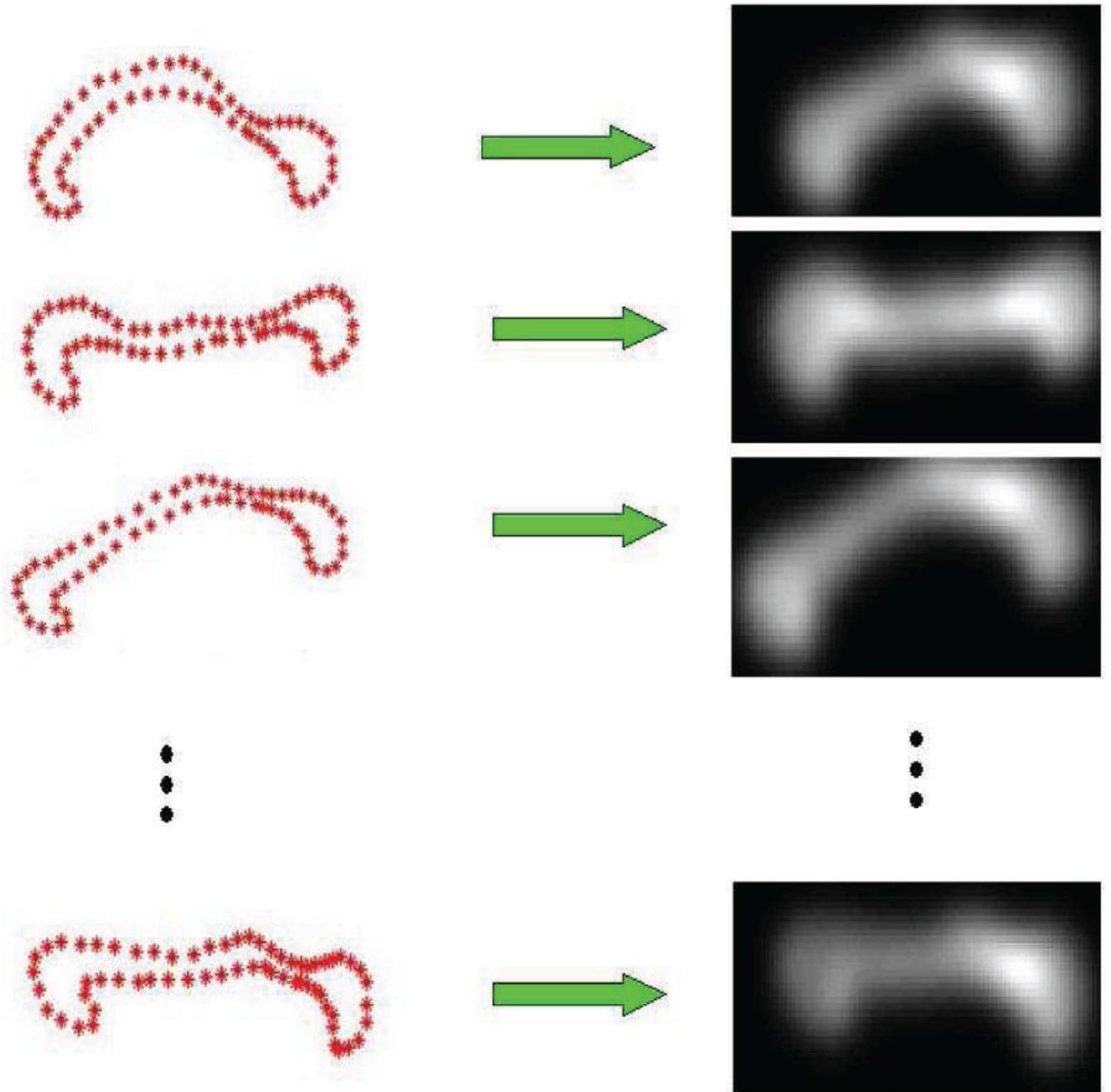
**Baba C. Vemuri** received the PhD degree in electrical and computer engineering from the University of Texas at Austin in 1987. He joined the Department of Computer and Information Sciences at the University of Florida, Gainesville, in 1987 and is currently a university research foundation professor of computer and information sciences and engineering. He was a coprogram chair of the 11th IEEE International Conference on Computer Vision (ICCV 2007). He has been an area chair and a program committee member of several IEEE conferences. He was an associate editor for several journals, including the *IEEE Transactions on Pattern Analysis and Machine Intelligence* (from 1992 to 1996) and the *IEEE Transactions on Medical Imaging* (from 1997 to 2003). He is currently an associate editor for the *Journal of Medical Image Analysis* and the *Journal of Computer Vision and Image Understanding*. His research interests include medical image analysis, computational vision, modeling for vision and graphics, and applied mathematics. For the last several years, his research work has primarily focused on information geometric methods. Along this theme, he has been developing algorithms for the analysis of diffusion weighted MRI and diffusion tensor MRI, 3D image segmentation, unimodal and multimodal image (rigid+nonrigid) registration, nonrigid registration of 3D point sets, metric learning, and large margin classifiers. He has published more than 100 refereed journal articles and conference proceedings on medical image analysis, computer vision, graphics, and applied mathematics. He is a fellow of the IEEE and a member of ACM SIGGRAPH. He received the US National Science Foundation Research Initiation Award (NSF RIA) in 1988 and the Whitaker Foundation Award in 1994. He was a recipient of the Best Peer Reviews at the Third European Conference on Computer Vision (ECCV 1994), the Best Poster Award at the 17th International Conference on Information Processing in Medical Imaging (IPMI 2001), and the Best Poster Award at IPMI 2005.



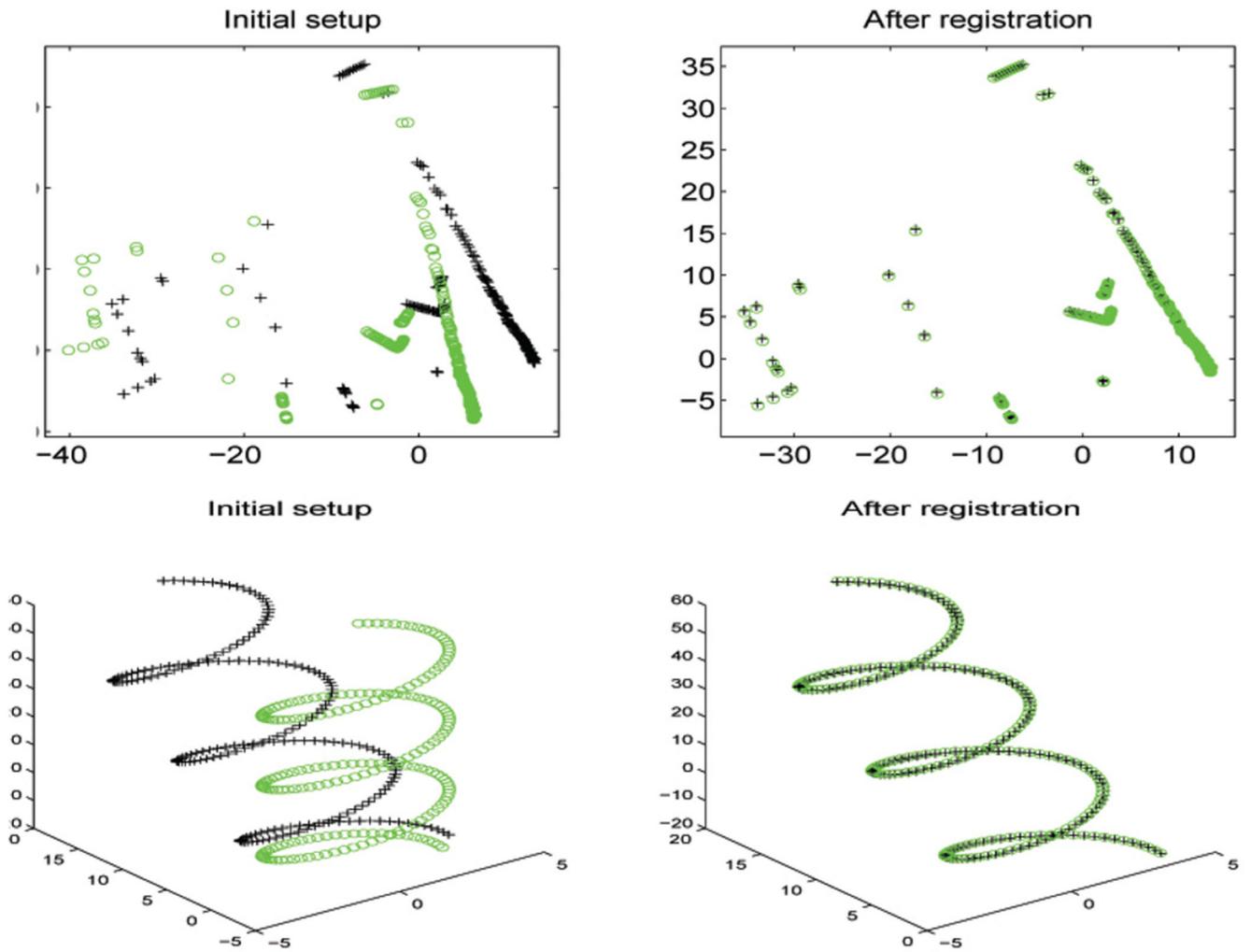
**Anand Rangarajan** received the BTech degree in electronics engineering from the Indian Institute of Technology, Madras, India, in 1984 and the PhD degree in electrical engineering from the University of Southern California in 1991. From 1990 to 1992, he was a postdoctoral associate in the Department of Diagnostic Radiology and the Department of Computer Science at Yale University and held a joint research faculty position from 1992 to 1995. From 1995 to 2000, he was an assistant professor with the Image Processing and Analysis Group (IPAG), Department of Diagnostic Radiology and the Department of Electrical Engineering at Yale University. He is currently an associate professor in the Department of Computer and Information Science and Engineering (CISE) at the University of Florida. He was the chair of the 1992 NIPS Workshop on Deterministic Annealing and Combinatorial Optimization and a cochair (with Eric Mjolsness) of the 1995 NIPS Workshop on Statistical and Structural Models in Network Vision. He has served on the program committees of the First International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR 1997), EMMCVPR 1999, and EMMCVPR 2001, the 1998 IEEE Workshop on Biomedical Image Analysis (WBIA), 1999 IEEE Workshop on Statistical and Computational Theories of Vision (SCTV), Third International Workshop on Biomedical Image Registration (WBIR 1999), IEEE Workshop on Motion and Video Computing 2002, 2000 IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), CVPR 2001, and CVPR 2003. He was also a cochair of the 2001 IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA) (with Larry Staib) and EMMCVPR 2003 (with Mario Figueiredo and Josiane Zerubia). He was active in the 1996, 1998, 2000, and 2002 Tucson Conferences Toward a Science of Consciousness. His research interests include the application of machine learning ideas and principles to medical image analysis and computer vision and the scientific study of consciousness. He is member of the IEEE.



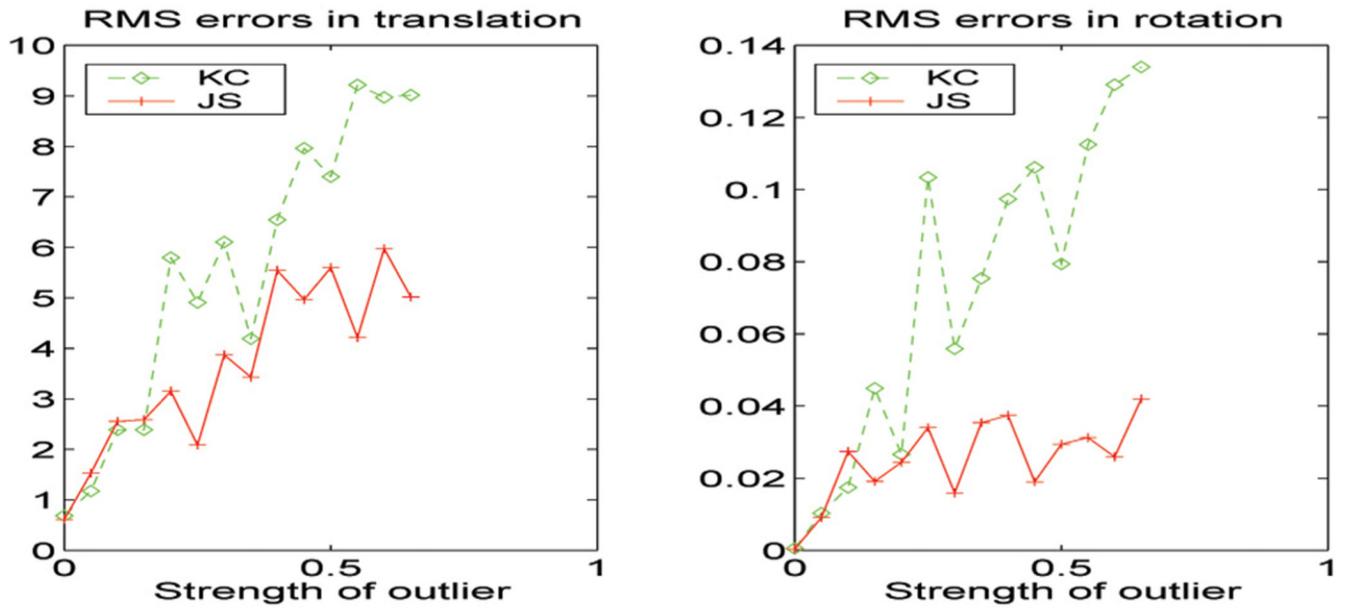
**Stephan J. Eisenschenk** is currently an associate professor in the Department of Neurology, University of Florida. He is fellowship trained in the neurophysiology in epilepsy and sleep medicine at the University of Florida. His services include the diagnosis of both epileptic syndromes and sleep disorders. His main research interests are the cellular and biochemical basis of seizures, including the propagation patterns of seizure activity, with applications to the treatment of patients with epilepsy. His current research includes the assessment of neuronal densities and correlation to seizure activity and patterns of spread throughout the brain, in conjunction with clinical outcomes of patients that have undergone seizure surgery. His research also includes the utilization of stereotactic radiosurgery for the treatment of focal epilepsy.



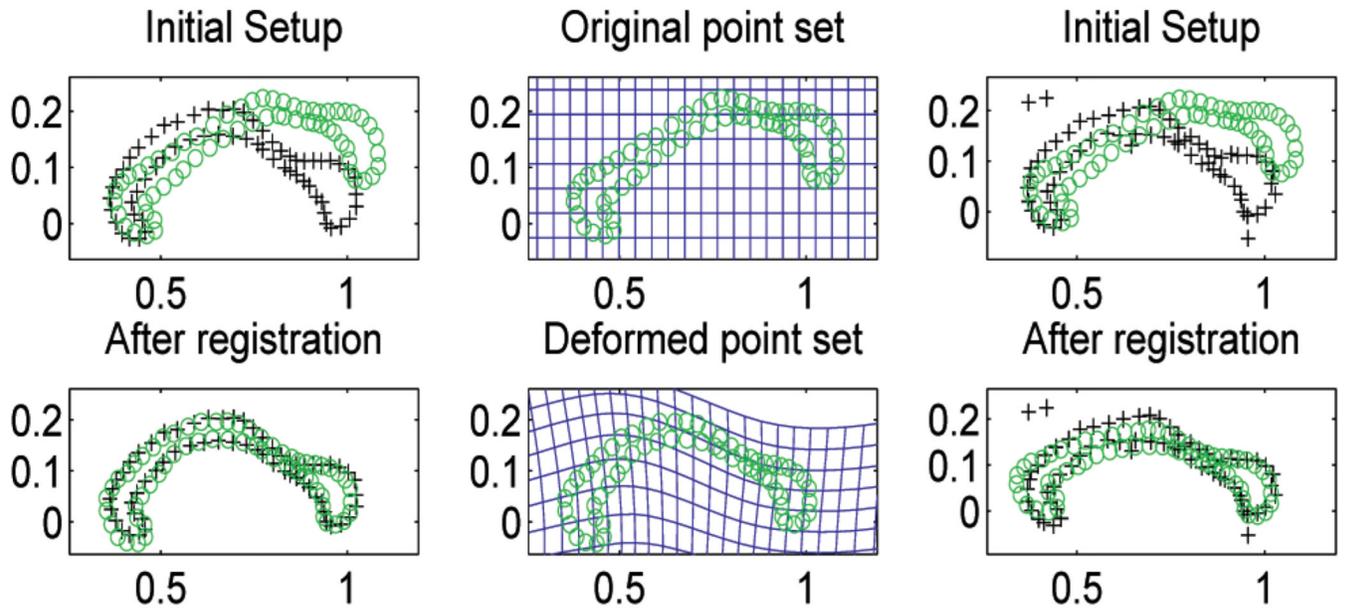
**Fig. 1.** Illustration of corpus callosum shapes (point sets) represented as density functions.



**Fig. 2.** Results of rigid registration in the noiseless case. “o” and “+” indicate the model and scene points, respectively.

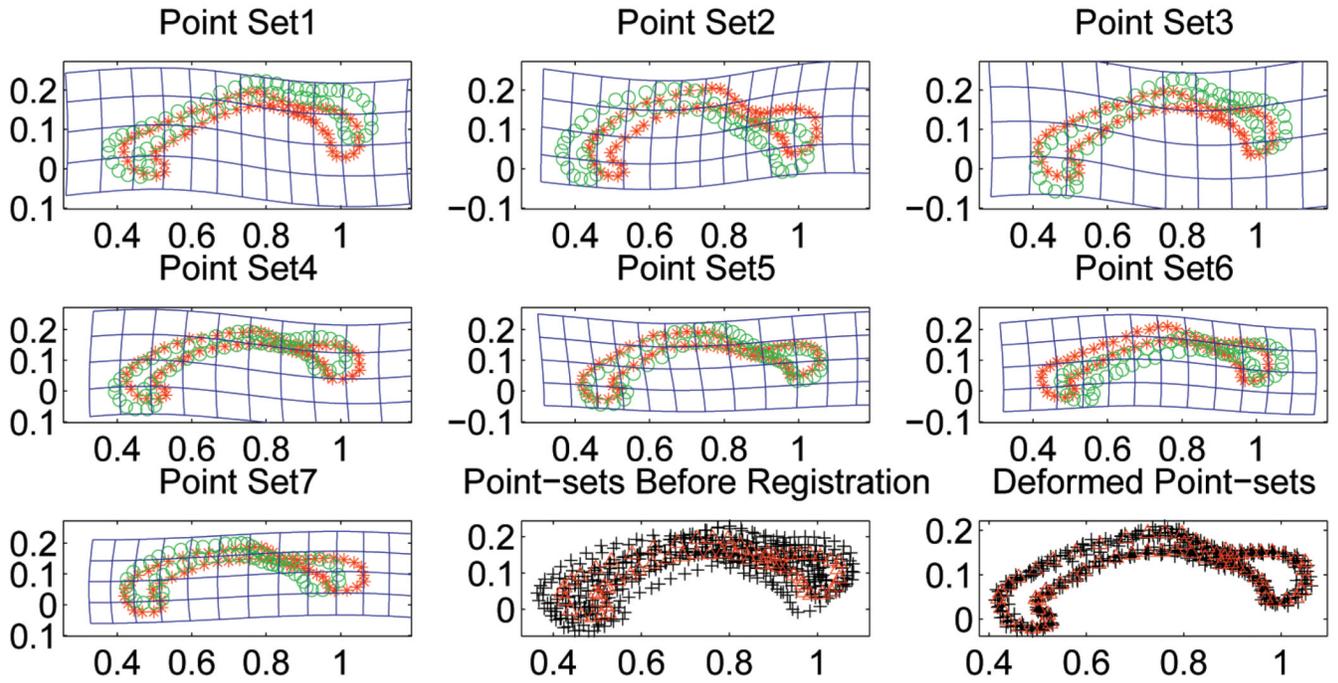


**Fig. 3.** Robustness to outliers in the presence of large noise. Errors in the estimated rigid transform versus a proportion of outliers  $((\tau - \eta)/(\eta))$  for both our method and the KC method.



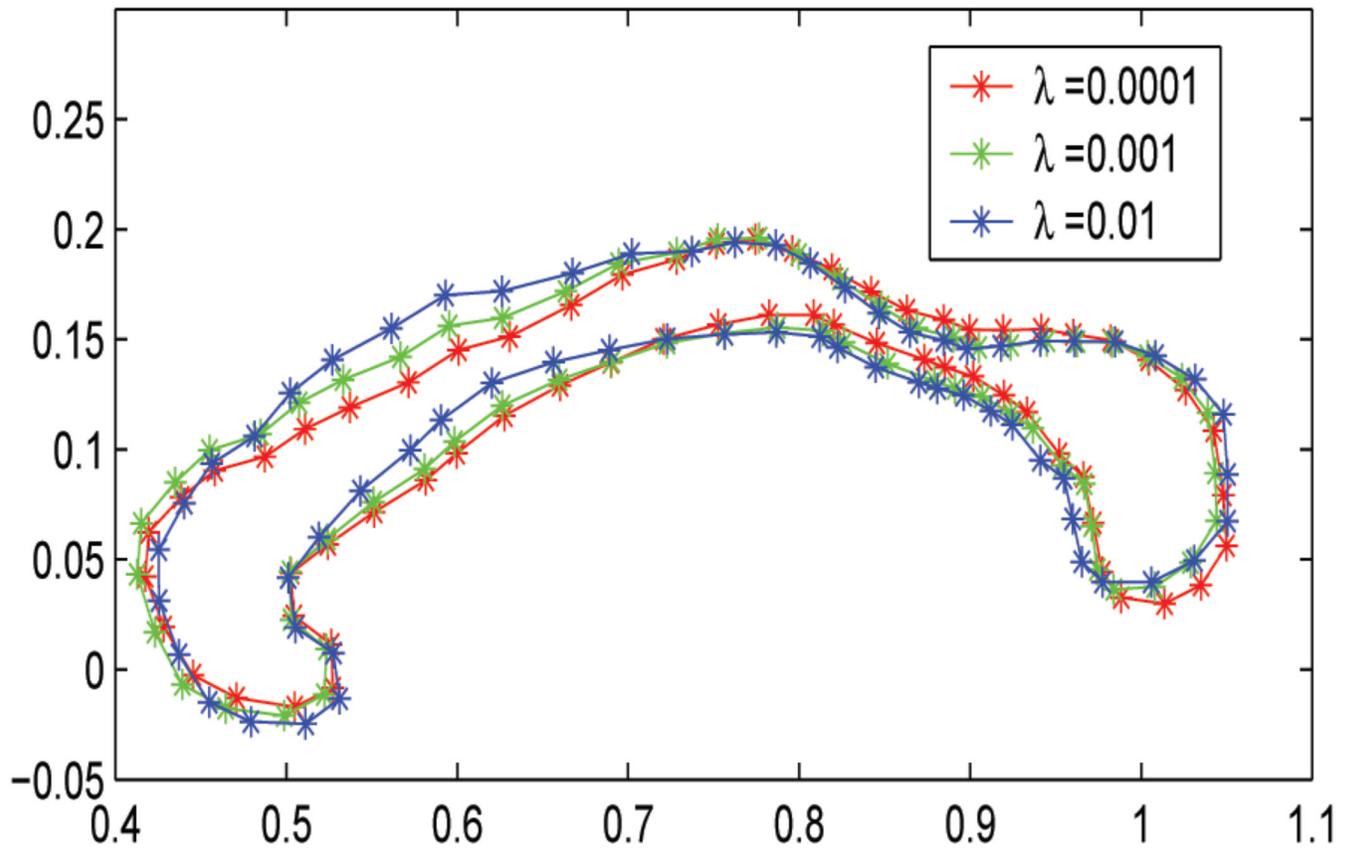
**Fig. 4.**

Nonrigid registration of the corpus callosum data. Left column: Two manually segmented corpus callosum slices before and after registration, respectively. “o” and “+” indicate the model and scene points, respectively. Middle column: Warping of the 2D grid using the recovered motion. Right column: Same slices, with one corrupted by noise and outliers, before and after registration.

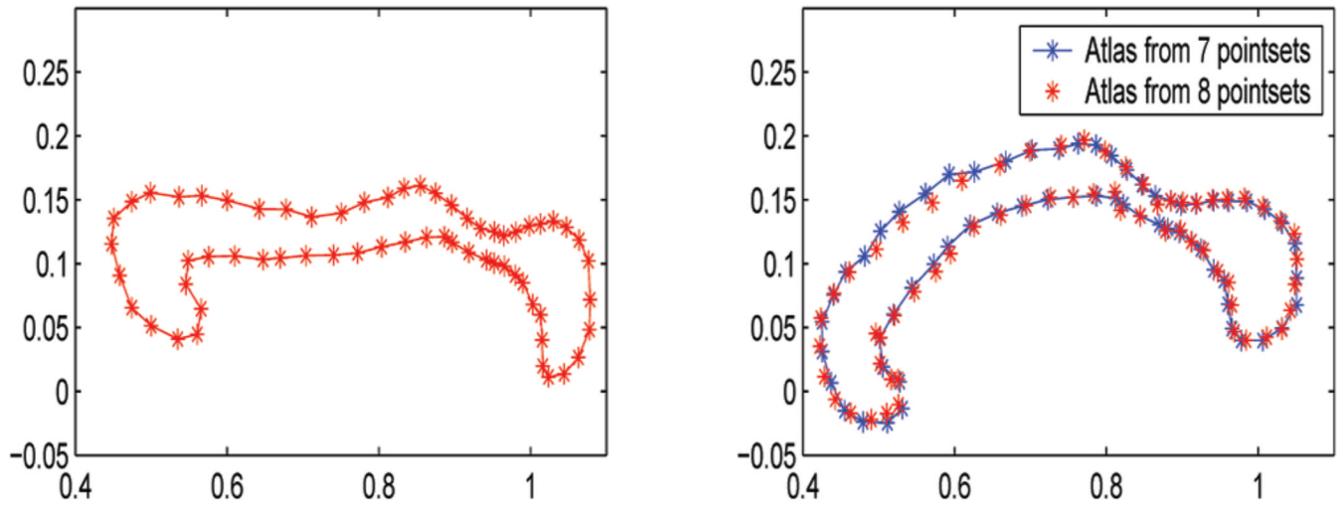


**Fig. 5.**

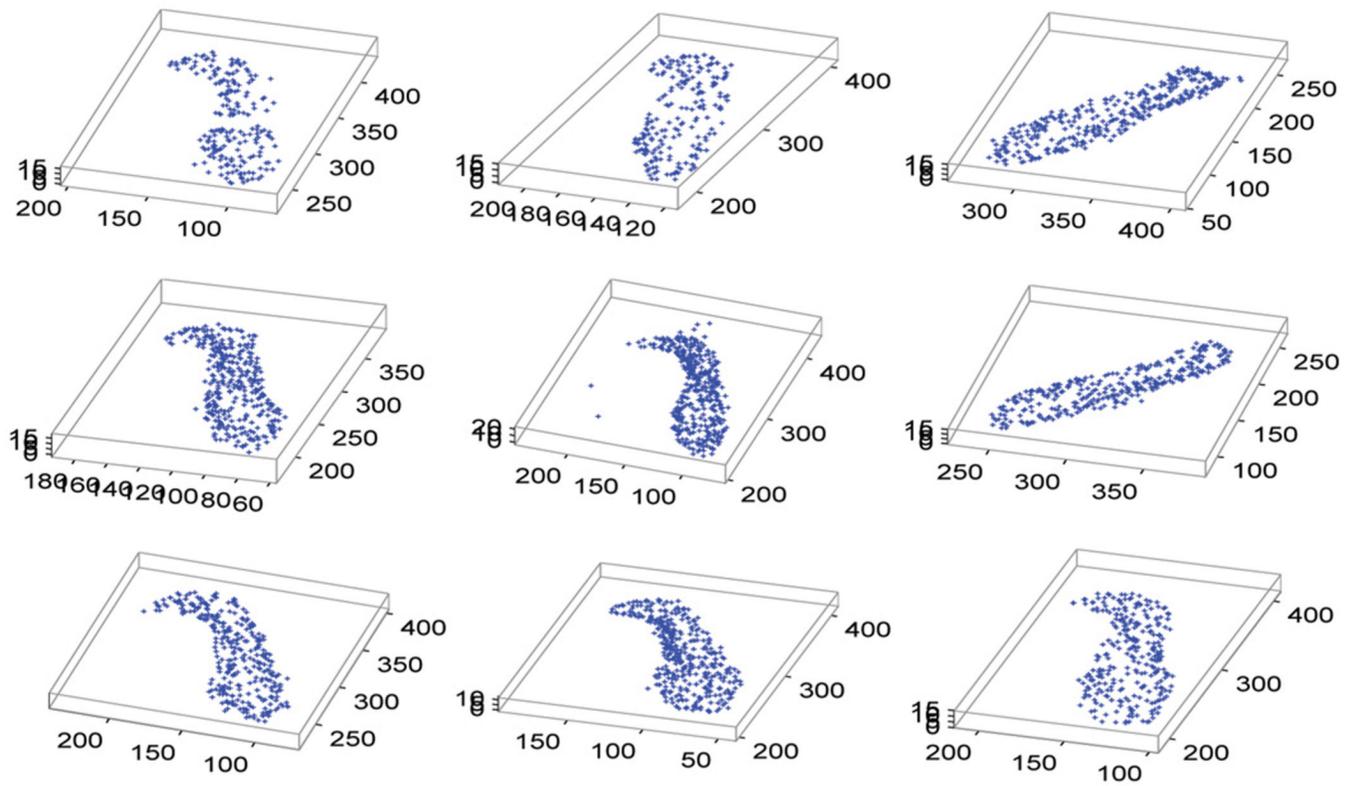
Experiment results on seven 2D corpus callosum point sets. The first two rows and the left image in the third row show the deformation of each point set to the atlas, superimposed with the initial point set (show in “o”) and deformed point set (shown in “\*”). The middle image in the third row shows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



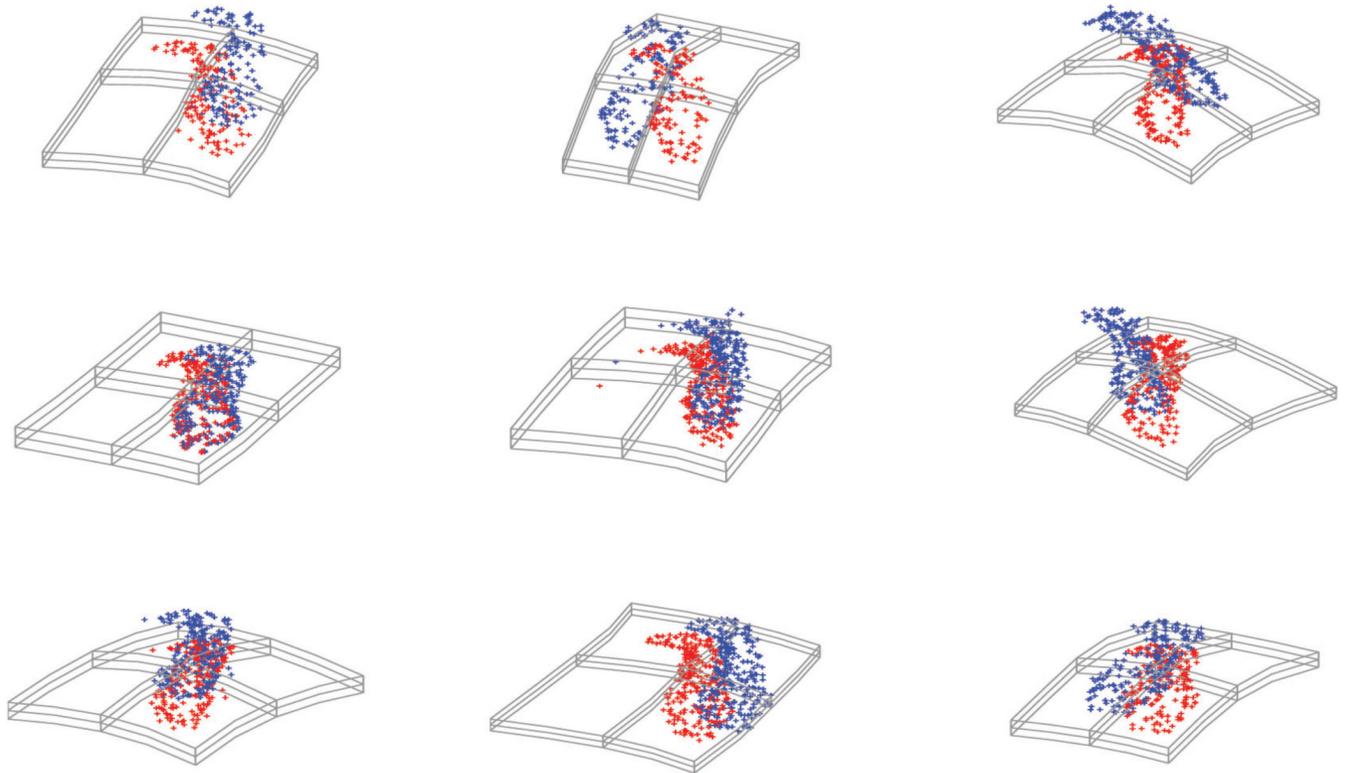
**Fig. 6.**  
Plots of the 2D atlas results with different regularization parameters of the TPS.



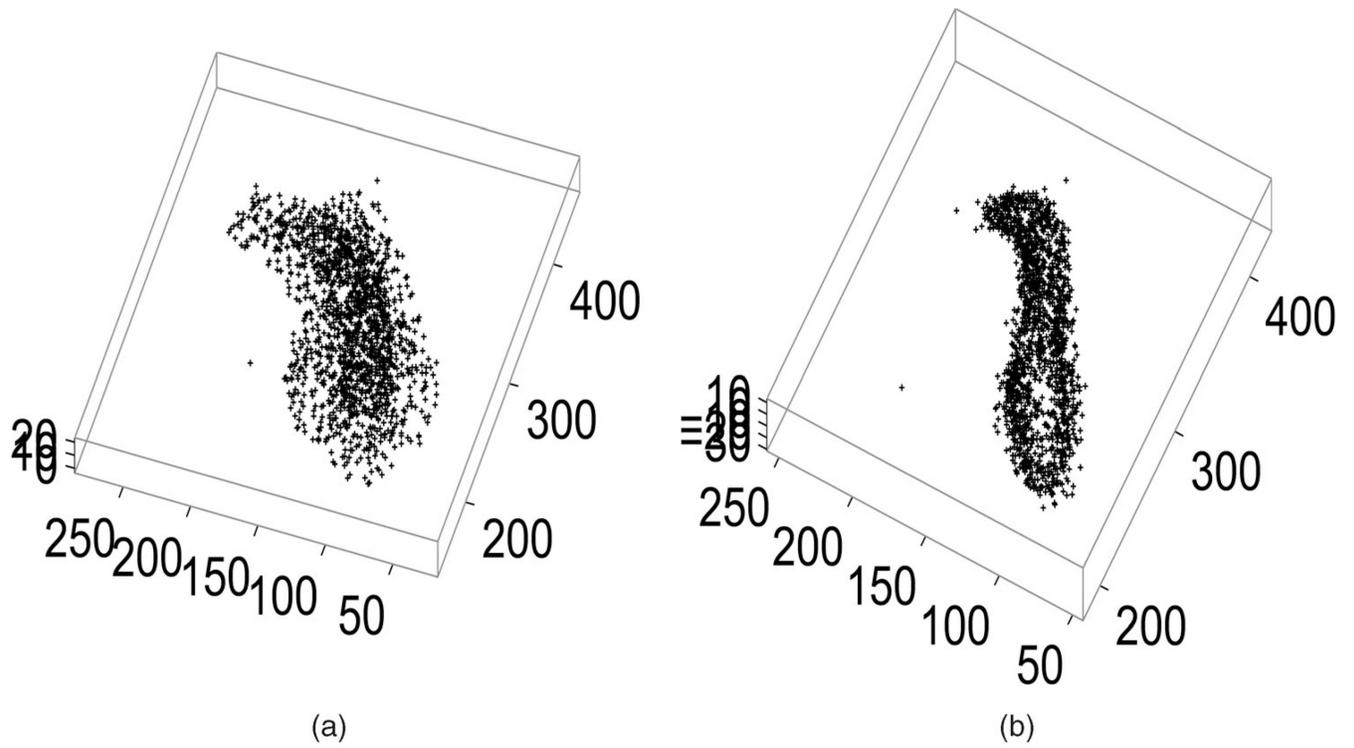
**Fig. 7.** Illustration of the effect of adding a point set that is not largely varying or different from the original set. On the left is the original seven point sets augmented with a point set. On the right is the resulting atlas compared with the atlas constructed from the original seven point sets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



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



**Fig. 9.** 3D hippocampal point sets. Nine (of the 10) hippocampal point sets are shown. The deformed point sets (shown in red “+”) are shown superimposed on the data (shown in blue “+”), along with the underlying space deformation. All of the point sets were subsampled for the purpose of display.



**Fig. 10.** Ten 3D hippocampal point sets. (a) Scatterplot of 10 3D hippocampal point sets. (b) Scatterplot of the 10 deformed point sets. Note that the point sets were subsampled for the purpose of display.