Group-wise Point-set registration using a novel CDF-based Havrda-Charvát Divergence

Ting Chen, Baba C. Vemuri, Anand Rangarajan and Stephan J. Eisenschenk *†

Abstract

This paper presents a novel and robust technique for group-wise registration of *point sets* with unknown correspondence. We begin by defining a Havrda-Charvát (HC) entropy valid for cumulative distribution functions (CDFs) which we dub the HC Cumulative Residual Entropy (HC-CRE). Based on this definition, we propose a new measure called the CDF-HC divergence which is used to quantify the dis-similarity between CDFs estimated from each point-set in the given population of point sets. This CDF-HC divergence generalizes the CDF based Jensen-Shannon (CDF-JS) divergence introduced earlier in the literature, but is much simpler in implementation and computationally more efficient.

A closed-form formula for the analytic gradient of the cost function with respect to the non-rigid registration parameters has been derived, which is conducive for efficient quasi-Newton optimization. Our CDF-HC algorithm is especially useful for *unbiased point-set* atlas construction and can do so without the need to establish correspondences. Mathematical analysis and experimental results indicate that this CDF-HC registration algorithm outperforms the previous *group-wise* point-set registration algorithms in terms of efficiency, accuracy and robustness.

1 Introduction

Point-set registration is a widely encountered problem in several fields namely, computer vision, computer graphics, medical imaging and pattern recognition. In computer vision, it is encountered in image mosaicing to form panoramas and in computer graphics it is required for fusing 3D range data sets obtained from different vantage points to form a 3D model. In medical imaging, registration is necessary to match landmarks in volume (MRI, CT etc.) scans for the purposes

*T. Chen, B. C. Vemuri and A. Rangarajan are with Department of CISE, University of Florida, Gainesville, FL 32601. S. J. Eisenschenk is with Department of Neurology, University of Florida.

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

of disease diagnosis or in the construction of image/shape atlases. Finally, in pattern recognition, point pattern matching is used to correlate feature clusters to assess the similarity between them.

The key problem in point-set registration regardless of the dimension in which the points are embedded in is to estimate the transformation between the coordinates used to represent the points in each set. The transformation may be characterized as being linear or nonlinear, parameterized or non-parameterized. The literature has many techniques to solve for the linear and nonlinear transformations required to register the point sets. Below, we briefly discuss some of the prominent methods and then establish motivation for the work reported here.

The prior work in point-set registration can be traced back to Baird's effort in 1985 [1], wherein a technique was proposed for pair-wise shape registration under a similarity transformation. He constructed a linear programming model to solve for the registration parameters and made effective use of the feasibility-testing algorithms, such as the Simplex algorithm and the Soviet ellipsoid algorithm, to allow for a systematic search for the correspondences. Theoretical analysis indicated that the runtime of the algorithm is asymptotically quadratic in the number of points n, but in practice it is linear in n for $n < 100$. However, rigid registration is too restrictive a requirement. Since then, abundant research on pair-wise non-rigid point-set registration can be found in literature. For instance, Belongie *et al.* [2] aimed at non-rigidly registering two shapes represented by points using shape contexts, by first solving for the correspondences. Their method is more tuned to shape indexing than registration. Chui and Rangarajan [3] proposed a method that jointly recovers the correspondence and the non-rigid registration between two point sets via deterministic annealing and softassign. Their work requires outlier rejection parameters to be specified and the use of deterministic annealing frequently results in a slow to converge algorithm in practice. Note that these two previously discussed non-rigid registration methods employ non-rigid spatial mappings, particularly thin-plate splines (TPSs) [5, 7] as the deformation model. In recent work, Glaunes *et al.* [4] attempt solving the point-set matching problem in a diffeomorphism setting. This successfully overcomes the drawbacks such as local folds and reflections induced by spline based models as in [6]. However, it requires a large amount of computation in 3D due to the need to compute a spatial integral. Furthermore, the method has not been extended to the group-wise setting.

Alternatively, there exists a class of methods that achieves more robustness with unknown correspondences in the presence of outliers. The general idea for this class of methods is to represent each point-set by a probability density function and compute the dis-similarity between

them using an information theoretic measure. This class of methods is closely related to our work reported here. The most illustrative example is the method proposed in [8] by Wang *et al.*, who attempted minimizing the relative entropy between two distributions estimated from the point sets w.r.t. the registration parameters, so as to register the two point sets. The main drawback of this approach is that only rigid pair-wise registration problem is addressed. Besides this, Tsin and Kanade [9] proposed a Kernel Correlation (KC) based point-set registration algorithm by maximizing the kernel correlation cost function over the transformation parameters, where the cost function is proportional to the correlation of the two kernel densities estimated. Jian and Vemuri in [10] modeled each of the point sets as a Gaussian mixture model (GMM) and minimized the $L2$ distance over the space of transformation parameters, yielding the desired transformation. While the method has attractive speed and robustness properties, the $L2$ distance is not a divergence measure and the overlay of the point-sets is not modeled, making it difficult to extend this to the unbiased registration of multiple point-sets.

To summarize, in all the techniques discussed thus far, one of the two given point sets is fixed as a reference which definitely leads to a bias in the deformation toward the chosen data set. Moreover, all the point-set registration methods mentioned above are designed to achieve pair-wise point-set registration, and are not easily generalizable to achieve group-wise registration of multiple point sets. Considering group-wise alignment algorithms, most of the efforts were dedicated to group-wise *image* registrations, i.e. constructing *image* atlas. For instance, in [11, 12, 13, 14], several non-rigid group-wise *image* registration methods were proposed. However, it is a nontrivial task to extend these elegant techniques to group-wise *point-set* registration. Therefore, we will not further discuss these *image* based methods in this paper and only focus on *point-set* atlas construction instead. Before moving to group-wise point-set registration, we need to briefly mention that rather than the feature point representation scheme, shape representation using curves or surfaces [15, 16] has also received attention in the literature. Since statistical shape analysis in curve/surface space is very difficult, methods using this representation have usually resorted to computing means etc. of spline parameters (used for curve/surface representation) [17] which is an extrinsic approach.

A $2D$ average shape modeling technique with automatic shape clustering and outlier detection was proposed by Duta *et al.* in [18]. Their matching method took the point sets extracted from the shape contours as input and performed pair-wise registration of two point sets without any requirements of setting the initial position/scale of the two objects or needing any manually

tuned parameters. However, the Procrustes analysis procedure imposed in their model requires the knowledge of correspondences. In [19], a group-wise point-set registration algorithm was proposed as a generalization of [3], but this method has the same shortcomings as [3] in that an explicit correspondence problem needs to be solved. Furthermore, the method is slow due to the use of deterministic annealing. In the recent past, several research articles on group-wise point-set registration have been published by Wang *et al.* [21, 20]. The main strength of their work is in simultaneously group-wise registering the data and computing the mean atlas shape for multiple unlabeled point sets without choosing any specific data set as a reference or solving for correspondences, thus yielding an unbiased group-wise registration as well as an atlas. Their approach is to minimize the JS divergence among the PDFs in [21] (or CDFs in [20]), estimated from the given population of point sets, with TPSs adopted as the deformation model. The main claim in [20] is that the CDF-JS is more immune to noise than PDF-JS because of the robustness property of the CDF (being an integral measure). However, the main bottleneck in their CDF-based work is the computational cost and the complexity in implementation, since they use cubic spline Parzen windows to estimate the PDF and CDF. Hence, a numerical approximation is involved in CDF estimation and extensive computation is involved in CDF entropy estimation since the entropies and their derivatives are not available in closed form. Unlike their work, in this paper we generalize the CDF-JS and develop an algorithm that is computationally much faster and more simple and accurate from an implementation perspective (than the work in [20]), without losing the inherent statistical robustness or accuracy in CDF based models. We will compare the computational complexity of CDF-JS and CDF-HC in section 4 and show that CDF-HC tremendously reduces the complexity compared to CDF-JS. We also demonstrate the robustness and accuracy of the CDF-HC method by showing a set of experimental results on CDF-HC, CDF-JS and PDF-JS methods.

The rest of the paper is organized as follows: In section 2, we present the definition of CDF-HC divergence and introduce our point-set registration model. Section 3 contains of the description of a novel technique for estimating the empirical CDF-HC which is used for the implementation of our algorithm. The algorithm is then analyzed and validated experimentally in Section 4 and we present concluding remarks in Section 5.

2 Point Pattern Registration Model

In this section, we present the details of our proposed non-rigid point-set registration model. The basic idea is to model each point-set by a survival function (complement of a cumulative distribution function abbreviated as CDF), and then quantify the distance between these probability distributions via an information-theoretic measure. This dis-similarity measure is then optimized over the space of coordinate transformation parameters. We will begin by deriving this new dissimilarity measure, namely the CDF-HC divergence.

2.1 Definition of CDF-HC divergence

The CDF-HC divergence in our paper parallels the definition of CDF-JS [20] and is defined over the Havrda-Charvát (HC) Cumulative Residual Entropy (HC-CRE), the definition of which will be given later. For convenience, we reproduce the definition of CRE, CDF-JS and Havrda Charvát differential entropy (HC) here,

Definition 1 (CRE, [22]): Let X be a random vector in R^d, the CRE of X is defined by

$$\mathcal{E}(X) = -\int_{R^d_+} (P(|X| > \lambda) \log P(|X| > \lambda)) d\lambda \tag{1}$$

where $X = \{x_1, x_2, \ldots, x_d\}$, $\lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_d\}$, and $|X| > \lambda$ means $|x_i| > \lambda_i$, $R^d_+ = \{x_i \in R^d; x_i \geq 0; i \in \{1, 2, \ldots, d\}\}$. .

Definition 2 (CDF-JS, [20]): Given N cumulative probability distributions P_k, $k \in \{1, \ldots, N\}$, the CDF-JS divergence of the set $\{P_k\}$ is defined as

$$JS(P_1, P_2, \ldots, P_N) = \mathcal{E}(\sum_k \pi_k P_k) - \sum_k \pi_k \mathcal{E}(P_k) \tag{2}$$

where $0 \leq \pi_k \leq 1$, $\sum_k \pi_k = 1$, and $\mathcal{E}$ is the Cumulative Residual Entropy (CRE) defined in [22].

Definition 3 (HC, [23]): The Havrda Charvát entropy is defined as

$$H_\alpha(X) = -\sum_{i=1}^{n} (\alpha - 1)^{-1} (p^\alpha(x_i) - p(x_i)) \tag{3}$$

where $x_1, \ldots, x_n$ are possible values for the random variable X, p denotes the probability mass function of X and α is its inherent parameter.

Now we define HC-CRE by replacing the density function in Eqn. 3 with the survival function. This definition parallels the Cumulative Residual Entropy $\mathcal{E}$ which is based on CDFs.

Definition 4 (HC-CRE): Let X be a random vector in R^d: We define the HC-CRE of X by

$$\mathcal{E}_{\mathcal{H}}(X) = -\int_{R_+^d} (\alpha - 1)^{-1}(P^{\alpha}(|X| > \lambda) - P(|X| > \lambda))d\lambda \tag{4}$$

where X, λ, and R_+^d are defined as in Definition 1.

Note that we borrow the notation $|X| > \lambda$ and its definition from [22], wherein $|X| > \lambda$ is taken to mean that $|x_i| > \lambda_i$ for each i. Since the CDF in 2D and above depends on the axis directions, we appear to have a problem. However, our CDF-HC formulation is independent of the axis definition, that is, any type of definition for $|X| > \lambda$ out of all the 2^d permutations is equivalent to each other, because the estimated cumulative distribution functions are equivalent after coordinate axis reflection of X. Again, in this paper we only refer to $|x_i| > \lambda_i$ as the definition of $|X| > \lambda$ for the simplicity of discussion.

The relationship between HC-CRE and CRE is straightforward. We now show that the HC-CRE approaches CRE as α tends to 1.

Property 1 (Convergence):

$$\lim_{\alpha \to 1} \mathcal{E}_{\mathcal{H}}(X) = \mathcal{E}(X) \tag{5}$$

Proof We use the familiar L'Hôpital rule for the limit. We take the derivative of the numerator and denominator with respect to α and evaluate the limit as α tends to one.

$$\begin{aligned}
& \lim_{\alpha \to 1} \mathcal{E}_{\mathcal{H}}(X) \\
= & \lim_{\alpha \to 1} -\int \frac{P^{\alpha}(X > \lambda) - P(X > \lambda)}{\alpha - 1} d\lambda \\
= & \lim_{\alpha \to 1} -\int P^{\alpha}(X > \lambda) \log P(X > \lambda) d\lambda \\
= & -\int P(X > \lambda) \log P(X > \lambda) d\lambda \\
= & \mathcal{E}(X)
\end{aligned}$$

Based on HC-CRE, we define the CDF-HC divergence between N probability distributions P_k, $k \in \{1, 2, \ldots, N\}$. Henceforth in this paper we will consider X to be in the R_+ domain and write X instead of $|X|$.

Definition 5 (CDF-HC divergence): The CDF-HC divergence is defined as

$$HC(P_1, P_2, \ldots, P_N) = \mathcal{E}_{\mathcal{H}}(\sum_k \pi_k P_k) - \sum_k \pi_k \mathcal{E}_{\mathcal{H}}(P_k) \tag{6}$$

We can rewrite $HC(P_1, P_2, \ldots, P_N)$ by substituting Eqn. 4 into Eqn. 6 to get a simplified version of Definition 5.

Property 2 (Simplification): Let P be the convex combination of $\{P_k\}$: $P = \sum_k \pi_k P_k$. We can simplify HC to be

$$\begin{aligned} & HC(P_1, P_2, \ldots, P_N) \\ = \; & -(\alpha-1)^{-1}\Big(\int_{R_+^d} P^\alpha(X>\lambda)d\lambda - \sum_k \pi_k \int_{R_+^d} P_k^\alpha(X_k>\lambda)d\lambda\Big) \end{aligned} \tag{7}$$

We use this simplified CDF-HC divergence formula in all our computations and implementations.

2.2 Group-wise Point-set Registration Model

Denote the N point sets to be registered as $\bar{X}_k, k \in \{1, 2, \cdots, N\}$. Each point-set $\bar{X}_k$ consists of points $\bar{\mathbf{x}}_k^i \in R^d, i \in \{1, 2, \cdots, D_k\}$, D_k being the number of points in the point-set $\bar{X}_k$. Assume each point-set $\bar{X}_k$ is related to the finally registered data X_k via an unknown transformation function f_k, and let $\mu_k \in R^{D_k} \times R^d$ be the set of transformation parameters associated with each function f_k, i.e. $X_k = f_k(\bar{X}_k)$ and each X_k consists of points $\mathbf{x}_k^i \in R^d, i \in \{1, 2, \cdots, D_k\}$.

To group-wise register all the given N point sets, we need to recover the transformation parameters μ_k, $k \in \{1, 2, \cdots, N\}$. This problem can be modeled as an optimization problem with the objective function being the CDF-HC divergence between the N survival functions computed from the deformed point sets, represented as $P_k = \mathcal{P}(X_k), k \in \{1, 2, \cdots, N\}$.

The group-wise registration problem a.k.a. the atlas construction problem can now be formulated as,

$$\begin{aligned} & \min_{\mu_k} HC(P_1, P_2, \ldots, P_N) + \eta \sum_{k=1}^{N} ||Lf_k||^2 \\ = \; & \min_{\mu_k} (\mathcal{E}_{\mathcal{H}}(\sum_k \pi_k P_k) - \sum_k \pi_k \mathcal{E}_{\mathcal{H}}(P_k)) + \eta \sum_{k=1}^{N} ||Lf_k||^2 \end{aligned} \tag{8}$$

In Eqn. 8, a standard regularization of the transformation functions $\{f_k\}$ is used. The parameter η is a positive constant, which acts as the trade off between the two energies. It prevents the data set from collapsing into a single point. By tuning η, we can control the degree of deformation, the demonstration of which is shown in the experiment.

Let L denote the regularization operator. For example, L could be a differential operator such as a second order linear differential operator corresponding to the thin-plate spline (TPS). In our implementation, we choose TPS as the non-rigid deformation. Given a set of control points in R^d, we write TPS as a general non-rigid mapping $f : R^d \rightarrow R^d$, such that $f(\mathbf{x}) = WU(\mathbf{x}) + A[\mathbf{x};\mathbf{I}]$, where $A[\mathbf{x};\mathbf{I}]$ is the affine part of the TPS transformation and the non-rigid part is determined by

the transformation parameters stored in the $d \times n$ matrix W. Here $U(\mathbf{x})$ is an $n \times 1$ vector consisting of n basis functions $U_i(\mathbf{x}) = U(\mathbf{x}, \mathbf{x}_i) = U(||\mathbf{x} - \mathbf{x}_i||)$. Let $r = ||\mathbf{x} - \mathbf{x}_i||$. Then $U(r)$ is the reproducing kernel of the thin-plate spline. Note that there exists a boundary condition $PW^T = 0$ [7] for TPS, where P is a $(d+1) \times n$ matrix with the first column being ones and the rest of the columns being the coordinates of the points in the point-set. This condition ensures that the non-rigid part of the transformation is zero at infinity. Using the constraint that W lies in the null-space of P, the dimension of W is reduced to $(d+1) \times d$. Hence, the TPS transformation parameter $\mu_k = [A, W]$ is a $d \times n$ matrix. Therefore, the objective function for non-rigid registration can be formulated as an energy functional in a regularization framework, where the bending energy of the TPS warping is explicitly given by $trace(WKW^T)$, with $K = (K_{ij})$, and $K_{ij} = U(\mathbf{x}_i, \mathbf{x}_j)$ depending on the spline kernel and the control point sets.

Having introduced the objective function and the transformation model, the task now is to design an efficient way to estimate the empirical CDF-HC divergence and derive the analytic gradient of the estimated divergence in order to efficiently achieve a good (albeit suboptimal) solution.

3 Estimating the Empirical CDF-HC

In this section, we propose a technique for estimating the empirical CDF-HC. As mentioned previously, in [20] the Parzen window technique is used. Specifically a cubic spline Parzen window is used to estimate the smoothed probability density function p of a given point-set. The cumulative residual distribution function is computed by integrating over p. This is a constructive method and requires numerical integration which can impact performance (as we will see in the experiments). However, in this paper, we present a novel technique to construct the CDF surface using the Dirac Mixture Model (DMM) which is computationally faster and simpler from an implementation perspective. We then derive the analytic gradient of CDF-HC, when the parameter α for HC equals to 2. Without loss of generality, we only discuss the derivation for the 2D case, since the derivation can be easily extended to the 3D case.

3.1 The Dirac Mixture Model

Mixture models [24] have been an effective tool for modeling shapes [10, 21, 25], especially when the shapes are represented by feature points or landmarks. Here we resort to the Dirac Mixture

Model for computing the cumulative residual function for a given point-set X_k. The DMM is obtained by constructing a Dirac Delta function for each point in the given point-set. A DMM is defined as a convex combination of Dirac Delta functions $\mathcal{D}(\mathbf{x}_k^i|m^i)$, where m^i is the mean vector of $x^i, i \in 1, 2, \cdots, D_k$. The probability density function is explicitly given as $p(\mathbf{x}_k^i) = \sum_{i=1}^{D_k} \phi^i \mathcal{D}(\mathbf{x}_k^i|m^i)$, where ϕ^i are the weights associated with the component functions. For each component, the mean vector is given by the location of each point. Without prior information, we can assume each component has the same weight $\frac{1}{D_k}$. The spatial transformation used in registration - the TPS in our case - is applied to the mean vector which coincides with the location of the points.

From the DMM's use of Dirac delta functions, we see that each such delta function can be integrated to obtain a Heaviside Step function for each point, that is by constructing a transformed Heaviside Step function H^i for every point $\mathbf{x}^i \in R^d, i \in \{1, 2, \cdots, D_k\}$. The CDF surface for the point-set X_k is achieved by summing up all the H^i functions.

$$P_k(X_k > \lambda) = \frac{1}{D_k} \sum_{i}^{D_k} H^i(\mathbf{x}, \mathbf{x}^i) \tag{9}$$

where the definition of $H(t, t_0)$ in one-dimension is given by

$$H(t, t_0) = \begin{cases} 1 & t \leq t_0 \\ 0 & t > t_0 \end{cases} \tag{10}$$

The separability of the Heaviside function in higher dimensions allows us to turn the multi-dimensional Heaviside function into the multiplication of one-dimensional Hs. Also note that the integral of $H(t, t_0)$, from 0 to ∞, is t_0. This simple but important property will be used in gradient computation in the next subsection.

3.2 Gradient Computation

Now we will derive the analytic gradient of CDF-HC when $\alpha = 2$. Note that it is the only case for which CDF-HC has a concise expression for estimating the CDF and a closed form solution for the gradient of the cost function. This will have ramifications in the optimization strategy since CDF-HC is not available in closed form for $\alpha \neq 2$. Therefore, we mainly focus on $\alpha = 2$ in this paper.

Let $\alpha = 2$ in Eqn. 7. We get

$$\begin{aligned} & HC(P_1, P_2, \ldots, P_N) \\ = \; & -\int_{R_+^d} P^2(X > \lambda) d\lambda + \sum_k \pi_k \int_{R_+^d} P_k^2(X_k > \lambda) d\lambda \end{aligned} \tag{11}$$

We begin by deriving the detailed formula for each term of this equation. Take the 2D case with $d = 2$ and $\mathbf{x}^i = [x_i, y_i]$ as an example in our discussion. First, we compute the second term of Eqn. 11. Since the Heaviside Function is separable, for a given point-set X^k, we have

$$\begin{aligned} P_k(X_k > \lambda) &= \frac{1}{D_k}\sum_{i=1}^{D_k} H(x, y, x_i, y_i) \\ &= \frac{1}{D_k}\sum_{i=1}^{D_k} H(x, x_i)H(y, y_i) \end{aligned} \tag{12}$$

Hence,

$$\begin{aligned} & P_k^2(X_k > \lambda) \\ = & \frac{1}{{D_k}^2}\sum_{i=1}^{D_k} H(x, y, x_i, y_i)\sum_{j=1}^{D_k} H(x, y, x_j, y_j) \\ = & \frac{1}{{D_k}^2}\sum_{i,j} H(x, \min(x_i, x_j))H(y, \min(y_i, y_j)) \end{aligned} \tag{13}$$

Now, we compute the first term, i.e. the convex combination term, of Eqn. 11. Recall that

$$P(X > \lambda) = \sum_{k=1}^{N} \pi_k P_k(X_k > \lambda) \tag{14}$$

We have

$$\begin{aligned} & P^2(X > \lambda) \\ = & (\pi_1 P_1(X_1 > \lambda) + \ldots + \pi_N P_N(X_N > \lambda))^2 \\ = & \sum_k (\pi_k P_k(X_k > \lambda))^2 + \sum_{l \neq s} \pi_l \pi_s P_l(X_l > \lambda) P_s(X_s > \lambda) \end{aligned} \tag{15}$$

The first part of Eqn. 15 coincides with Eqn. 13. For the second part, we have a similar expression, i.e.

$$\begin{aligned} & P_l(X_l > \lambda)P_s(X_s > \lambda) \\ = & \frac{1}{D_l}\sum_{i=1}^{D_l} H(x, y, x_i, y_i)\frac{1}{D_s}\sum_{j=1}^{D_s} H(x, y, x_j, y_j) \\ = & \frac{1}{D_l D_s}\sum_{i,j} H(x, \min(x_i, x_j))H(y, \min(y_i, y_j)) \end{aligned} \tag{16}$$

where

$$\min(x, y) = -\frac{1}{\beta}\log(e^{-\beta x} + e^{-\beta y}) \tag{17}$$

Note that we employ an analytical form of the $\min$ operator in our previous expressions so that we get the analytical gradient.

$$\frac{\partial \min(x,y)}{\partial x} = \{ \begin{array}{ll} \frac{e^{-\beta x}}{e^{-\beta x}+e^{-\beta y}} & x \neq y \\ 1 & x = y \end{array} \tag{18}$$

which is later used in this paper. It is obvious that the two equations, Eqn. 13 and Eqn. 16, share a uniform expression except for a scaling constant c.

$$g(x,y) = c\sum_{i,j} H(x, \min(x_i, x_j))H(y, \min(y_i, y_j)) \tag{19}$$

where c is a stand-in for the two different constants $\frac{1}{{D_k}^2}$ or $\frac{1}{D_l D_s}$. Since we are working in 2D Euclidean space, the integral in Eqn. 11 is replaced with the 2D integral in the R_+^2 domain.

$$\begin{aligned} & HC(P_1, P_2, \ldots, P_N) \\ = & -\iint_{R_+^2} P^2(X > \lambda) d\lambda_x d\lambda_y + \sum_k \pi_k \iint_{R_+^2} P_k^2(X_k > \lambda) d\lambda_x d\lambda_y \end{aligned} \tag{20}$$

Since the Heaviside function has a straightforward integral expression, we compute the 2D integral of the expression as

$$\begin{aligned} G(X) &= \int_0^\infty \int_0^\infty g(x,y) dx dy \\ &= \int_0^\infty \int_0^\infty c\sum_{i,j} H(x, \min(x_i, x_j))H(y, \min(y_i, y_j)) dx dy \\ &= c\sum_{i,j} \int H(x, \min(x_i, x_j)) dx \int H(y, \min(y_i, y_j)) dy \\ &= c\sum_{i,j} \min(x_i, x_j) \min(y_i, y_j) \end{aligned} \tag{21}$$

Therefore, the key issue remaining is to derive the analytic gradient for $G(X)$, since the cost function is a linear combination of $G(X)$. We use the chain rule to get

$$\frac{\partial G(X)}{\partial \mu_k} = \frac{\partial G(X)}{\partial X_k} \frac{\partial X_k}{\partial \mu_k} \tag{22}$$

where

$$\frac{\partial G(X)}{\partial X_k} = [\frac{\partial G(X)}{\partial \mathbf{x}_k^1}, \ldots, \frac{\partial G(X)}{\partial \mathbf{x}_k^{D_k}}]$$

$$\frac{\partial G(X)}{\partial \mathbf{x}_k^i} = [\frac{\partial G(X)}{\partial x_k^i}, \frac{\partial G(X)}{\partial y_k^i}]$$

and

$$\frac{\partial G(X)}{\partial x_k^i} = c \sum_{i,j} \min(y_i, y_j) \frac{\partial \min(x_i, x_j)}{\partial x_i}$$

Note that in Eqn. 13, x_i and x_j refer to the x-coordinates of points from the same point-set, whereas in Eqn. 16, they refer to the x-coordinates from different point sets.

Let B_k be the TPS basis matrix computed in advance from the given point-set $\bar{X}_k$. The first three columns of B_k span the affine basis and the remaining columns span the nonlinear warping basis. The transformation therefore can be formed as $X_k = B_k \mu_k$. Hence,

$$\frac{\partial X_k}{\partial \mu_k} = B_k' \tag{23}$$

Finally, we produce the gradient of the objective function as follows,

$$\begin{aligned} \frac{\partial HC}{\partial \mu_k} &= \frac{\partial HC}{\partial X_k} \frac{\partial X_k}{\partial \mu_k} \\ &= [\frac{\partial HC}{\partial \mathbf{x}_k^1}, \frac{\partial HC}{\partial \mathbf{x}_k^2}, \ldots, \frac{\partial HC}{\partial \mathbf{x}_k^{D_k}}] B_k' \end{aligned} \tag{24}$$

where

$$\begin{aligned} \frac{\partial HC}{\partial x_k^i} &= \partial(-\sum \pi_k^2 \int\int P_k^2(X_k > \lambda) dxdy \\ &- \sum_{l \neq s} \pi_l \pi_s \int\int P_l(X_l > \lambda) P_s(X_s > \lambda) dxdy \\ &+ \sum_k \pi_k \int\int P^2(X_k > \lambda) dxdy) / \partial x_k^i \\ &= -\sum \pi_k^2 \frac{1}{{D_k}^2} \sum_{i,j} \min(y_i, y_j) \frac{\partial \min(x_i, x_j)}{\partial x^i} \\ &- \sum_{l \neq s} \pi_l \pi_s \frac{1}{D_l D_s} \sum_{i,j} \min(y_i, y_j) \frac{\partial \min(x_i, x_j)}{\partial x^i} \\ &+ \sum_k \pi_k \frac{1}{{D_k}^2} \sum_{i,j} \min(y_i, y_j) \frac{\partial \min(x_i, x_j)}{\partial x^i} \end{aligned} \tag{25}$$

The expression is also similar with $\frac{\partial HC}{\partial y_k^i}$. The above derivation is directly extensible to higher dimensions as a result of the separability of the Heaviside Function.

We can see that the equation for the objective function Eqn. 21 and gradient Eqn. 24 are simple to implement and computationally fast. With the analytical gradient being explicitly derived, we can use the gradient-based numerical optimization methods such as quasi-Newton [26] to yield a good solution. Meanwhile, and from the overall perspective, robustness is achieved by using

a CDF based objective function. Note that our algorithm can also be applied to yield a biased registration if we fix one of the data sets as the model and estimate the transformation from the scene data sets to the model. In the next section, we will use this "byproduct" - a biased group-wise registration - to propose a series of comparison experiments, by taking the fixed model data set as ground truth.

4 Complexity Analysis and Experimental Results

We first briefly analyze the computational complexity of both CDF-JS and CDF-HC algorithms and show that our model greatly reduces the computational complexity—in terms of CDF estimation—and thus increases the efficiency of the overall approach. Next, we show a demonstrative example of $2D$ atlas construction and illustrate the effect of the regularization parameter η by showing the group-wise registration results for a variety of η. To demonstrate the accuracy and robustness of our CDF based method over the corresponding PDF based approach, a set of comparison experiments were carried out on both synthetic and real 2D data sets. Finally, we show the results of group wise $3D$ registration and a non-rigid group wise registration assessment method is proposed to evaluate the registration without knowing the ground truth.

4.1 Computational Complexity Analysis

Here, we compare the computational complexity for the objective functions of CDF-JS and CDF-HC. Without loss of generality in the mathematical analysis, we assume there exist N point sets with dimension d, each consisting of n points. Also we assume the weights (i.e. π_k in Definition 1 & 4) for all the point sets to be equal to $\frac{1}{N}$. Since both methods use TPS as the non-rigid transformation model, we only compare the complexity of the information theoretic measure part of their cost functions, since the TPS regularization part has the same computational complexity. Taking the $d = 2$ case as an example, we reproduce the cost functions for CDF-JS (Eqn. (7) in [20] but using our notation) and CDF-HC here:

$$\begin{aligned} \mathbf{CDF-JS}: \quad & C(P_1, P_2, \ldots, P_N) \\ = \quad & -\sum_{\lambda_x}^{N_{\lambda_x}} \sum_{\lambda_y}^{N_{\lambda_y}} P \log P + \frac{1}{N} \sum_{k=1}^{N} \sum_{\lambda_x}^{N_{\lambda_x}} \sum_{\lambda_y}^{N_{\lambda_y}} P_k \log P_k \end{aligned} \tag{26}$$

where $P_k = \sum_{i=1}^{n} \Phi(x_i^k)\Phi(y_i^k)$, $P = \frac{1}{N}\sum_{k=1}^{N} P_k$ and $\Phi()$ is the cumulative residual function of the cubic spline kernel used to compute CDF-JS.

$$\begin{aligned} \mathbf{CDF-HC}: \quad & HC(P_1, P_2, \ldots, P_N) \\ = \quad & -\int\int_{R_+^2} P^2 d\lambda_x \lambda_y + \frac{1}{N}\sum_{k=1}^{N} \int\int_{R_+^2} P_k^2 d\lambda_x \lambda_y \end{aligned} \tag{27}$$

where $P = \frac{1}{N}\sum_{k=1}^{N} P_k$ and $\int\int_{R_+^2} P_{k_1} P_{k_2} d\lambda_x \lambda_y = \sum_{i=1}^{n}\sum_{j=1}^{n} min(x_i, x_j) min(y_i, y_j)$.

Since the computational complexity for the functions $\Phi()$, $\min()$ and $\log()$ do not depend on the number of the inputs, it is valid to assume that those functions have a fixed $O(1)$ cost. Therefore, it is clear that the complexity for the CDF-JS algorithm is $O(N_\lambda N_\gamma N n)$ and the complexity for the CDF-HC algorithm is $O(Nn^2)$. Since in the Parzen window estimation, N_λ and N_γ are the number of discrete coordinate values in the x and y axis respectively, we thus have $N_\lambda N_\gamma \gg n$ and, consequently, we conclude that the computational complexity for CDF-JS is asymptotically much larger than that for CDF-HC. (As a brief aside, we would also like to mention that our experience with both algorithms—CDF-JS and CDF-HC—has been overwhelmingly in favor of the latter, but this perspective is driven by our choice of quasi-Newton-based optimization algorithms and the fact that the objective function and gradient are available in closed form for $\alpha = 2$ for CDF-HC.)

4.2 Group-wise Registration for Atlas Construction

In this section, we first show a demonstrative example of our CDF-HC algorithm for *unbiased* 2D atlas construction on a real Corpus Callosum (CC) data set. In this experiment, we manually extracted 63 points on the outer contour of the CC from seven normal subjects. Our algorithm can simultaneously align multiple shapes into a mean shape as shown in Fig. 1. Next, we perturbed the seventh data set and added outliers to it (as shown in the first figure of Fig. 2, denoted by '+'). The registration results of both PDF-JS and CDF-HC are shown in the Fig. 2. We found that the CDF-HC method can better register the outlier shape to the emerging mean shape. In these experiments, the initialization of the non-rigid registration parameters are simply [**I;0**] for all the affine parts and 0 for all the non-rigid ones. All the experiments in this paper were implemented in MATLAB® using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-Newton method of optimization with a mixed quadratic and cubic line search procedure. This quasi-Newton method uses the BFGS formula for updating the approximation of the Hessian matrix. When analytic gradients are used, (and this is the base case for our approach with $\alpha = 2$) cubic line searches

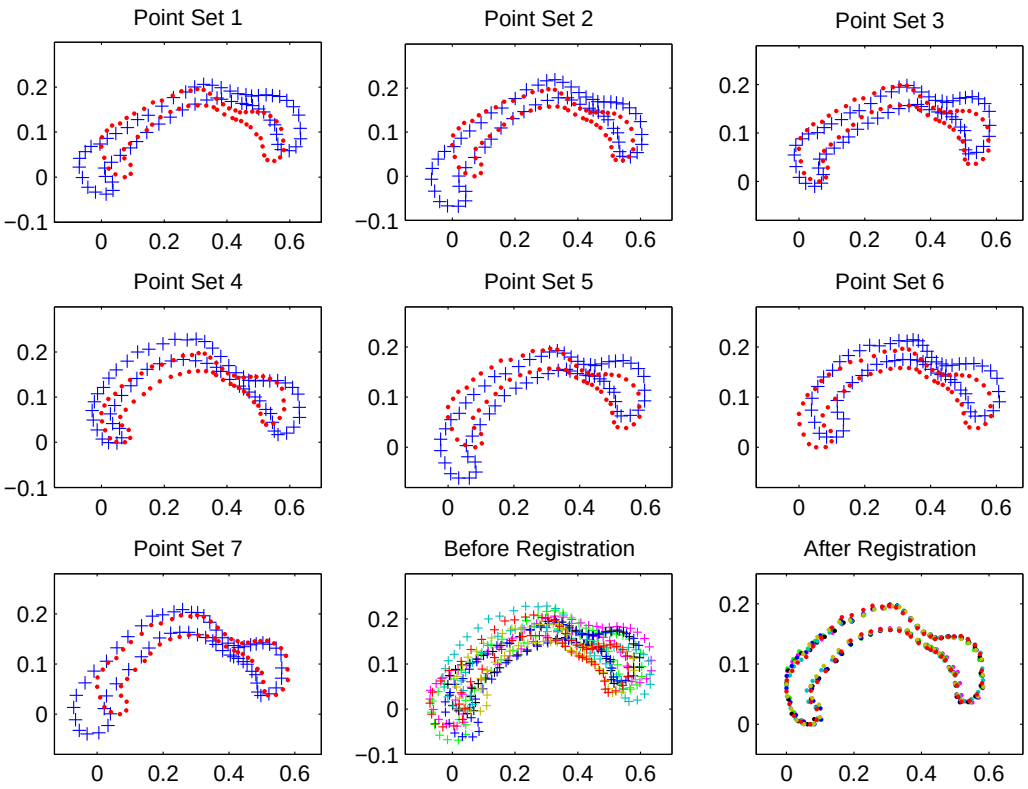

Figure 1: Unbiased group-wise non-rigid registration via CDF-HC on real CC data sets. The first and second rows and the leftmost image in the third row show the deformation of each point set to the atlas. The initial point set is denoted with '+' and the deformed one '·'. The middle image in the third row shows the superimposed point sets before registration. The rightmost image in the third row shows the superimposed point sets after registration.

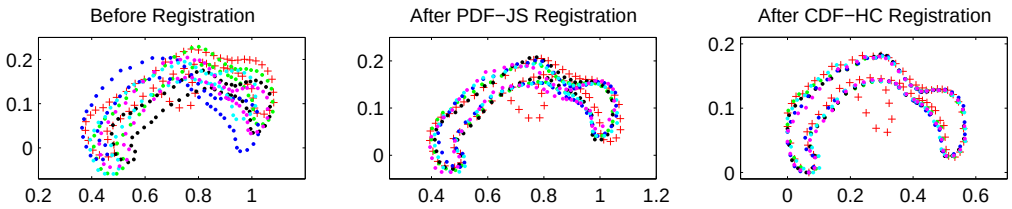

Figure 2: Unbiased group-wise non-rigid registration via PDF-JS and CDF-HC on CC data sets, with the seventh being the outlier. We denote the inliers with '·' and the outlier with '+'.

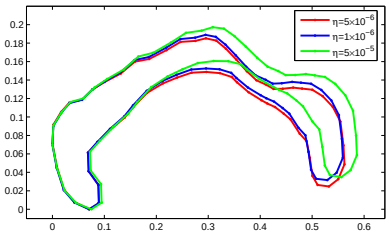

Figure 3: Examples of different regularization parameters of TPS that produce a relatively stable atlas.

are preferred. When numerical approximation of the gradients is used, the quadratic line search method is preferred since it requires fewer gradient evaluations.

Our algorithm requires us to choose a good value of the regularization parameter η. To illustrate the effect of registration via altering η , we finally construct a set of atlases for different η. The same Corpus Callosum data sets were used in this experiment. Fig. 3 shows the atlas with different η. Experimental results indicate that the η that produce relatively stable atlas is in the range of $[0.000005, 0.00005]$—an order of magnitude range.

In the next set of experiments, we examined the (anecdotal) variability w.r.t. α. In these experiments, we used the numerical gradient in the BFGS quasi-Newton method even for $\alpha = 2$ for the sake of a fair comparison. While it is difficult to reach a conclusion, we observed that the optimization process was much longer especially since each result had to be obtained for the best setting of the regularization parameter. The increased difficulty of the optimization—due to the absence of the analytical cost function and gradient—appeared to narrow the range (of regularization parameter values) over which we obtained good registrations. We observed a deterioration of the quality of the registration results as α is increased beyond 3.

To demonstrate the accuracy and robustness to noise of our algorithm over CDF-JS and PDF-JS, we designed the following procedure to perform a *biased* 2D atlas construction on synthetic data sets with and without outliers using both methods. We first manually extract 113 points on the outer contour of the Beijing 2008 Olympic Logo [1], namely point set B, and randomly generate 6 sets of TPS transformation parameters. We then applied these transformations to B to get 6 randomly transformed data sets (as shown in the first figure in Row number 1 of Fig. 5, indicated by $'+'$). The same procedure was applied to a fish shape taken from GatorBait 100[2], namely point-

[1]This dataset is available at http://en.beijing2008.cn/en_index.shtml

[2]This dataset is available under the terms of the GNU General Public License (GPL version 2) at

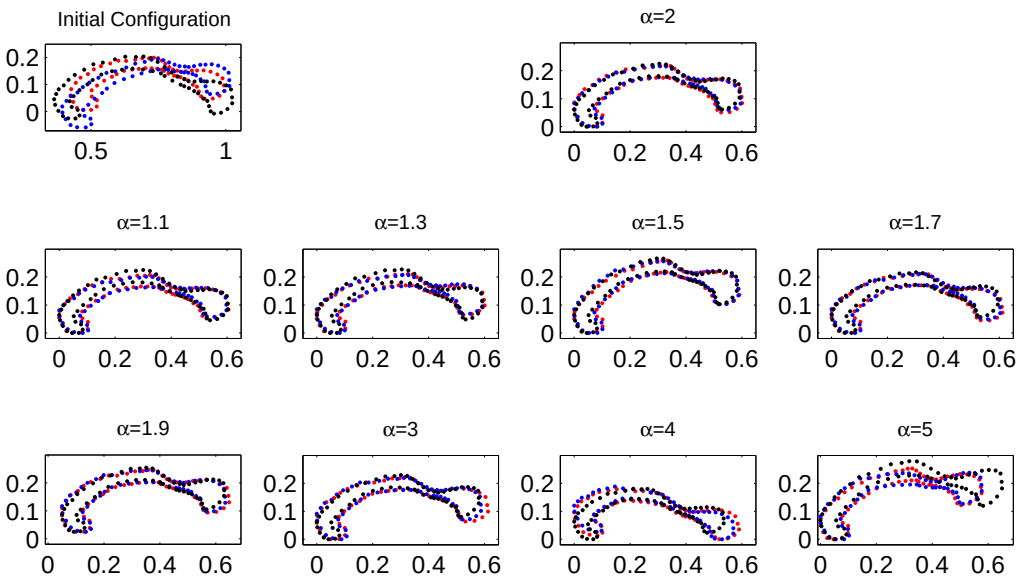

Figure 4: Registration performed using different values of α. In all these experiments, the numerical gradient was used.

set F, so as to get another 6 non-rigid transformed data sets (the first figure in Row number 2 of Fig. 5, indicated by $'+'$), each containing 150 points. We also disturb the data sets with 10 randomly generated point jitter for each of the fish data sets. The CDF-HC, CDF-JS and PDF-JS algorithms are used for group-wise registration of the 6 newly created data sets to the original B or F. For all three methods, we use the same initialization for the optimization parameters, that is we initialize the affine portion with [**I;0**] and 0 for the non-rigid parameters. By implementing the above procedure, we established the ground truth atlases for both Olympic Logo and fish data sets, i.e. B and F, and hence we are able to compare the three methods with the same fixed ground truth.

The Kolmogorov-Smirnov (KS) statistic [27] was computed to measure the difference between the CDFs of the ground truth point-set and the newly registered point sets. A natural question to ask at this juncture is why we are using a different statistic to gauge registration accuracy. The reason is simple. We do not want to use the same registration measure - the HC divergence - to also gauge the registration accuracy. Furthermore, the HC divergence is not an established measure (though we're toiling as hard as we can to change that) whereas the KS statistic is extremely well

http://www.cise.ufl.edu/ãnand/GatorBait_100.tgz

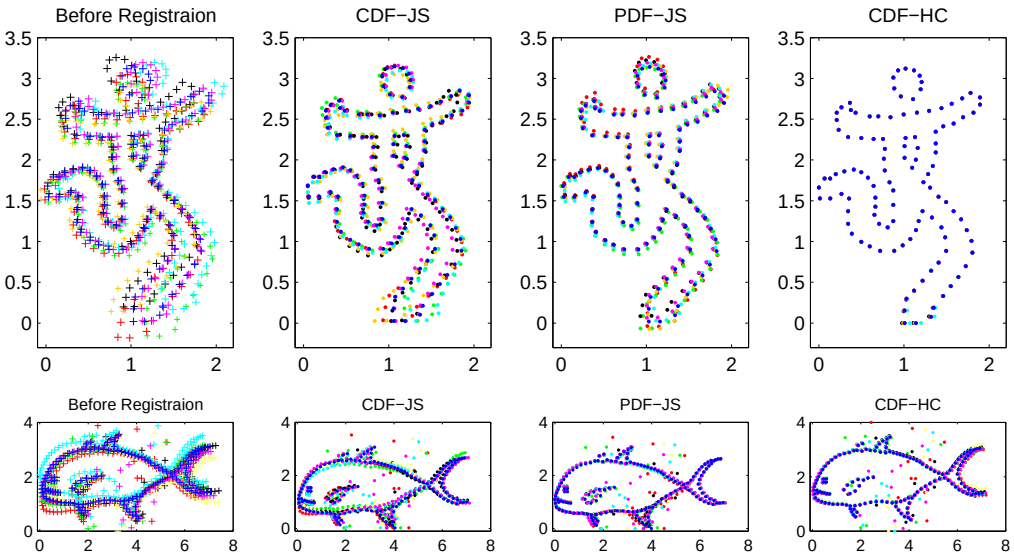

Figure 5: Biased group-wise non-rigid registration via CDF-JS, PDF-JS and CDF-HC on Olympic Logo and 2D fish data sets. The initial point sets are denoted with '+' and the deformed ones '·'. From left to right: The first column shows the superimposed point sets before registration. The second column shows the superimposed point sets after registration using the CDF-JS method, the third shows the same except with PDF-JS and the last column shows the results after CDF-HC registration.

known and widely used (albeit usually for significance testing). Note that here we are not doing statistical significance tests but instead using the KS-statistic as a measure of the dis-similarity between two underlying probability distributions. While the one-dimensional KS-statistic denoted as $D^{(1)}$ is independent of the special form of the distribution this is not true in higher dimensions. A popular $2D$ extension $D^{(2)}$ was proposed by Peacock in [28], wherein he showed that for most cases, Peacock's version of the $2D$ KS statistic was sufficiently distribution-free. Later in [29], Gosset generalized $D^{(2)}$ and proposed a $3D$ KS statistic $D^{(3)}$. Note that in $2D$, there are 3 independent directions (4 directions in total) to perform the cumulative integration and in $3D$ there are a total of 8 such directions. The procedures by Peacock and Gosset were to adopt the largest differences between the two empirical cumulative distribution functions (eCDF), among all the 4 or 8 cumulative directions in 2D and 3D respectively.

In this experiment, we constructed the $2D/3D$ eCDFs from the point-sets following the formulae proposed in [29]. An example of the $2D$ eCDF computed from one of the 4 directions is shown in Fig. 6. We computed the KS-statistic between eCDFs from the ground truth and the registered point sets. Let F_g be the eCDF of the ground truth and F_k be the eCDF estimated from the k_{th} registered point set. The average difference of the eCDF between the ground truth point-set and registered point-sets are evaluated using $\frac{1}{N}\sum_{k=1}^{N} D^{(2)}(F_g, F_k)$. The KS-statistics for both Olympic Logo and fish data sets are presented in Table 1 and the registration results are shown in Fig. 5. They clearly indicate that the CDF-HC method yields a smaller KS-statistic, hence better registration and more immune to noise.

We also present the comparison using the average nearest neighbor distance[3] in Table 2, which also favors the CDF-HC method. For a pair of given point sets, the average nearest neighbor distance is defined by finding the nearest neighbor from the second point set for each point in the first point set and vice versa, and then computing the average distance over all the points. Here, we compute the distance between each point set and the ground truth point set and then take the average.

Table 1: KS statistic

KS-statistic	**CDF-JS**	**PDF-JS**	**CDF-HC**
Olympic Logo	0.1103	0.1018	**0.0324**
Fish	0.1314	0.1267	**0.0722**

[3]We acknowledge an anonymous reviewer for this suggestion.

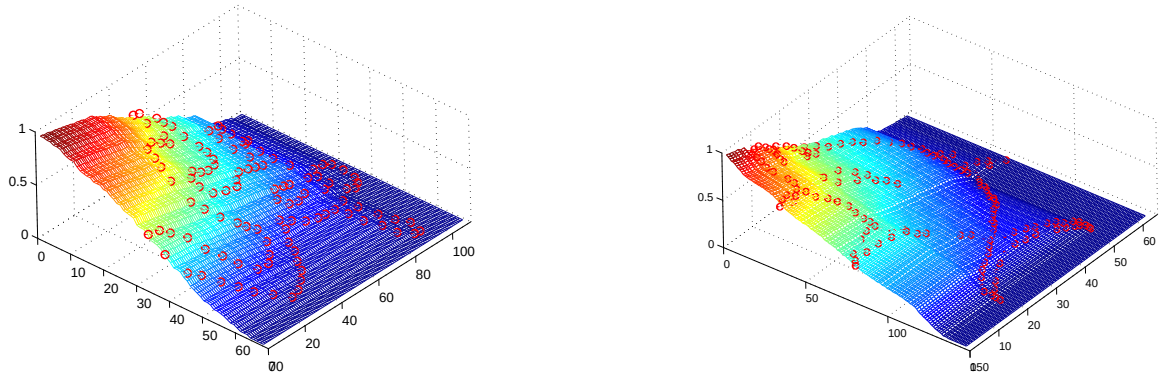

Figure 6: Example meshes of the 2D empirical cumulative distribution function estimated from the Olympic Logo and the fish data used for computing the KS-statistics with the integral performed from one of the four directions.

Table 2: Average Nearest Neighbor Distance

ANN Distance	**CDF-JS**	**PDF-JS**	**CDF-HC**
Olympic Logo	0.0367	0.0307	**0.0019**
Fish	0.0970	0.0610	**0.0446**

Finally, we present the results for 3D atlas construction. The initialization in the optimization here is similar to the previous experiments, i.e. **[I;0]** for affine and 0 for non-rigid parameters. The experiments were carried out on the hippocampus and duck data sets, the latter of which is extracted from a web-based 3D data set[4]. Each data set contains 4 point sets. The *unbiased* group-wise registration results for the hippocampus and duck data sets using the CDF-HC algorithm are shown in Fig. 7 and Fig. 8, respectively. These experiments clearly demonstrate that our point-set registration algorithm can simultaneously register multiple point sets, which can be used to compute a meaningful mean shape/atlas.

4.3 Group-wise Registration Assessment Without Ground Truth

In the previous section, a set of atlas construction experiments for real data were presented. However, there is no standard validation method to evaluate the "goodness" of a computed atlas shape (or in our case an atlas probability distribution). Therefore, we present a group-wise registration

[4] $http://www.3dxtras.com$

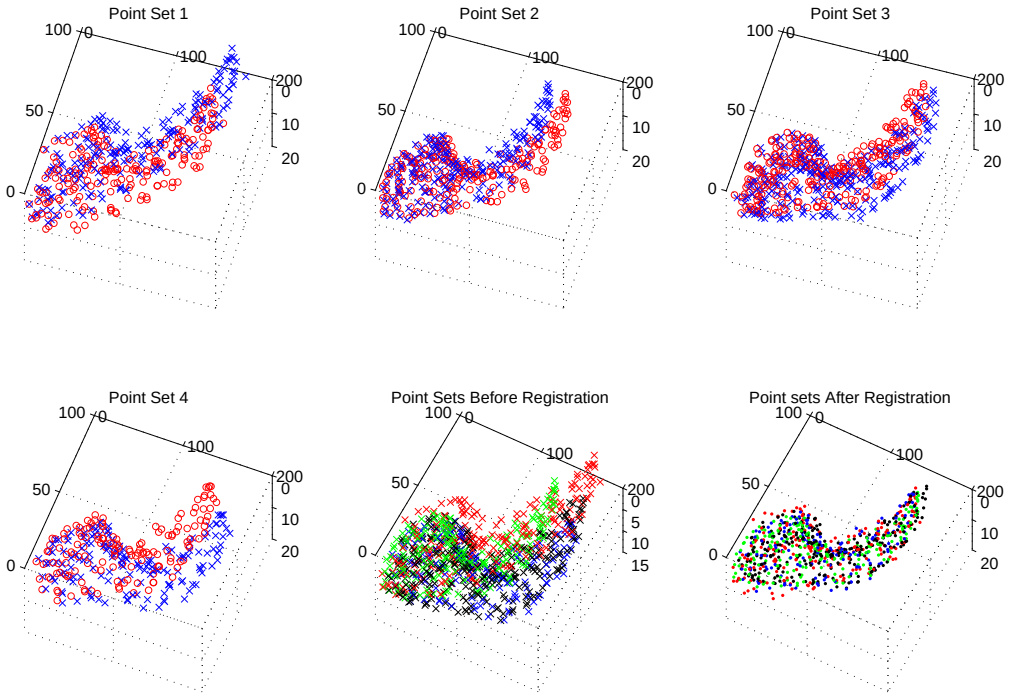

Figure 7: Atlas construction from four 3D hippocampi data sets. Each point set contains $400, 429, 554, 310$ points respectively. The first row and the leftmost image in the second row show the deformation of each point set to the atlas. The initial point set is denoted with $'\times'$ and the deformed one 'o'. The middle image in the second row shows the superimposed point sets before registration. The rightmost image in the second row shows the superimposed point sets after registration.

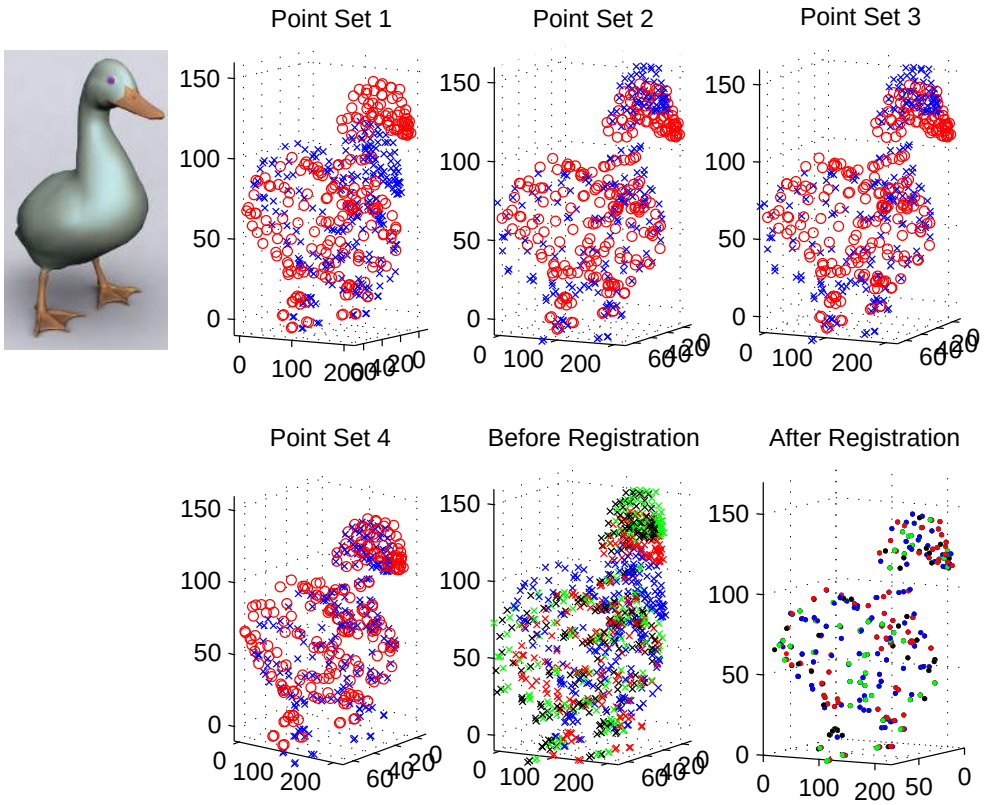

Figure 8: Atlas construction from 3D Duck data sets. Each point set contains 235 points. The first row and the leftmost image in the second row show the deformation of each point set to the atlas. The initial point set is denoted with '×' and the deformed one 'o'. The middle image in the second row shows the superimposed point sets before registration. The rightmost image in the second row shows the superimposed point sets after registration.

assessment criterion, with the ground truth unknown, in order to validate our registration.

Since the KS-statistic is a standard measure of dis-similarity between two CDFs, we decided to generalize the KS-statistic to measure the quality of unbiased group-wise registration. A procedure that is similar to the one presented in previous subsection 4.2 is used for computing the empirical cumulative distribution function in the estimation of the KS-statistic. Intuitively speaking, if the point sets are better registered, the estimated 2D/3D eCDF should be more similar to each other, and hence, we should obtain a smaller KS-statistic between each eCDF estimated from the registered point sets pair-wisely than the eCDF estimated from the initial point set pairs. We use the following measure to evaluate the CDF-HC atlas construction:

$$\mathcal{K} = \frac{1}{N^2} \sum_{k,s=1,k\neq s}^{N} D(F_k, F_s) \tag{28}$$

where F_k is the eCDF from the one point set out of the N point sets. D is $D^{(2)}$ or $D^{(3)}$ in 2D or 3D case respectively, that is, the KS-statistic between two 2D/3D eCDFs. $\mathcal{K}$ thus evaluates the average pairwise KS-statistic among these point sets.

The results for CDF-HC group wise registration assessment using $\mathcal{K}$ for the 2D and 3D *unbiased* atlas construction experiments are listed in Table 3. Our assessment measure $\mathcal{K}$ is computed for each data set before and after registration.

Similarly, the average nearest neighbor distance is also computed here in Table 4 for reference.

Table 3: Non-rigid group-wise registration assessment without ground truth using KS statistics

	Before Registration	**After Registration**
Corpus Callosum	0.3226	**0.0635**
Corpus Callosum with outlier	0.3180	**0.0742**
Olympic Logo	0.1559	**0.0308**
Fish with outlier	0.1102	**0.0544**
Hippocampus	0.2620	**0.0770**
Duck	0.2287	**0.0160**

Obviously, after CDF-HC registration, the point sets achieve a much lower value for both $\mathcal{K}$ as well as the average nearest neighbor distance, as compared to the measures before registration. This indicates that the newly registered point sets more resemble each other than the point sets before registration.

Table 4: Non-rigid group-wise registration assessment without ground truth using average nearest neighbor distance

	Before Registration	**After Registration**
Corpus Callosum	0.0291	**0.0029**
Corpus Callosum with outlier	0.0288	**0.0092**
Olympic Logo	0.0825	**0.0022**
Fish with outlier	0.1461	**0.0601**
Hippocampus	13.7679	**3.1779**
Duck	15.4725	**0.3280**

5 Conclusions

In this paper, we presented a novel and robust algorithm that utilizes an information theoretic measure—the CDF-based Havrda Charvát (CDF-HC) divergence—to simultaneously register multiple unlabeled point-sets with unknown correspondence. Inspired by the separability of the Heaviside Function, we model each point-set using a Dirac Mixture Model so as to employ the Heaviside Function in the CDF surface construction, which greatly simplifies the computation. We also discovered that the cost function has a closed-form solution for its derivative if the parameter for the CDF-HC divergence is set to $\alpha = 2$. This enabled us to reach simple but elegant formulae for both the cost function and its derivatives. The advantages of our algorithm over existing techniques are that CDF-HC can be used for unbiased group-wise point-set registration and it is robust, computationally faster and much simpler from an implementation perspective. We compared the computational complexity of objective function computation for CDF-HC and CDF-JS, showing that CDF-HC is much more efficient. We also compared the performance of CDF-HC, CDF-JS and PDF-JS methods and showed 2D experimental results on a variety of data sets to demonstrate the advantage of correctness and robustness of our CDF-HC algorithm over the corresponding CDF- and PDF-based approaches. Finally, we defined a KS-statistic based measure to evaluate the quality of the group wise registration for real data sets for the case when the ground truth atlas is unknown. Note that we use the TPS as our transformation model since this model is good enough for the deformations of the point sets in this atlas construction problem and we did not observe any local folding. A promising immediate avenue for future research is the incorporation of a diffeomorphism model of deformation [30]. As this effort expands to larger

data sets, we may have to consider simultaneously learning multiple exemplar atlases.

References

[1] Baird, H. S. (1985). *Model-based Image Matching using Location.* Cambridge: MIT Press.

[2] Belongie, S., Malik, J., & Puzicha, J. (2002). Shape Matching and Object Recognition using Shape Contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4), 509-522.

[3] Chui, H., & Rangarajan, A. (2003). A New Point Matching Algorithm for Non-rigid Registration. *Computer Vision and Image Understanding*, 89(2-3), 114-141.

[4] Glaunes, J., Trouvé, A., & Younes, L. (2004). Diffeomorpic Matching of Distributions: A New Approach for Unlabeled Point-Sets and Sub-Manifolds Matching. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, pp. 712-718.

[5] Bookstein, F. L. (1989). Principal Warps: Thin-Plate Splines and the Decomposition of Deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6), 567-585.

[6] Wahba, G. (1990). *Spline Models for Observational Data*. Philadelphia: SIAM.

[7] Rohr, K. (2001). *Landmark-Based Image Analysis: Using Geometric and Intensity Models.* 1st edition. New York: Springer.

[8] Wang, Y., Woods, K., & McClain, M. (2002). Information-Theoretic Matching of Two Point Sets. *IEEE Transactions on Image Procesing*, 11(8), 868-872.

[9] Tsin, Y., & Kanade, T. (2004). A Correlation-based Approach to Robust Point Set Registration. In *Proc. European Conf. on Computer Vision (ECCV)*, Springer LNCS 3024, pp. 558-569.

[10] Jian, B., & Vemuri, B. (2005). A Robust Algorithm for Point Set Registration using Mixture of Gaussians. In *Proc. 10th IEEE Intl. Conf. on Computer Vision (ICCV)*, vol. 2, pp. 1246-1251.

[11] Rohlfing, T., Brandt, R., Maurer, C. R. Jr., & Menzel, R. (2001). Bee Brains, B-splines and Computational Democracy: Generating an Average Shape Atlas, In *Proc. IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA)*, pp. 187-194.

[12] Twining, C. J., Cootes, T. F., Marsland, S., Petrovic, V., Schetowitz, R., & Taylor, C. J. (2006). Information-Theoretic Unification of Groupwise Non-rigid Registration and Model Building. In *Proc. Medical Image Understanding and Analysis (MIUA)*, vol. 2, pp. 226-230.

[13] Lorenzen, P., Prastawa, M., Davis, B., Gerig, G., Bullitt, E., & Joshi, S. (2006). Multi-Modal Image Set Registration and Atlas Formation. *Medical Image Analysis*, 10(3), 440-451.

[14] Sabuncu, M. R., Shenton, M. E., & Golland, P. (2007). Joint Registration and Clustering of Images. In *Proc. MICCAI Workshop on Statistical Registration Workshop: Pair-wise and Group-wise Alignment and Atlas Formation*, 10(WS), pp. 47-54.

[15] Sebastian, T. B., Crisco, J. J., Klein, P. N., & Kimia, B. B. (2000). Constructing 2D Curve Atlases. In *Proc. IEEE Workshop on Mathematical Methods in Biomedical Imaging Analysis (MMBIA)*, pp. 70-77.

[16] Klassen, E., Srivastava, A., Mio, M., & Joshi, S. H. (2004). Analysis of Planar Shapes Using Geodesic Paths on Shape Spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(3), 372-383.

[17] Blake, A., & Isard, M. (1998). *Active Contours: The Application of Techniques from Graphics, Vision, Control Theory and Statistics to Visual Tracking of Shapes in Motion.* New York: Springer-Verlag.

[18] Duta, N., Jain, A. K., & Dubuisson-Jolly, M. -P. (2001). Automatic Construction of 2D Shape Models, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(5), 433-446.

[19] Chui, H., Rangarajan, A., Zhang, J., & Leonard, C. M. (2004). Unsupervised Learning of an Atlas from Unlabeled Point Sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(2), 160-172.

[20] Wang, F., Vemuri, B. C., & Rangarajan, A. (2006). Group-wise Point Pattern Registration Using a Novel CDF-based Jensen-Shannon Divergence. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 1283-1288.

[21] Wang, F., Vemuri, B. C., Rangarajan, A., Schmalfuss, I., & Eisenschek, S. (2008). Simultaneous Nonrigid Registration of Multiple Point Sets and Atlas Construction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(11), 2011-2022.

[22] Wang, F., Vemuri, B. C., Rao, M., & Chen, Y. (2003). A New and Robust Information Theoretic Measure and its Application to Image Alignment. In *Proc. Information Processing in Medical Imaging (IPMI)*, Springer LNCS 2732, pp. 388-400.

[23] Havrda, J., & Charvát, F. (1967). Quantification Method of Classification Processes: Concept of Structural α-Entropy. *Kybernetika*, 3, 30-35.

[24] McLachlan, G. J., & Basford, K. E. (1988). *Mixture Models: Inference and Applications to Clustering*. New York: Marcel Dekker.

[25] Cootes, T., & Taylor, C. (1999). A Mixture Model for Representing Shape Variation. *Image and Vision Computing*, 17(8), 567-573.

[26] Nocedal, J., & Wright, S. J. (2006). *Numerical Optimization (Springer Series on Operations Research and Financial Engineering)*, 2nd edition, New York: Springer.

[27] Feller, W. (1948). On the Kolmogorov-Smirnov Limit Theorems for Empirical Distributions. *The Annals of Mathematical Statistics*, 19(2), 177-189.

[28] Peacock, J. A. (1983). Two-dimensional Goodness-of-Fit Testing in Astronomy. *Royal Astronomical Society Monthly Notices*, 202, 615-627.

[29] Gosset, E. (1987). A Three-dimensional Extended Kolmogorov-Smirnov Test As a Useful Tool in Astronomy. *Astronomy and Astrophysics*, 188(1), 258-264.

[30] Guo, H., Rangarajan, A., & Joshi, S. (2005). 3D Diffeomorphic Shape Registration Using Hippocampal Datasets. In *Proc. Intl. Conf. on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, Springer LNCS 3750, pp. 984-991.