Groupwise point pattern registration using a novel CDF-based Jensen-Shannon Divergence

Fei Wang, Baba C. Vemuri and Anand Rangarajan
Department of Computer and Information Science and Engineering
University of Florida, Gainesville, FL, 32611 USA
{fewang, vemuri, anand}@cise.ufl.edu [*]

Abstract

In this paper, we propose a novel and robust algorithm for the groupwise non-rigid registration of multiple unlabeled point-sets with no bias toward any of the given point-sets. To quantify the divergence between multiple probability distributions each estimated from the given point sets, we develop a novel measure based on their cumulative distribution functions that we dub the CDF-JS divergence. The measure parallels the well known Jensen-Shannon divergence (defined for probability density functions) but is more regular than the JS divergence since its definition is based on CDFs as opposed to density functions. As a consequence, CDF-JS is more immune to noise and statistically more robust than the JS.

We derive the analytic gradient of the CDF-JS divergence with respect to the non-rigid registration parameters for use in the numerical optimization of the groupwise registration leading a computationally efficient and accurate algorithm. The CDF-JS is symmetric and has no bias toward any of the given point-sets, since there is NO fixed reference data set. Instead, the groupwise registration takes place between the input data sets and an evolving target dubbed the pooled model. This target evolves to a fully registered pooled data set when the CDF-JS defined over this pooled data is minimized. Our algorithm is especially useful for creating atlases of various shapes (represented as point distribution models) as well as for simultaneously registering 3D range data sets without establishing any correspondence. We present experimental results on non-rigid registration of 2D/3D real point set data.

1. Introduction

Matching point patterns is ubiquitous in many fields of Engineering and Science e.g., medical imaging, sports science, archaeology, and others. Point sets are widely used in computer vision to represent boundary points of shapes contained in images or any other salient features of objects contained in images. Given two or more images represented using the salient features contained therein, most often than not, one is interested in matching these (feature) point patterns to determine a linear or a nonlinear transformation between the coordinates of the feature point sets. Such transformations capture the changes in the pattern geometry characterized by the given feature point set. An important problem in this context is how can we achieve an unbiased registration of the two or more point sets given i.e., if one arbitrarily chooses any one of the given data sets as a reference, the estimated registration transformation would be biased toward this chosen reference and it would be desirable to avoid such a bias. In this paper, we address this problem of achieving an unbiased (non-rigid) registration of the given multiple (feature point) data sets.

There are numerous techniques in literature on point pattern matching. Iterated closest point (ICP) technique [1] is a popular approach in the vision community to achieve point set registrations using the assumption that closest points correspond to each other. Despite its popularity, ICP has several shortcomings. The non-differentiable cost function associated with ICP introduces the local convergence problem which requires sufficient overlap between the data-sets and a close initialization. Also, a naive implementation of ICP is know to be prone to outliers which prompted more robust variation such as [5].

In [2], Chui et al., present unlabeled non-rigid registration algorithm for a pair of unlabeled point sets. The main strength of their work is the ability to jointly determine the correspondences and non-rigid transformation between two point sets using a combination of techniques such as deterministic annealing and soft-assign. Their work however is not stable in the presence of outliers and computational complexity of the technique is rather high. Recently, Glaunes et al. [6] have approached the problem of matching point sets by converting the problem into an image match-

ing problem. That is, they represent each point-set as a sum of Dirac delta functions and then "soften" the Dirac delta functions by using a Gaussian kernel. Subsequently, they use an L_2 norm on the "images" and minimize this distance (plus a diffeomorphic regularizer). Since "smoothed" delta functions are used to represent each point-set, this approach suffers from not having a noise model and there is essentially no way to estimate the degree of smoothness of the Gaussian kernel.

Another class of approaches involve methods that align the two given point sets without explicitly establishing a point correspondence. The idea is to use a kernel density estimator to estimate the density function fit to each of the two given data sets and then apply an information theoretic measure to compute the (dis)similarity between them. Recently, Tsin and Kanade [9] propose a kernel correlation based point set registration approach where the cost function is proportional to the correlation of two kernel density estimates. In [7], Bing et al., introduced a novel algorithm for rigidly and non-rigidly registering pairs of data sets. They derived a closed form expression for the distance between mixtures of Gaussians reprepresenting the point set data. This analytic cost function is then minimized over the space of desired transformations. Their algorithm is very fast in comparison to existing methods on point set registration and the results shown are quantitatively satisfactory.

All the methods discussed thus far are restricted to registering pairs of point set data. They are not directly extendible to groupwise registration of multiple point sets. Groupwise point set registration has become more popular recently . In [4], a groupwise registration method is presented. However, in their work, the stability of the registration result is not guaranteed in the case of data with outliers, and hence a good stopping criterion is required. Unlike their approach, we do not need to first solve a correspondence problem in order to subsequently solve the non-rigid registration problem.

The rest of paper is organized as follows: In section 2, we present the derivation of the CDF-JS and also the analytic derivative of the CDF-JS with respect to parameters of the non-rigid transformations required to register the multiple data sets. This is followed by a description of the groupwise registration algorithm. In section 3, we present several examples of the groupwise registration algorithm applied to several synthetic as well as real data sets in 2D and 3D. Finally, we conclude in section 4.

2. The Groupwise Registration Model

In this section, we present the details of our proposed non-rigid registration model. The basic idea is to model each point set by a probability distribution, then quantify the distance between these probability distributions using an information-theoretic measure. This similarity measure

is then optimized over the space of coordinate transformations yielding the desired transformations. We will begin by deriving this new similarity measure, namely the CDF-JS.

2.1. CDF-based Jensen-Shannon Divergence

Jensen-Shannon (JS) divergence serves as a measure of cohesion between multiple probability distributions. Given N probability distributions $\mathbf{P_a}$, $a \in \{1, ..., N\}$, the JS-divergence of $\mathbf{P_a}$ is defined by

$$JS(\mathbf{P_1}, \mathbf{P_2}, ..., \mathbf{P_N}) = H(\sum_a \pi_a \mathbf{P_a}) - \sum_a \pi_a H(\mathbf{P_a}) \quad (1)$$

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

It is easy to put the Jensen-Shannon divergence in to the framework of hypothesis testing. To see this, we construct a likelihood ratio between i.i.d. samples drawn from a mixture ($\sum_a \pi_a \mathbf{P_a}$) and i.i.d. samples drawn from a heterogeneous collection of distributions ($\mathbf{P_1}, \mathbf{P_2}, ..., \mathbf{P_N}$) with the samples being indexed by the specific member distribution 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,

$$\Lambda = \frac{\prod_{k=1}^{M} \sum_a \pi_a \mathbf{P_a}(X_k)}{\prod_{a=1}^{N} \prod_{k_a=1}^{N_a} \mathbf{P_a}(X_{k_a})}. \quad (2)$$

In contrast to the typical statistical test relative to a threshold, we seek the maximum of (2) which is equivalent to minimizing the negative logarithm of (2). That is, we seek to maximize the probability that the samples are drawn from the mixture rather than from separate members of the family ($\mathbf{P_1}, \mathbf{P_2}, ..., \mathbf{P_N}$). In the context of groupwise matching of point-sets, this makes eminent sense since maximizing the above ratio is tantamount to increasing the chance that all of the observed point-sets are warped versions of the same underlying warped and pooled data model. *The notion of the pooled data model is defined as follows. In our process of groupwise registration, the warping does NOT have a fixed target data set. Instead, the warping is between the input data sets and an evolving target which we call the pooled model. The target evolves to a fully registered pooled data set at the end of the optimization (of the CDF-JS) process. The pooled model then consists of input data sets which have undergone groupwise matching and are now fully registered with each other.* The connection to the JS-divergence arises from the fact that the negative logarithm of the above ratio (Eqn. 2) asymptotically converges

to the JS-divergence when the samples are assumed to be drawn from the mixture $\sum_a \pi_a \mathbf{P_a}$.

The JS divergence has many important properties. However, extension of this notion to continuous distribution poses some challenges. A straight forward extension of the discrete case to continuous distributions is to replace the Shannon entropy by the differential entropy. This definition raises the following concerns: (1) The entropy of a discrete distribution is always positive, while the differential entropy of a continuous random variable may take any value on the extended real line. Therefore, entropies are not well defined when defining JS for mixed (continuous and discrete) distributions. (2) It is "inconsistent" in the sense that the differential entropy of a uniform distribution in an interval of length a is $\log a$, which is zero if $a = 1$, negative if $a < 1$, and positive if $a > 1$. For additional discussion on some of these issues the reader is referred to [8]. In this work we propose an alternative measure of divergence between multiple random variables based on cumulative distribution functions and dub it the CDF-JS divergence. The main objective of our study is to extend the classical JS-divergence definition to random variables with continuous distributions. The concept we propose here overcomes the problems mentioned above, while retaining many of the important properties of JS-divergence.

The key idea in the definition is to use the cumulative distribution function in place of the density function in Shannon's definition of entropy, resulting in the recently introduced Cumulative Residual Entropy(CRE) in [10] and reproduced here for convenience. Let $\mathbf{x}$ be a random variable in $\mathcal{R}$, the cumulative residual entropy (CRE) of $\mathbf{x}$, is defined as:

$$\mathcal{E}(\mathbf{x}) = -\int_{\mathcal{R}_+} P(|\mathbf{x}| > \lambda) \log P(|\mathbf{x}| > \lambda) \mathrm{d}\lambda \quad (3)$$

Where $\mathcal{R}_+ = \left(\mathbf{x} \in \mathcal{R}; x \geq 0 \right)$. The distribution function is more regular because it is defined in an integral form unlike the density function, which is defined as the derivative of the distribution. The definition also preserves the well established principle that the logarithm of the probability of an event should represent the information content in the event. Based on CRE, we define the CDF-JS divergence between N probability distributions $\mathbf{P_a}$, $a \in \{1, ..., N\}$,

$$\mathcal{C}(\mathbf{P_1}, \mathbf{P_2}, ..., \mathbf{P_N}) = \mathcal{E}(\sum_a \pi_a \mathbf{P_a}) - \sum_a \pi_a \mathcal{E}(\mathbf{P_a}) \quad (4)$$

2.2. CDF-JS Divergence for Groupwise Point-sets Registration

Let the point-set data being registered be denoted by $\{X^a, a \in \{1, ..., N\}\}$. Each point-set X^a consists of points $\{\mathbf{x}_i^a \in \mathcal{R}_D, i \in \{1, ..., n_a\}\}$. Assume that each point set X^a is related to the pooled and registered data via an un-

known function transformation f^a, $\boldsymbol{\mu}^a$ is the set of the transformation parameters associated with each function f^a. To register the all these point data sets, we need to recover these transformation parameters. This problem can modeled as an optimization problem with the objective function being the CDF-JS divergence between the distributions of the deformed point-sets, represented as $\mathbf{P}_a = p(f^a(X^a))$, the atlas construction problem can now be formulated as,

$$\min_{\boldsymbol{\mu}^i} \mathcal{C}(\mathbf{P}_1, \mathbf{P}_2, ..., \mathbf{P}_N) + \lambda \sum_{a=1}^{N} ||Lf^a||^2$$

$$= \min_{\boldsymbol{\mu}^a} \mathcal{E}(\sum_{a=1}^{N} \pi_a \mathbf{P}_a) - \sum_{a=1}^{N} \pi_a \mathcal{E}(\mathbf{P}_a) + \lambda \sum_{a=1}^{N} ||Lf^a||^2 \quad (5)$$

In Eqn. (5), the weight parameter λ is a positive constant, the operator L determines the kind of regularization imposed. For example, L could correspond to a thin-plate spline, a Gaussian radial basis function, etc. Each choice of L is in turn related to a kernel and a metric of the deformation.

Following the approach in [3], 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 $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$ represented by thin-plate spline can be written analytically as: $f(x) = \mathbf{W}\mathbf{U}(\mathbf{x}) + \mathbf{A}\mathbf{x} + \mathbf{t}$. Here $\mathbf{A}\mathbf{x} + \mathbf{t}$ is the linear part of f. The nonlinear part is determined by a $d \times n$ matrix, $\mathbf{W}$. And $\mathbf{U}(\mathbf{x})$ is an $n \times 1$ vector consisting of n basis functions $U_i(\mathbf{x}) = U(\mathbf{x}, \mathbf{x}_i) = U(||\mathbf{x} - \mathbf{x}_i||)$ where $U(r)$ is the kernel function of thin-plate spline. For example, if the dimension is 2 ($d = 2$) and the regularization functional is defined on the second derivatives of f, we have $U(r) = 1/(8\pi)r^2 ln(r)$.

Therefore, the cost function for non-rigid registration can be formulated as an energy functional in a regularization framework, where the regularization term in Eqn. (5) is governed by the bending energy of the thin-plate spline warping and can be explicitly given by $trace(\mathbf{W}\mathbf{K}\mathbf{W}^T)$ where $\mathbf{K} = (K_{ij})$, $K_{ij} = U(\mathbf{x}_i, \mathbf{x}_j)$ 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 $\mathbf{W}$.

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

2.3. Estimating the Empirical CDF-JS

We will use the parzen window technique to estimate the cumulative distribution function and its derivative. The calculation of CDF-JS divergence requires the estimation

of the cumulative probability distributions of each point-set. Without lost of generality, we only discuss the derivation for the 2D case, which can be extended to 3D case easily. For each point $\mathbf{x}_i^a$ with the coordinates $[x_i^a, y_i^a]$ in the point set X^a, it is transformed by the function f^a to $f^a(\mathbf{x}_i^a, \boldsymbol{\mu}^a) = [f^a(x_i^a, \boldsymbol{\mu}^a), f^a(y_i^a, \boldsymbol{\mu}^a)]$. Let $\beta^{(3)}$ be a cubic spline Parzen window. The smoothed probability density function $p^a(l, k; \boldsymbol{\mu}^a)$ of the point-set $X^a, a \in \{1, ..., N\}\}$ is given by

$$p^a(l, k; \boldsymbol{\mu}^a) = \alpha^a \sum_i^{n_a} \beta^{(3)}\Big(l - \frac{f^a(x_i^a, \boldsymbol{\mu}^a) - x^0}{\triangle b_X}\Big) \\ \beta^{(3)}\Big(k - \frac{f^a(y_i^a, \boldsymbol{\mu}^a) - y^0}{\triangle b_Y}\Big) \quad (6)$$

where α^a is a normalization factor that ensures $\int \int p(l, k) \mathrm{d}l \mathrm{d}k = 1$, $[l, k]$ are the coordinate values in the X and Y axis respectively, the transformed point coordinates $[f^a(x_i^a, \boldsymbol{\mu}^a), f^a(y_i^a, \boldsymbol{\mu}^a)]$ is normalized by the minimum coordinate value, x^0, y^0, and the range of each bin, $\triangle b_X$, $\triangle b_Y$. From the density function, we can calculate the cumulative residual distribution function by the formula $P^a(l > \lambda, k > \gamma; \boldsymbol{\mu}^a) = \int_\lambda^\infty \int_\gamma^\infty p^a(l, k; \boldsymbol{\mu}^a) \mathrm{d}l \mathrm{d}k$, where $\lambda \in L_x, \gamma \in L_y$, L_x, L_y are discrete sets of coordinate values in the X and Y axis respectively. To express it in further detail, we have the following,

$$P^a(\lambda, \gamma; \boldsymbol{\mu}^a) = \alpha^a \sum_i^{n_a} \int_\lambda^\infty \beta^{(3)}\Big(l - \frac{f^a(x_i^a, \boldsymbol{\mu}^a) - x^0}{\triangle b_X}\Big) \mathrm{d}l \\ \int_\gamma^\infty \beta^{(3)}\Big(k - \frac{f^a(y_i^a, \boldsymbol{\mu}^a) - y^0}{\triangle b_Y}\Big) \mathrm{d}k \\ = \alpha^a \sum_i^{n_a} \Phi\Big(\lambda - \frac{f^a(x_i^a, \boldsymbol{\mu}^a) - x^0}{\triangle b_X}\Big) \Phi\Big(\gamma - \frac{f^a(y_i^a, \boldsymbol{\mu}^a) - y^0}{\triangle b_Y}\Big) \quad (7)$$

where $\Phi(\ )$ is the cumulative residual function of cubic spline kernel which is defined as follows,

$$\Phi(v) \quad = \quad \int_v^\infty \beta^{(3)}(u) \mathrm{d}u$$

Note that $\frac{\mathrm{d}\Phi(u)}{\mathrm{d}u} = -\beta^{(3)}(u)$.

Having specified the distribution function of the data, we can then rewrite Eqn. (5) as follows, (For simplicity, we choose $\pi_a = \frac{1}{N}, \forall a = \{1, 2, ..., N\}$.)

$$\mathcal{C}(\mathbf{P}_1, \mathbf{P}_2, ..., \mathbf{P}_N) = \mathcal{E}(\sum_{a=1}^N \pi_a \mathbf{P}_a) - \sum_{a=1}^N \pi_a \mathcal{E}(\mathbf{P}_a) \\ = -\sum_\lambda \sum_\gamma P \log P + \frac{1}{N} \sum_\lambda \sum_\gamma P^a \log P^a \quad (8)$$

where P is the cumulative residual distribution function for the density function $\frac{1}{N} \sum p^a(l, k; \boldsymbol{\mu}^a)$, which can be ex-

pressed as

$$P(\lambda, \gamma; \boldsymbol{\mu}^a) = \frac{1}{N} \sum_{a=1}^N \alpha^a \sum_i^{n_a} \Phi\Big(\lambda - \frac{f^a(x_i^a, \boldsymbol{\mu}^a) - x^0}{\triangle b_X}\Big) \\ \Phi\Big(\gamma - \frac{f^a(y_i^a, \boldsymbol{\mu}^a) - y^0}{\triangle b_Y}\Big) \quad (9)$$

2.4. Optimizing the CDF-JS Divergence

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 CDF-JS divergence (our cost function):

$$\nabla \mathcal{C} = [\frac{\partial \mathcal{C}}{\partial \boldsymbol{\mu}^1}, \frac{\partial \mathcal{C}}{\partial \boldsymbol{\mu}^2}, ..., \frac{\partial \mathcal{C}}{\partial \boldsymbol{\mu}^N}] \quad (10)$$

Each component of the gradient maybe found by differentiating Eqn (8) with respect to the transformation parameters. It can be easily shown that $\frac{\partial P(\lambda, \gamma; \boldsymbol{\mu}^a)}{\partial \boldsymbol{\mu}^a} = \frac{1}{N} \frac{\partial P^a(\lambda, \gamma; \boldsymbol{\mu}^a)}{\partial \boldsymbol{\mu}^a}$. Based on these facts, it is straight forward to derive the gradient of the CDF-JS divergence with respect to the transformation parameters $\boldsymbol{\mu}^a$, which is given by

$$\frac{\partial \mathcal{C}}{\partial \boldsymbol{\mu}^a} = -\sum_\lambda \sum_\gamma [1 + \log P] \frac{\partial P(\lambda, \gamma; \boldsymbol{\mu}^a)}{\partial \boldsymbol{\mu}^a} \\ + \frac{1}{N} \sum_\lambda \sum_\gamma [1 + \log P^a] \frac{\partial P^a(\lambda, \gamma; \boldsymbol{\mu}^a)}{\partial \boldsymbol{\mu}^a} \quad (11) \\ = \sum_\lambda \sum_\gamma \frac{1}{N} \frac{\partial P^a(\lambda, \gamma; \boldsymbol{\mu}^a)}{\partial \boldsymbol{\mu}^a} \log \frac{P^a}{P}$$

As a byproduct of the groupwise registration using the CDF-JS divergence, we get the atlas of the given population of data sets, which is simply obtained by substituting the estimated transformation functions f_a in to the formula for the atlas $p(\mathcal{A}) = \sum_{a=1}^N \pi_a \mathbf{P}_a$. Note that our algorithm can be applied to yield a biased registration in situations that demand such a solution. This is achieved by fixing one of the data sets (say the reference) and estimating the transformation from this to the novel scene data. We will present experimental results on point-set alignment between two given point-sets as well as atlas construction from multiple point-sets in the next section.

3. Experiment Results

We now present experimental results on the application of our algorithm to both synthetic and real data sets. First, to demonstrate the robustness and accuracy of our algorithm, we show the alignment results by applying the CDF-JS divergence to the matching of pairs of point-sets. Then,

we will present the atlas construction results achieved via groupwise registration of multiple point sets, in the second part of this section.

3.1. Pair-wise Point Set Alignment

First, 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 sensitive 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 outliers 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 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 1. A 3D example is also presented in Figure 2.
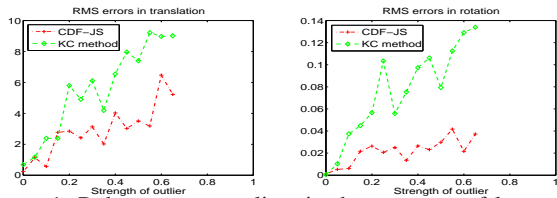

Figure 1. Robustness to outliers in the presence of large noise. Errors in estimated rigid transform vs. proportion of outliers $((\tau - \rho)/(\rho))$ for both our method and KC method.
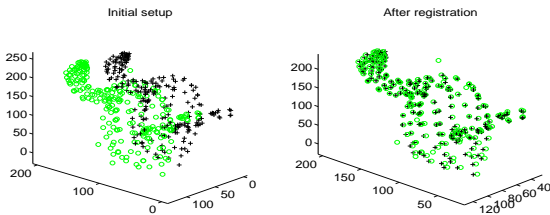

Figure 2. Robustness test on 3D swan data. 'o' and '+' indicate the model and scene points respectively. Note that the scene point-set is corrupted by noise and outliers.

3.2. Groupwise Registration for Atlas Construction Results

In this section, we begin with a simple yet illustrative example of the application of our algorithm for 2D atlas construction. After this example, we describe a 3D implementations on real hippocampal data sets. The structure we are interested in this experiment is the corpus callosum as it appears in MR brain images. Constructing an atlas for the corpus callosum and subsequently analyzing the individual shape variation from "normal" anatomy has been regarded as potentially valuable for the study of brain diseases such as agenesis of the corpus callosum(ACC), and fetal alcohol syndrome(FAS).

We manually extracted points on the outer contour of the corpus callosum from six normal subjects, (as shown Figure 3, indicated by "o"). The recovered deformation between each point-set and the mean shape are superimposed on the first two rows and the left image in third row in Figure 3. The resulting atlas (mean point-set) is shown in third row of Figure 3, and is superimposed over all the point-sets. As we described earlier, all these results are computed simultaneously and automatically. This example clearly demonstrate that our joint matching and atlas construction algorithm can simultaneously align multiple shapes (modeled by sample point-sets) and compute a meaningful atlas/mean shape.

Next, we present results on 3D hippocampal point-sets. Four 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 from the 3D brain MRI scans of 4 subjects. The point-sets differ in shape, with the number of points $450, 421, 376, 307$ in each point-set respectively. In the first four images of Figure 4, the recovered nonrigid deformation between each hippocampal point-set to the atlas is shown along with a superimposition on all of the original data sets. In second row of the Figure 4, we also show the scatter plot of original point-sets along with all the point-sets after the non-rigid warping. An examination of the two scatter plots clearly shows the efficacy of our recovered non-rigid warping. Note that validation of what an atlas shape ought to be in the real data case is a difficult problem and we relegate its resolution to a future paper.

4. Conclusions

In this paper, we presented a novel and robust algorithm that utilize an information theoretic measure, namely CDF based Jensen-Shannon divergence, to simultaneously register multiple unlabeled point-sets (represented by probability distribution). The primary technical challenge in groupwise non-rigid registration of point-sets is the correspondence problem. Our work avoids the correspondence problem by minimizing the CDF based Jensen-Shannon (JS) di-
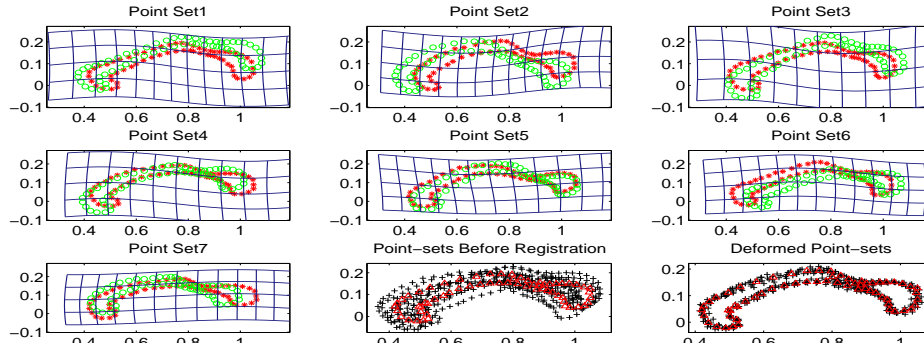
Figure 3. Experiment results on 7 2D corpus collasum point sets. The first two rows and the left image in third row show the deformation of each point-set to the atlas, superimposed with initial point set (show in 'o') and deformed point-set (shown in '*'). Middle image in the third row: The estimated atlas is shown superimposed over all the point-sets. Right: The estimated atlas is shown superimposed over all the deformed point-sets.
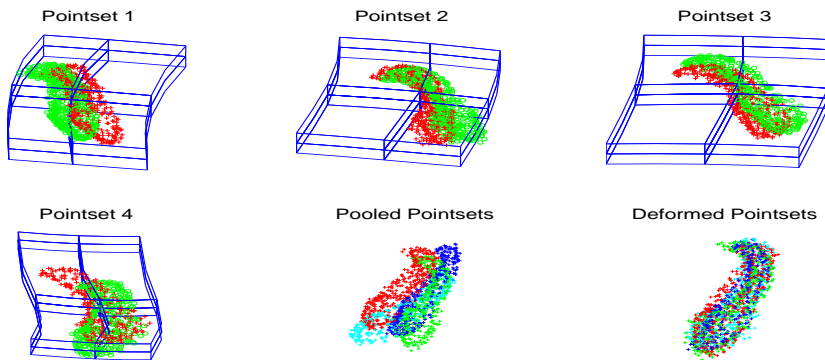

Figure 4. Atlas construction from four 3D hippocampal point sets. The first row and the left image in second row shows the deformation of each point-set to the atlas (represented as cluster centers), superimposed with initial point set (show in green 'o') and deformed point-set (shown in red '+'). Left image in the second row: Scatter plot of the original four hippocampal point-sets. Right: Scatter plot of all the warped point-sets.

vergence between the point sets represented as probability distributions. Other advantages of using the CDF-JS divergence over existing methods in literature for groupwise point-sets non-rigid registration is that, the CDF-JS divergence is symmetric, and allows for use of unequal cardinality of the given point sets to be registered. The cost function optimization is achieved very efficiently by computing analytic gradients of the same and utilizing them in a quasi-Newton scheme. We compared our algorithm performance with competing methods on real and synthetic data sets and showed significantly improved performance in the context of robustness to noise and outliers in the data. Experiments were depicted with both 2D and 3D point sets from medical and non-medical domains. Our future work will focus on generalizing the non-rigid deformations to diffeomorphic mappings.

References

[1] P. J. Besl and N. D. McKay. A method for registration of 3-D shapes. *IEEE Trans. Patt. Anal. Machine Intell.*, 14(2):239–256, February 1992. 1

[2] H. Chui and A. Rangarajan. A new algorithm for non-rigid point matching. In *CVPR2000*, pages 2044–2051, 2000. 1

[3] H. Chui and A. Rangarajan. A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding (CVIU)*, 89:114–141, 2003. 3

[4] 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.*, 26(2):160–172, 2004. 2

[5] A. W. Fitzgibbon. Robust registration of 2D and 3D point sets. In *BMVC2001*, 2001. 1

[6] J. Glaunes, A. Trouve, and L. Younes. Diffeomorphic matching of distributions: A new approach for unlabelled point-sets and sub-manifolds matching. In *CVPR*, volume 2, pages 712–718, 2004. 1

[7] B. Jian and B. Vemuri. A robust algorithm for point set registration using mixture of gaussians. In *ICCV2005*, pages 1246–1251, 2005. 2

[8] G. Jumarie. *Relative Information*. Springer, 1990. 3

[9] Y. Tsin and T. Kanade. A correlation-based approach to robust point set registration. In *ECCV2004(3)*, pages 558–569, 2004. 2

[10] F. Wang, B. C. Vemuri, M. Rao, and Y. Chen. A new &
     robust information theoretic measure and its application to
     image alignment. In *IPMI*, pages 388–400, 2003. 3