Information Geometry for Landmark Shape Analysis: Unifying Shape Representation and Deformation

Adrian M. Peter [Student Member, IEEE] and **Anand Rangarajan [Member, IEEE]**
The authors are with the Department of Computer and Information Science and Engineering, University of Florida, E301 CSE Building, PO Box 116120, Gainesville, FL 32611-6120

Adrian M. Peter: adrian.peter@gmail.com; Anand Rangarajan: anand@cise.ufl.edu

Abstract

Shape matching plays a prominent role in the comparison of similar structures. We present a unifying framework for shape matching that uses mixture models to couple both the shape representation and deformation. The theoretical foundation is drawn from information geometry wherein information matrices are used to establish intrinsic distances between parametric densities. When a parameterized probability density function is used to represent a landmark-based shape, the modes of deformation are automatically established through the information matrix of the density. We first show that given two shapes parameterized by Gaussian mixture models (GMMs), the well-known Fisher information matrix of the mixture model is also a Riemannian metric (actually, the Fisher-Rao Riemannian metric) and can therefore be used for computing shape geodesics. The Fisher-Rao metric has the advantage of being an intrinsic metric and invariant to reparameterization. The geodesic—computed using this metric—establishes an intrinsic deformation between the shapes, thus unifying both shape representation and deformation. A fundamental drawback of the Fisher-Rao metric is that it is *not* available in closed form for the GMM. Consequently, shape comparisons are computationally very expensive. To address this, we develop a new Riemannian metric based on generalized ϕ-entropy measures. In sharp contrast to the Fisher-Rao metric, the new metric is available in closed form. Geodesic computations using the new metric are considerably more efficient. We validate the performance and discriminative capabilities of these new information geometry-based metrics by pairwise matching of corpus callosum shapes. We also study the deformations of fish shapes that have various topological properties. A comprehensive comparative analysis is also provided using other landmark-based distances, including the Hausdorff distance, the Procrustes metric, landmark-based diffeomorphisms, and the bending energies of the thin-plate (TPS) and Wendland splines.

Index Terms

Information geometry; Fisher information; Fisher-Rao metric; Havrda-Charvát entropy; Gaussian mixture models; shape analysis; shape matching; landmark shapes

1 INTRODUCTION

Shape analysis is a key ingredient to many computer vision and medical imaging applications that seek to study the intimate relationship between the form and function of natural, cultural, medical, and biological structures. In particular, landmark-based

For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.

deformable models have been widely used in quantified studies [1], requiring size and shape similarity comparisons. Shape comparison across subjects and modalities require the computation of similarity measures, which in turn rely upon nonrigid deformation parameterizations. Almost all of the previous work in this area uses separate models for shape representation and deformation. The principal goal of this paper is to show that shape representations beget shape deformation parameterizations [2], [3]. This unexpected unification directly leads to a shape comparison measure.

A brief, crosscutting survey of existing work in shape analysis illustrates several taxonomies and summaries. Shape deformation parameterizations range from Procrustean metrics [4] to spline-based models [5], [6] and from PCA-based modes of deformation [7] to landmark diffeomorphisms [8], [9]. Shape representations range from unstructured point sets [10], [11] to weighted graphs [12] and include curves [13], surfaces [14], and other geometric models. These advances have been instrumental in solidifying the shape analysis landscape. However, one commonality in virtually all of this previous work is the use of separate models for shape representation and deformation. For example, this decoupling between shape representation and deformation is evident in the spline-based planar landmark-matching model:

$$E(f) = \sum_{a=1}^{K} \|\mathbf{v}_a - f(\mathbf{u}_a)\|^2 + \lambda \|Lf\|^2.$$

(1)

Minimizing (1) the results in a nonrigid mapping f that takes landmarks $\mathbf{u}_a$ on to $\mathbf{v}_a$. However, the mapping f did not come about from the landmarks, which are just points in $\mathbb{R}^2$; furthermore, the class of admissible maps is controlled by our choice of the differential operator L. The framework presented in this article directly addresses this issue of decoupling the representation from deformation, yielding a model that enables us to warp landmarks without the use of splines. This is expanded upon in Section 2.

In this paper, we use probabilistic models for shape representation. Specifically, Gaussian mixture models (GMMs) are used to represent unstructured landmarks for a pair of shapes. Since the two density functions are from the same parameterized family of densities, we show how a Riemannian metric arising from their information matrix can be used to construct a geodesic between the shapes. We first discuss the Fisher-Rao metric, which is actually the Fisher information matrix of the GMM. To motivate the use of the Fisher-Rao metric, assume for the moment that a deformation applied to a set of landmarks creates a slightly warped set. The new set of landmarks can also be modeled using another mixture model. In the limit of infinitesimal deformations, the Kullback-Leibler (KL) distance between the two densities is a quadratic form with the Fisher information matrix playing the role of the metric tensor. Using this fact, we can compute a geodesic distance between two mixture models (with the same number of parameters).

A logical question arose out of our investigations with the Fisher information matrix: Must we always choose the Fisher-Rao Riemannian metric when trying to establish distances between parametric probabilistic models? (Remember in this context that the parametric models are used to represent shapes.) The metric's close connections to Shannon entropy and the concomitant use of the Fisher information in parameter estimation have cemented it as the incumbent information measure. It has also been proliferated by research efforts in information geometry, where one can show its proportionality to popular divergence measures such as KL. However, the algebraic form of the Fisher-Rao metric tensor makes it very difficult to use when applied to multiparameter spaces like mixture models. For instance, it is not possible to derive closed-form solutions for the metric tensor or its

derivative. To address many of these computational inefficiencies that arise when using the standard information metric, we introduce a new Riemannian metric based on the generalized notion of a ɸ-entropy functional. We take on the challenge of improving (computationally) the initial Fisher-based model by incorporating the notion of generalized information metrics as first shown by Burbea and Rao [15].

The rich differential geometric connections associated with representing shapes as mixture models enables a flexible shape analysis framework. In this approach, several of the drawbacks often associated with contemporary methods are remedied, i.e., shape matching under this model:

- Provides a unified model for shape representation and deformation—no spline model needed for deforming landmark shapes.

- Does not place constraints on shape topology, i.e., shapes are not required to be simple curves.

- Allows mixture model representations of shapes to be analyzed on the manifold of densities, thus respecting the natural geometry associated with the representation.

- Utilizes a generalized method to develop new information metrics—the new metric we develop has significant computational savings over the Fisher-Rao metric and, for the first time, provides a closed-form metric for parametric Gaussian mixtures.

We begin in Section 1.1 by providing further motivation for our approach and cover a few related methods (the remaining are cited throughout the text). In Section 2, we discuss the probabilistic representation model for landmark shapes. We show how it is possible to go from a landmark representation to one using GMMs. We look at the underlying assumptions and their consequences, which play a vital role in interpreting the analysis. Section 3 illustrates the theory and intuition behind how one directly obtains a deformation model from the representation. It provides a brief summary of the necessary information geometry background needed to understand all subsequent analysis. We illustrate connections between the Fisher information and its use as a Riemannian metric to compute the shortest path between two densities. We then motivate generalizations by discussing Burbea and Rao's work on obtaining differential metrics using the ɸ-entropy functional in parametric probability spaces. The use of a specific ɸ-function α-order entropy first introduced by Havrda and Charvát [16]. This can in turn be utilized to develop a new metric (α-*order entropy metric*) that leads to closed-form solutions for the Christoffel symbols when using GMMs for coupling shape representation and deformation. This enables almost an order-of-magnitude performance increase over the Fisher-Rao-based solution. Section 4 validates the Fisher-Rao and α-order entropy metrics by using them to compute shape distances between corpus callosum data and provides extensive comparative analysis with several other popular landmark-based shape distances.

1.1 Motivation and Related Work

There are a number of advantages when mixture models are used to represent shape landmarks or shape point sets in general. The first is the alleviation of the correspondence problem. Other benefits of the mixture representation include the inherent robustness to noise and localization error of the shape features and landmarks. A shape distance is obtained by computing a distance between probability density functions. Moreover, in a manner that is highly reminiscent to comparing distance transforms of shapes, the probability density functions can be compared at every point in $\mathbb{R}^2$ for two-dimensional shapes. In the literature, we find several instances of using divergence measures [17], [18], [11] and closed-form L_2 distances between mixture models [19] as stand-ins for shape distance measures. In all of these previous approaches, the objective function minimized is a

combination of a distance measure between mixture densities and a spline regularization of the nonrigid warping. The spline driven nonrigid warping attempts to make a source shape mixture density as close as possible to a fixed target shape mixture density. These approaches can be succinctly summarized as minimizing:

$$E(f) = D\left(p\left(\mathbf{x}|\Theta^{(1)}\right), p\left(\mathbf{x}|\Theta^{(2)}(f)\right)\right) + \lambda \|Lf\|^2,$$

(2)

where $\Theta^{(1)}$ is the set of parameters of the first (fixed) shape's mixture model, and $\Theta^{(2)}(f)$ is the set of (warped) parameters of the second shape's mixture model. The choice of spline—a thin-plate spline (TPS) or Wendland spline for example—is determined by the choice of the differential operator L. This set of approaches aims to discover the best nonrigid warping function f (whose spatial smoothness properties are determined by the choice of L) that takes $p(\mathbf{x}|\Theta^{(2)}(f))$ as close as possible to $p(\mathbf{x}|\Theta^{(1)})$. (When a diffeomorphism is sought, the second term is modified to accommodate an infinitesimal generator of a group of transformations.) As previously mentioned, the mixture density distance measure can be a divergence measure like the popular KL (or Jensen-Shannon) measures [20] or a more straightforward closed-form L_2 distance. Moreover, when we examine this notion of shape distance from a wider perspective, these distances are not that different from those obtained using distance transforms [21] or distributions (generalized functions) [22].

Turning our focus to the spline-based regularization term, we observe an interesting disconnect especially from the vantage point of information geometry. In (2), we have the combination of a distance measure D between two mixture density functions and a spline-based regularization term $\|Lf\|^2$. These two terms are independent of each other, and this is reflected in the fact that we can choose any distance measure (KL, L_2, etc.) and any spline (TPS, Wendland, Gaussian radial basis, etc.), resulting in a crossproduct of choices. This decoupling of shape representation (mixture model in this case), and shape deformation is also present in other (nonprobabilistic) landmark diffeomorphism frameworks [9], [8]. For example, in [8], the landmark diffeomorphism objective function takes the form

$$E(\{\Theta(t), f_t\}) = \sum_{a=1}^{K} \int \left\| \frac{d\phi_a}{dt} - f_t(\Theta(t)) \right\|^2 dt + \lambda \int \|L f_t\|^2 dt,$$

(3)

where $f_t(\Theta(t))$ is a velocity field, and $\Theta(t)$ is the set of landmark positions at time t. However, and in a preview of the central idea in this paper, there is a strong relationship between the two terms. The distance measure D gives us a scalar measure of the similarity between two mixture densities, and the regularization operator L forces f to be spatially smooth in order to generate transformations close to identity. From the information geometry perspective, there is a geodesic path from the second shape's probability density $p(\mathbf{x}|\Theta^{(2)})$ to the first shape's probability density $p(\mathbf{x}|\Theta^{(1)})$

Why cannot we unify the two terms—distance measure and spatial smoothness—and directly find the geodesic on a suitably defined probabilistic manifold that gives the shortest possible path between $p(\mathbf{x}|\Theta^{(1)})$ and $p(\mathbf{x}|\Theta^{(2)})$? If this can be achieved, there would be no reason to have two separate terms, one for a shape distance measure and one for a regularization of the nonrigid deformation. Instead, by computing a geodesic between the two probability densities, all we would need to do is move from $p(\mathbf{x}|\Theta^{(1)})$ to $p(\mathbf{x}|\Theta^{(2)})$ on the shortest path connecting the two shapes. This gives the distance measure (length of geodesic) and the warp (intermediate points along the geodesic) *all without the need for a spline-based spatial mapping regularization term*. The distance measure D would be

modified to be a geodesic objective function serving the dual role of shape distance and shape regularization.

2 THE REPRESENTATION MODEL: FROM LANDMARKS TO MIXTURES

In this section we describe the use of probabilistic models, specifically mixture models, for shape representation. Suppose we are given two planar shapes, S_1 and S_2, consisting of K landmarks:

$$S_1=\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_K\}, S_2=\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_K\}, \tag{4}$$

where $\mathbf{u}_a=\left[u_a^1, u_a^2\right]^T, \mathbf{v}_a=\left[v_a^1, v_a^2\right]^T \in \mathbb{R}^2, \forall a \in \{1, \ldots, K\}$. Typical shape-matching representation models consider the landmarks as a collection of points in $\mathbb{R}^2$ or as a vector in $\mathbb{R}^{2K}$. A consequence with these representations is that if one wishes to perform deformation analysis between the shapes, a separate model needs to be imposed, e.g., TPS [23] or landmark diffeomorphisms [8], to establish a map from one shape to the other. (In landmark matching, the correspondence between the shapes is assumed to be known.) In Section 3, we show how the probabilistic shape representation we present in the current section provides an *intrinsic* warping between the shapes—thus unifying both shape representation and deformation.

Mixture model representations have been used to solve a variety of shape analysis problems, e.g., [18] and [24]. We select the most frequently used mixture model to represent our shapes by using a K-component GMM, where the shape landmarks are the centers (i.e., the ath landmark position serves as the ath mean for a specific bivariate component of the GMM). This parametric, GMM representation for the shapes is given by [25]

$$p(\mathbf{x}|\mathbf{\Theta})=\frac{1}{2\pi\sigma^2 K}\sum_{a=1}^{K}\exp\left\{-\frac{\|\mathbf{x} - \phi_a\|^2}{2\sigma^2}\right\}, \tag{5}$$

where $\mathbf{\Theta}$ is the set consisting of all landmarks,

$$\phi_a=[\theta^{(2a-1)}, \theta^{(2a)}]^T, \mathbf{x}=[x^{(1)}, x^{(2)}]^T \in \mathbb{R}^2$$

and equal weight priors are assigned to all components, i.e., $\frac{1}{K}$. (Note: the planar landmarks $\mathbf{u}_a$ or $\mathbf{v}_a$ are mapped to the corresponding GMM component mean ϕ_a.) Though we only discuss planar shapes, it is mathematically straightforward to extend to 3D. In addition, the number of landmarks can be selected either manually or through the use of model selection [26], depending on the application.

The variance σ^2 can capture uncertainties that arise in landmark placement and/or natural variability across a population of shapes. Incorporating full component-wise elliptical covariance matrices provides the flexibility to model structurally complicated shapes. The equal weighting on the componentwise priors is acceptable in the absence of any a priori knowledge. Fig. 1 illustrates this representation model for three different values of σ^2. The input shapes consists of 63 landmarks drawn by an expert from MRI images of the corpus callosum and 233 landmarks manually extracted from an image of a fish. The variance is a free parameter in our shape-matching algorithm, and in practice, it is selected to control the

size of the neighborhood of influence for nearby landmarks. As evident in the figure, another interpretation is that larger variances blur locations of high curvature present in the corpus callosum and fish shapes. Thus, depending on the application, we can dial-in the sensitivities to different types of local deformations. Even though it may seem that as σ^2 increases we lose detailed resemblance to the original shape, it is still valid to compare two shapes with large variance since their representation as mixtures is still unique with respect to the locations of the GMM components. In addition, recall that the variance allows us to handle errors in the landmark locations. Due to these desirable properties, the choice of the variance is currently a free parameter in our algorithm and is isotropic across all components of the GMM. So far, we have only focused on the use of GMMs for landmarks. However, they are also well suited for dense point cloud representation of shapes. In such applications, the mean and covariance matrix can be directly estimated from the data via standard parameter estimation techniques.

The real advantage in representing a shape using a parametric density is that it allows us to perform rich geometric analysis on the density's parameter space. Section 3 covers how this interpretation in the theoretical setting of information geometry allows us to use the same representation model to deform shapes.

3 THE DEFORMATION MODEL: RIEMANNIAN METRICS FROM INFORMATION MATRICES OF MIXTURES

We now address the issue of how the same landmark shape representation given by (5) can also be used to enable the computation of deformations between shapes. The overarching idea will be to use the parametric model to calculate the information matrix, which is a Riemannian metric on the parameter space of densities. If any two shapes are represented using the same family of parametric densities, the metric tensor will allow us to take a "walk" between them. Section 3.1 expands on our use of the terminology intrinsic and extrinsic to describe the analysis under our probabilistic framework. We then use the Fisher-Rao metric to motivate some key ideas from the information geometry used in subsequent parts of the paper. Immediately following, we discuss how to apply the popular Fisher-Rao metric to shape matching and develop the fully intrinsic deformation framework. Next, we show how it is possible to derive other information matrices starting from the notion of a generalized entropy. Section 3.5 puts forth a possible solution on how movement of landmarks on the intrinsic space can be used to drive the extrinsic space deformation, a necessity for applying these methods to applications such as point-set shape registration.

3.1 Intrinsic versus Extrinsic Analysis

In the context of using mixture models to represent and deform shapes, we will often use the words *intrinsic* and *extrinsic*. These terms are analogous to their use in differential geometry, where intrinsic describes analysis strictly derived from the surface properties of the manifold and extrinsic refers to the use of the space ambient to the manifold. In the present framework, the K landmarks of a single shape correspond to the centers of a K-component GMM, which in turn give the coordinates of a single point on the manifold of mixture densities. Similarly, another shape with K landmarks will also have the same interpretation as a point on the manifold. Thus, our technique, as described in the next section, enables you to directly use this representation to obtain a warp from one shape onto the other shape without requiring one to arbitrarily introduce a deformable model such as a spline. Since we always stay on the manifold and use the intrinsic property of the metric tensor to obtain our path between densities, which is also the warp between shapes, we refer to this as intrinsic analysis.

Our reference to warping of the extrinsic space arises from the fact that often shape data are realized as point sets, not just landmarks. For a pair of point-set shapes, landmarks can be extracted by a variety of methods such as manual assignment or clustering. Once we have landmarks representations of the shapes, an intrinsic warp can be established, as described above. However, this warp only describes the movement of the landmarks from one shape onto the other. How does one move the shape points, i.e. the extrinsic space consisting of *points surrounding the landmarks*, based on the movement of the landmarks? Though not the focal point of this paper, for completeness we provide one possible solution in Section 3.5. To warp these extrinsic points it will be necessary to introduce an external regularizer but the formulation is commensurate with our theme, using the GMM to drive the warping. Fig. 2 illustrates a fish shape consisting of several thousand points (light gray) from which we have extracted 233 landmarks (black points)—the extrinsic points are surrounding the landmarks, while the landmarks are used as the intrinsic coordinates. Echoing our claim: for *landmark* matching, our framework is completely intrinsic, providing a path (consequently a warp) from one landmark shape onto another without the need of a spline regularizer. Only if the application dictates the need to warp the extrinsic space do we employ the use of a spline model, and even then, the warps are still driven by movement along the intrinsic path determined by the intermediate landmark shapes.

3.2 Backgrounder on Information Geometry

It was Rao [27] who first established that the Fisher information matrix satisfies the properties of a metric on a Riemannian manifold. This is the reasoning behind our nomenclature of *Fisher-Rao* metric whenever the Fisher information matrix is used in this geometric manner. The Fisher information matrix arises from multiparameter densities, where the (i, j) entry of the matrix is given by

$$g_{ij}(\theta) = \int p(\mathbf{x}|\theta) \frac{\partial}{\partial \theta^i} \log p(\mathbf{x}|\theta) \frac{\partial}{\partial \theta^j} \log p(\mathbf{x}|\theta) d\mathbf{x}.$$

(6)

The Fisher-Rao metric tensor (6) is an intrinsic measure, allowing us to analyze a finite n-dimensional statistical manifold M without considering how M sits in an $\mathbb{R}^{2n+1}$ space [28]. In this parametric statistical manifold, $p \in M$ is a probability density with its local coordinates defined by the model parameters. For example, a bivariate Gaussian density can be represented as a single point on a four-dimensional manifold with coordinates $\theta = (\mu^{(1)}, \mu^{(2)}, \sigma^{(1)}, \sigma^{(2)})^T$, where as usual, these represent the mean and standard deviation of the density. (The superscript labeling of coordinates is used to be consistent with differential geometry references.) For the present interest in landmark matching, $\dim(M) = 2K$ because we only use the means of a GMM as the manifold coordinates for a K landmark shape. (Recall that σ is a free parameter in the analysis).

The exploitation of the Fisher-Rao metric on statistical manifolds is part of the overarching theory of information geometry [29], [30]. It can be shown that many of the common metrics on probability densities (e.g., KL, Jensen-Shannon, etc.) can be written in terms of the Fisher-Rao metric given that the densities are close [30]. For example, the KL divergence between two parametric densities with parameters θ and $\theta + \delta\theta$, respectively, is proportional to

$$D(p(x|\theta+\delta\theta)\|p(x|\theta)) \approx \frac{1}{2}(\delta\theta)^T g \delta\theta.$$

(7)

In other words, the KL divergence is equal to, within a constant, a quadratic form with the Fisher information matrix g playing the role of the Hessian. The use of the information matrix to measure the distance between distributions has popularized its use in several applications in computer vision and machine learning. In [31], the authors have used it to provide a more intuitive geometric explanation of model selection criteria such as the minimum description length (MDL) criterion. To our knowledge, there are only few other recent uses of the Fisher-Rao metric for computer vision related analyses. Maybank [32] utilizes Fisher information to analyze projective transformations of the line. Mio et al. [33] apply nonparametric Fisher-Rao metrics for image segmentation. Lenglet et al. [34] successfully demonstrated the use of the Fisher-Rao metric on multivariate normal densities in the analysis of diffusion tensor imaging data. Srivastava et al. [35] have studied applications of the nonparametric Fisher-Rao metric to curve-based shape classification. In their nonparameteric framework, they have cleverly used the $\sqrt{p}$ representation, which enables all analyses (geodesics, means, etc.) to take place on the unit hypersphere. Finally, in [46], we have developed a shape matching technique using the parametric representation of $\sqrt{p}$, where $\sqrt{p}$ is expanded in an orthogonal wavelet basis, leading to a similar unit hypersphere geometry as the nonparametric model.

Information geometry incorporates several other differential geometry concepts in the setting of probability distributions and densities. Besides having a metric, we also require the construct of connections to move from one tangent space to another. The connections are facilitated by computing Christoffel symbols of the first kind, $\Gamma_{k,ij} \overset{\text{def}}{=} \frac{1}{2} \left\{ \frac{\partial g_{ik}}{\partial \theta^j} + \frac{\partial g_{kj}}{\partial \theta^i} - \frac{\partial g_{ij}}{\partial \theta^k} \right\}$, which rely on the partial derivatives of the metric tensor. It is also possible to compute Christoffel symbols of the second kind, which involve the inverse of the metric tensor. Since all analysis is intrinsic, i.e., on the surface of the manifold, finding the shortest distance between points on the manifold amounts to finding a geodesic between them. Recall that in the context of shape matching, points on the manifold are parametric densities which in turn represent landmark shapes. Fig. 3 illustrates this overall idea. The two shapes are represented using mixture models, the parameters of which map to points on the manifold. The goal is to use the metric tensor to find a geodesic between them. Walking along the geodesic will give us intermediate landmark shapes, and the geodesic length will give us an intrinsic shape distance.

3.3 Fisher-Rao Metric for Intrinsic Shape Matching

To discover the desired geodesic between two GMM represented landmark shapes (4), we can use the Fisher-Rao metric (6) to formulate an energy between them as

$$s = \int_0^1 g_{ij} \dot{\theta}^i \dot{\theta}^j dt,$$

(8)

where the standard Einstein summation convention (where summation symbols are dropped) is assumed, and $\dot{\theta}^i = \frac{d\theta^i}{dt}$ is the parameter time derivative. Technically, (8) integrates the square of the infinitesimal length element but has the same minimizer as $\int_0^1 \sqrt{g_{ij} \dot{\theta}^i \dot{\theta}^j} dt$ [36] (which is the true geodesic distance). Note that we have introduced a geodesic curve parameter t, where $t \in [0, 1]$. The geodesic path is denoted as $\theta(t)$, and at $t = 0$ and $t = 1$, we have the end points of our path on the manifold, for instance

$$
\theta(0) \overset{\text{def}}{=}
\begin{bmatrix}
\theta^{(1)}(0) \\
\theta^{(2)}(0) \\
\theta^{(3)}(0) \\
\theta^{(4)}(0) \\
\vdots \\
\theta^{(2K-1)}(0) \\
\theta^{(2K)}(0)
\end{bmatrix}
=
\begin{bmatrix}
u_1^{(1)} \\
u_1^{(2)} \\
u_2^{(1)} \\
u_2^{(2)} \\
\vdots \\
u_K^{(1)} \\
u_K^{(2)}
\end{bmatrix}.
\tag{9}
$$

θ(1) is defined similarly, and as shown, they represent the landmarks of the reference and target shape, respectively. The functional (8) is minimized using the standard calculus of variations techniques, leading to the following Euler-Lagrange equations:

$$
\frac{\delta \mathscr{E}}{\delta \theta^k} = -2g_{ki}\ddot{\theta}^i + \left\{ \frac{\partial g_{ij}}{\partial \theta^k} - \frac{\partial g_{ik}}{\partial \theta^j} - \frac{\partial g_{kj}}{\partial \theta^i} \right\} \dot{\theta}^i \dot{\theta}^j = 0.
\tag{10}
$$

This can be rewritten in the more standard form

$$
g_{ki}\ddot{\theta}^i + \Gamma_{k,ij}\dot{\theta}^i\dot{\theta}^j = 0.
\tag{11}
$$

This is a system of second-order ODEs and not analytically solvable when using GMMs. One can use a gradient descent to find a local solution to the system with update equations:

$$
\theta_{\tau+1}^k(t) = \theta_\tau^k(t) - \alpha_{(\tau+1)}\frac{\delta E}{\delta \theta_\tau^k(t)}, \forall t,
\tag{12}
$$

where τ represents the iteration step and α the step size. It is worth noting that one can apply other optimization techniques to minimize (8). To this end, in [37], the authors have proposed an elegant technique based on numerical approximations and local eigenvalue analysis of the metric tensor. Their proposed method works well for shapes with a small number of landmarks, but the speed of convergence can degrade considerably when the cardinality of the landmarks is large. This is due to the requirement of repeatedly computing eigenvalues of large matrices. Alternate methods, e.g., quasi-Newton algorithms, can provide accelerated convergence while avoiding expensive matrix manipulations. In the next section, we investigate a general class of information matrices that also satisfy the property of being Riemannian metrics. Thus, the analysis presented above to find the geodesic between two shapes holds and simply requires replacing the Fisher-Rao metric tensor by the new $g_{i,j}$.

3.4 Beyond Fisher-Rao: φ-Entropy and α-Order Entropy Metrics

Rao's seminal work and the Fisher information matrix's relationship to the Shannon entropy have entrenched it as the metric tensor of choice when trying to establish a distance metric between two parametric models. However, Burbea and Rao went on to show that the notion of distances between parametric models can be extended to a large class of generalized metrics [15]. They defined the generalized φ-*entropy functional* as

$$H_\phi(p) = -\int_\chi \phi(p)\, d\mathbf{x},$$

(13)

where χ is the measurable space (for our purposes $\mathbb{R}^2$), and ϕ is a C^2-convex function defined on $\mathbb{R}_+ \equiv [0, \infty)$. (For readability, we will regularly replace $p(\mathbf{x}|\theta)$ with p.) The metric on the parameter space is obtained by finding the Hessian of (13) along a direction in its tangent space. The directional derivative of (13) in the direction of v is given by

$$
\begin{aligned}
D_v H_\phi &= \frac{d}{dt} H_\phi(p+tv)|_{t=0}, t \in \mathbb{R} \\
&= -\int \phi'(p) v\, d\mathbf{x},
\end{aligned}
$$

(14)

which we differentiate once more to get the Hessian:

$$D_v^2 H_\phi = -\int \phi''(p) v^2\, d\mathbf{x}.$$

(15)

Assuming sufficient regularity properties on $\theta = \{\theta^1,\ldots,\theta^n\}$, the direction in the tangent space of this parameter set can be obtained by taking the total differential of $p(\mathbf{x}|\theta)$ with respect to θ:

$$dp(\theta) = \sum_{k=1}^n \frac{\partial p}{\partial \theta^k} d\theta^k.$$

(16)

This results in the Hessian being defined as

$$\Delta_\theta H_\phi(p) = -\int_\chi \phi''(p) [\, dp(\theta)]^2 d\mathbf{x},$$

(17)

where we have replaced v with dp, which directly leads to the following differential metric satisfying Riemannian metric properties:

$$ds_\phi^2(\theta) = -\Delta_\theta H_\phi(p) = \sum_{i,j=1}^n g_{i,j}^\phi d\theta^i d\theta^j,$$

(18)

where

$$g_{i,j}^\phi = \int_\chi \phi''(p) \left(\frac{\partial p}{\partial \theta^i}\right) \left(\frac{\partial p}{\partial \theta^j}\right) d\mathbf{x}.$$

(19)

(We refer the reader to [15] for more detailed derivations of the above equations.) The metric tensor in (19) is called the ϕ-*entropy matrix*. By letting

$$\phi(p) = p \log p,$$

(20)

(13) becomes the familiar Shannon entropy, and (19) yields the Fisher information matrix. One major drawback of using the Fisher-Rao metric is that the computation of geodesics is very inefficient as they require numerical calculation of the integral in (6).

We now discuss an alternative choice of ϕ that directly leads to a new Riemannian metric and enables us to derive closed-form solutions for (19). Our desire to find a computationally efficient information metric was motivated by noticing that if the integral of the metric could be reduced to just a correlation between the partials of the density with respect to θ^i and θ^j,

i.e., $\int \frac{\partial p}{\partial \theta^i} \frac{\partial p}{\partial \theta^j} d\mathbf{x}$, then the GMM would reduce to separable one-dimensional Gaussian integrals for which the closed-from solution exists. In the framework of generalized ϕ-entropies, this idea translated to selecting a ϕ such that ϕ'' becomes a constant in (19). In [16], Havrda and Charvát introduced the notion of a α-order entropy using the convex function:

$$\phi(p) = (\alpha - 1)^{-1}(p^\alpha - p), \alpha \neq 1. \tag{21}$$

As $\lim_{\alpha \to 1} \phi(p)$, (21) tends to (20). To obtain our desired form, we set $\alpha = 2$, which results in $\frac{1}{2}\phi'' = 1$. (The 1/2 scaling factor does not impact the metric properties.) Thus, the new metric is defined as

$$g_{i,j}^\alpha = \int_\chi \left(\frac{\partial p}{\partial \theta^i}\right)\left(\frac{\partial p}{\partial \theta^j}\right) d\mathbf{x}, \tag{22}$$

and we refer to it as the α-*order entropy metric tensor*. The reader is referred to the Appendix in [3], where we provide some closed-form solutions to the α-order entropy metric tensor and the necessary derivative calculations needed to compute (22). Though we were computationally motivated in deriving this metric, it will be shown via experimental results that it has shape discriminability properties similar to that of the Fisher-Rao and other shape distances. Deriving the new metric also opens the door for further research into applications of the metric to other engineering solutions. Under this generalized framework, there are opportunities to discover other application-specific information matrices that retain Riemannian metric properties.

3.5 Extrinsic Deformation

The previous sections illustrated the derivations of the probabilistic Riemannian metrics, which led to a completely intrinsic model for establishing the geodesic between two landmark shapes on a statistical manifold. Once the geodesic has been found, traversing this path yields a new set of θ's at each discretized location of t, which in turn represents an intermediate intrinsically deformed landmark shape. We would also like to use the results of our intrinsic model to go back and warp the extrinsic space.

Notice that the intrinsic deformation of the landmarks only required our θ's to be parameterized by time. Deformation of the ambient space $\mathbf{x} \in \mathbb{R}^2$, i.e., our shape points, can be accomplished via a straightforward incorporation of the time parameter on to our extrinsic space,

$$p(\mathbf{x}(t)|\theta(t)) = \frac{1}{K}\sum_{a=1}^{K}\frac{1}{2\pi\sigma^2}\exp\left\{-\frac{1}{2\sigma^2}\|\mathbf{x}(t) - \phi_a(t)\|^2\right\}. \tag{23}$$

We want to deform the $\mathbf{x}(t)$'s of extrinsic space through the velocities induced by the intrinsic geodesic and simultaneously preserve the likelihood,

$$p(\mathbf{x}(t)|\theta(t)) = p(\mathbf{x}(t+\delta t)|\theta(t+\delta t)),$$

of all these ambient points relative to our intrinsic θs. Instead of enforcing this condition on $L = p(\mathbf{x}(t)|\theta(t))$, we use the negative log-likelihood $-\log L$ of the mixture and set the total derivative with respect to the time parameter to zero:

$$\frac{d\log L}{dt} = (\nabla_{\theta^1}\log L)^T \dot{\theta}^1 + (\nabla_{\theta^2}\log L)^T \dot{\theta}^2 + \frac{\partial\log L}{\partial x^1(t)} u + \frac{\partial\log L}{\partial x^2(t)} \upsilon = 0, \tag{24}$$

where $u(t) = \dfrac{dx^1}{dt}$, and $\upsilon(t) = \dfrac{dx^2}{dt}$ represent the probabilistic flow field induced by our parametric model. The notation ∇_{θ^1} is used to reflect the partial derivative with respect to the first coordinate location of each of the K components of the mixture density, and similarly, ∇_{θ^2} are the partials with respect to the second coordinate for each of the K components. Note that this formulation is analogous to the one we find in optical flow problems [38]. Similar to optical flow, we introduce a TPS regularizer to smooth the flow field:

$$\int \left[(\nabla^2 u)^2 + (\nabla^2 \upsilon)^2 \right] d\mathbf{x}. \tag{25}$$

We note that is also possible to use the quadratic variation instead of the Laplacian as the regularizer. On the interior of the grid, both of these satisfy the same biharmonic, but the quadratic variation yields smoother flows near the boundaries.

The overall extrinsic space deformation can be modeled using the following energy functional:

$$E(u,\upsilon) = \int \left(\lambda \left[(\nabla^2 u)^2 + (\nabla^2 \upsilon)^2 \right] + \left[\frac{d\log L}{dt} \right]^2 \right) d\mathbf{x}, \tag{26}$$

where λ is a regularization parameter that weighs the error in the extrinsic motion relative to the departure from smoothness. The minimal flow fields are obtained via the Euler-Lagrange equation of (26). As formulated, the mapping found through the thin-plate regularizer is not guaranteed to be diffeomorphic. This can be enforced if necessary and is currently under investigation for future work. In this section, we have shown that selecting the representation model (23) immediately gave the likelihood preserving data term used to drive the warping of extrinsic shape points, thus continuing our theme of unified shape representation and deformation.

4 EXPERIMENTAL RESULTS AND ANALYSIS

Even though we cannot visualize the abstract statistical manifold on which we impose our two metrics, we have found it helpful to study the resulting geodesics of basic transformations on simple shapes (Figs. 4 and 6). In all figures, the dashed straight line represents the initialization path, and the solid bell-shaped curve shows the final geodesic

between shapes. Fig. 4 shows a straight-line shape consisting of 21 landmarks that has been slightly collapsed like a hinge. Notice that the resulting geodesic is bent, indicating the curved nature of the statistical manifolds. Even though the bending in Fig. 4b is not as visually obvious, a closer look at the landmark trajectories for a couple of the shape's landmarks (Fig. 5) illustrates how the intermediate landmark positions have repositioned themselves from their uniform initialization. It is the velocity field resulting from these intermediate landmarks that enables a smooth mapping from one shape to another [11]. Fig. 6 illustrates geodesics obtained from matching a four-landmark square to one that has been rotated 210 degrees clockwise. The geodesics obtained by the Fisher-Rao metric are again smoothly curved, illustrating the hyperbolic nature of the manifold with this specified information matrix [39], whereas the α-order entropy metric displays sharper, abrupt variations. In both cases, we obtained well-behaved geodesics with curved geometry.

As we have noted, one of the strengths of this framework is that it does not topologically constrain the shapes, allowing us to obtain warps and similarity measures between shapes that exhibit features such as interior structures and disconnected components. To showcase this desirable feature, we matched six fish shapes [40] shown in Fig. 7. For each vertical pair of fish, we extracted equal number of landmarks. The landmark locations for each fish serve as the means of the Gaussian mixture. Since each fish has now been converted to its mixture density representation, we can apply our framework to find geodesics between the pairs. Once the geodesic is found, we can obtain the warp that takes one shape onto another by taking intermediate points (each of which is a valid mixture density) along the geodesic. We are able to accomplish this *without the use of a spline model* because the shapes, under the density representation, are on the manifold of mixture densities; obtaining intermediate shapes amounts to treating the mean components of the intermediate mixtures as the landmarks of the shapes. In Fig. 8, we show eight intermediate shapes for each matching pair from that in Fig. 7. The geodesics were computed with the α-order entropy metric. We compare these deformations to ones produced using the landmark diffeomorphism technique [8]. This is a fairly recent technique with the metric arising from the minimum energy of fitting iterated splines to the infinitesimal velocity vectors that diffeomorphically take one shape onto the other. It is worth noting that in [8], the authors implemented a discrete approximation to their proposed energy functional. In order to avoid any numerical approximation issues and experimental variability, our implementation obtains a gradient descent solution directly on the analytic Euler-Lagrange equations for their functional. Notice that the intermediate deformations, in comparison to our method, are very similar; however, the key differentiator is that landmark diffeomorphisms (Fig. 9) require the use of splines to obtain these intermediate warps, *whereas our method does not*. (Note that we selected the λ parameter in landmark diffeomorphisms such that it would yield intermediate deformations similar to the ones obtained with our method for a particular value of σ. For both methods, varying their respective free parameters can yield different intermediate deformations.)

For applications in medical imaging, we have evaluated both the Fisher-Rao and α-order entropy metrics on real data consisting of nine corpora callosa with 63 landmarks each, as shown in Fig. 10. These landmarks were acquired via manual marking by an expert from different MRI scans. As with all landmark-matching algorithms, correspondence between shapes is known. We performed pairwise matching of all shapes in order to study the discriminating capabilities of the metrics. Since both the Fisher-Rao and α-order entropy metric are obtained from GMMs, we tested both metrics with three different values of the free parameter σ^2. In addition to the two proposed metrics, we performed a comparative analysis with several other standard landmark distances and similarity measures. The distance metrics included are Procrustes [4], [41], symmetrized Hausdorff [42], and landmark diffeomorphisms [8]. The first two distance metrics have established themselves

as a staple for shape comparison, while the third is more recent and was used in the previous discussion for deformation analysis. The shape similarity measures (which are not metrics) incorporated in the study use the bending energy of spline-based models to map the source landmarks to the target. We used two spline models: the ubiquitous TPS [5], which has basis functions of infinite support and the more recently introduced Wendland spline [43], which has compactly supported bases. For the sake for brevity, we will refer to all measures as metrics or distances with the understanding that the bending energies do not satisfy the required properties of a true metric.

The results of pairwise matching of all nine shapes is listed in Table 1, containing the actual pairwise distances. The distances show a global trend among all of the metrics. For example, shape 1 and 8 have the smallest distance under all the metrics except α-order entropy metric with $\sigma^2 = 0.1$ and the TPS bending energy. However, shape 8 is the second best match under both these, clearly illustrating a similar performance to the others. In addition, almost all the metrics rank pair (4,7) as the worst match. The single discrepancy comes from the Hausdorff metric. However, it lists (4,7) as the penultimately worst match, which globally is in overall agreement with the others. We then used each of these metrics to perform hierarchical clustering, Fig. 11, on the nine shapes. Fig. 11 clearly shows a global trend in the groupings among the different metrics. One can interpret this agreement as a reflection of obvious similarities or dissimilarities among the shapes.

The interesting properties unique to each of these metrics arise in the differences that are apparent in the local trend. We attribute a majority of these local rank differences due to the inherent sensitivities of each metric. These sensitivities are a direct consequences of how they are formulated. For example, it is well known that the Hausdorff metric is biased to outliers due to the *max-min* operations in its definition. The bending energy of the spline models is invariant to affine transformations between shapes, and its increase is a reflection of how one shape has to be "bent" to the other. The differences among the spline models can be attributed to the compact (Wendland) versus infinite (TPS) support of the basis functions. We refer the reader to the aforementioned references for more through discussions of the respective metrics and their formulations.

Though we are in the early stages of investigating the two new metrics and their properties, these results clearly validate their use as a shape metric. The choice of $\sigma^2 = \{0.1, 0.5, 1.5\}$ impacted the local rankings among the two metrics. As Fig. 1 illustrated, σ^2 gives us the ability to "dial-in" the local curvature shape features. When matching shapes, selecting a large value of σ^2 implies that we do not want the matching influenced by localized high-curvature points on the shape. Similarly, a low value of σ^2 reflects our desire to incorporate such features. As an illustration of this, consider the first three dendrograms in the top row of that in Fig. 11. The first two dendrograms were computed using the Fisher-Rao metric with variance parameter $\sigma^2 = \{0.1, 0.5\}$, resulting in shape 6 being ranked as the next best match to pair (1,8). When we set $\sigma^2 = 1.5$, shape 3 now becomes the next best match to (1,8). Hence, we see that σ^2 impacts the shape distance. However, it affects it in such a way that is discernibly natural—meaning that the ranking was not drastically changed, which would not coincide with our visual intuition. The differences between Fisher-Rao and α-order entropy metric arise from the structural differences in their respective metric tensors $g_{i,j}$. The off-diagonal components (corresponding to intralandmarks) of the α-order entropy metric tensor are zero. This decouples the correlation between a landmark's own *x*- and *y* –coordinates, though correlations exist with the coordinates of other landmarks. Intuitively, this changes the curvature of the manifold and shows up visually in the shape of the geodesic [3], which in turn impacts the distance measure.

The α-order entropy metric provided huge computational benefits over the Fisher-Rao metric. The Fisher-Rao metric requires an extra $O(N^2)$ computation of the integral over $\mathbb{R}^2$, where we have assumed an *N* point discretization of the *x*- and *y*-axes. This computation must be repeated at each point along the evolving geodesic and for every *pair* of landmarks. The derivatives of the metric tensor that are needed for geodesic computation require the same $O(N^2)$ computation for every landmark *triple* and at each point on the evolving geodesic. Since our new ɸ-entropy metric tensor and derivatives are in closed form, this extra $O(N^2)$ computation is not required. Please note that the situation only worsens in 3D, where $O(N^3)$ computations will be required for the Fisher-Rao metric (and derivatives), while our new metric (and derivatives) remain in closed form. It remains to be seen if other closed-form information metrics can be derived, which are meaningful in the shape-matching context.

The comparative analysis with other metrics illustrated the utility of Fisher-Rao and α-order entropy metrics as viable shape distance measures. In addition to their discriminating capabilities, these two metrics have several other advantages over the present contemporaries. The representation model based on densities is inherently more robust to noise and uncertainties in the landmark positions. In addition, we showcased the ability of these metrics to deform shapes with various topologies—thus enabling landmark analysis for anatomical forms with interior points or disjoint parts. Most importantly, the deformation is directly obtained from the shape representation, eliminating an arbitrary spline term found in some formulations. The robustness and flexibility of this model has good potential for computational medical applications such as computer-aided diagnosis and biological growth analysis. As a general shape similarity measure, our metrics are yet another tool for general shape recognition problems.

5 CONCLUSIONS

In this paper, we have presented a unified framework for shape representation and deformation. Previous approaches treat representation and deformation as two distinct problems. Our representation of landmark shapes using mixture models enables immediate application of information matrices as Riemannian metric tensors to establish an intrinsic geodesic between shape pairs. To this end, we discussed two such metrics: the Fisher-Rao metric and the new α-order entropy metric. To our knowledge, this is the first time these information geometric principles have been applied to shape analysis. In our framework, shapes modeled as densities live on a statistical manifold, and intrinsic distances between them are readily obtained by computing the geodesic connecting two shapes. Our development of the α-order entropy was primarily motivated by the computational burdens of working with the Fisher-Rao metric. Given that our parameter space comes from GMMs, the Fisher-Rao metric suffers serious computational inefficiencies as it is not possible to get closed-form solutions to the metric tensor or the Christoffel symbols. The new α-order entropy metric, with α = 2, enables us to obtain closed-form solutions to the metric tensor and its derivatives and therefore alleviates this computational burden. We also illustrated how to leverage the intrinsic geodesic path from the two metrics to deform the extrinsic space, important to applications such as registration. Our techniques were applied to matching 2D corpus callosum and fish [40] landmark shapes, illustrating the usefulness of this framework for shape discrimination and deformation analysis. Test results show the applicability of the new metrics to shape matching, providing discriminability similar to several other metrics. Admittedly, we are still in the early stages of working with these metrics and have yet to perform statistical comparisons on the computed shape geodesic distances. These metrics also do not suffer from topological constraints on the shape structure (thus enabling their applicability to a large class of image analysis and other shape analysis applications).

Our intrinsic, coupled representation and deformation framework is not only limited to landmark shape analysis, where correspondence is assumed to be known. The ultimate practicality and utility of this approach will be realized upon extension of these techniques to unlabeled point sets, where correspondence is unknown. Existing solutions to this more difficult problem have only been formulated via models that decouple the shape representation and deformation, e.g., [10]. Though the metrics presented in this work result from second order analysis of the generalized entropy, it is possible to extend the framework to incorporate other probabilistic, Riemannian metrics. For example, one can perform intrinsic analysis on the manifold of von Mises mixture densities, which is particularly useful for unit vector data [44].

The immediate next step is to move beyond landmarks and model shape point sets using GMMs thereby estimating the free parameter σ^2 directly from the data. It is also possible to incorporate the full covariance matrix enabling the mixture density representation to have richer descriptive power for point-set shapes. Our future work will focus on extending this framework to incorporate diffeomorphic warping of the extrinsic space and investigation into other information metrics—especially ones that leverage the $\sqrt{p}$ representation [30], [45], [35], [46] since this results in geodesics on hyperspheres. Extensions to 3D shape matching are also possible.

Acknowledgments

REFERENCES

1. Bookstein, FL. Morphometric Tools for Landmark Data: Geometry and Biology. Cambridge Univ. Press; 1991.

2. Peter, A.; Rangarajan, A. Shape Matching Using the Fisher-Rao Riemannian Metric: Unifying Shape Representation and Deformation; Proc. IEEE Int'l Symp. Biomedical Imaging (ISBI '06); 2006. p. 1164-1167.

3. Peter, A.; Rangarajan, A. A New Closed-Form Information Metric for Shape Analysis; Proc. Int'l Conf. Medical Image Computing and Computer Assisted Intervention (MICCAI '06); 2006. p. 249-256.

4. Small, C. The Statistical Theory of Shape. Springer; 1996.

5. 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.

6. Rohr K, Stiehl HS, Sprengel R, Buzug TM, Weese J, Kuhn MH. Landmark-Based Elastic Registration Using Approximating Thin-Plate Splines. IEEE Trans. Medical Imaging 2001 June;vol. 20(no. 6):526–534.

7. Davies, RH.; Twining, C.; Cootes, TF.; Taylor, CJ. An Information Theoretic Approach to Statistical Shape Modelling; Proc. British Machine Vision Conf. (BMVC '01); 2001. p. 3-12.

8. Camion, V.; Younes, L. Geodesic Interpolating Splines; Proc. Int'l Conf. Energy Minimization Methods for Computer Vision and Pattern Recognition (EMMCVPR '01); 2001. p. 513-527.

9. Joshi S, Miller M. Landmark Matching via Large Deformation Diffeomorphisms. IEEE Trans. Image Processing 2000 Aug.;vol. 9(no. 8):1357–1370.

10. Chui H, Rangarajan A. A New Point Matching Algorithm for Non-Rigid Registration. Computer Vision and Image Understanding 2003 Mar.;vol. 89(nos. 2–3):114–141.

11. Guo, H.; Rangarajan, A.; Joshi, S. 3-D Diffeomorphic Shape Registration on Hippocampal Data Sets; Proc. Int'l Conf. Medical Image Computing and Computer Assisted Intervention (MICCAI '05); 2005. p. 984-991.

12. Siddiqi, K.; Shokoufandeh, A.; Dickinson, SJ.; Zucker, SW. Shock Graphs and Shape Matching; Proc. IEEE Int'l Conf. Computer Vision (ICCV '98); 1998. p. 222-229.

13. Srivastava A, Joshi S, Mio W, Liu X. Statistical Shape Analysis: Clustering, Learning and Testing. IEEE Trans. Pattern Analysis and Machine Intelligence 2005 Apr.;vol. 27(no. 4):590–602.

14. Thompson P, Toga AW. A Surface-Based Technique for Warping Three-Dimensional Images of the Brain. IEEE Trans. Medical Imaging 1996 Aug.;vol. 5(no. 4):402–417.

15. Burbea J, Rao R. Entropy Differential Metric, Distance and Divergence Measures in Probability Spaces: A Unified Approach. J. Multivariate Analysis 1982;vol. 12:575–596.

16. Havrda ME, Charvát F. Quantification Method of Classification Processes: Concept of Structural α-entropy. Kybernetica 1967;vol. 3:30–35.

17. Woods YWK, McClain M. Information-Theoretic Matching of Two Point Sets. IEEE Trans. Image Processing 2002 Aug.;vol. 11(no. 8):868–872.

18. Wang, F.; Vemuri, BC.; Rangarajan, A.; Schmalfuss, IM.; Eisenschenk, SJ. Simultaneous Nonrigid Registration of Multiple Point Sets and Atlas Construction; Proc. European Conf. Computer Vision (ECCV '06); 2006. p. 551-563.

19. Jian, B.; Vemuri, BC. A Robust Algorithm for Point Set Registration Using Mixture of Gaussians; Proc. IEEE Int'l Conf. Computer Vision (ICCV '05); 2005. p. 1246-1251.

20. Lin J. Divergence Measures Based on the Shannon Entropy. IEEE Trans. Information Theory 1991 Jan.;vol. 37(no. 1):145–151.

21. Paragios N, Rousson M, Ramesh V. Non-Rigid Registration Using Distance Functions. Computer Vision and Image Understanding 2003 Mar.;vol. 89(nos. 2–3):142–165.

22. Glaunes, J.; Trouvé, A.; Younes, L. Diffeomorphic Matching of Distributions: A New Approach for Unlabeled Point-Sets and Sub-Manifolds Matching; Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR); 2004. p. 712-718.

23. Wahba, G. Spline Models for Observational Data. Soc. for Industrial and Applied Math.; 1990.

24. Cootes, T.; Taylor, C. A Mixture Model for Representing Shape Variation; Proc. British Machine Vision Conf. (BMVC '97); 1997. p. 110-119.

25. McLachlan, GJ.; Basford, KE. Mixture Models: Inference and Applications to Clustering. Marcel Dekker; 1988.

26. Figueiredo MAT, Jain AK. Unsupervised Learning of Finite Mixture Models. IEEE Trans. Pattern Analysis and Machine Intelligence 2002 Mar.;vol. 24(no. 3):381–396.

27. Rao CR. Information and Accuracy Attainable in Estimation of Statistical Parameters. Bull. of the Calcutta Math. Soc 1945;vol. 37:81–91.

28. Boothby, WM. An Introduction to Differentiable Manifolds and Riemannian Geometry. Academic Press; 2002.

29. Čencov NN. Statistical Decision Rules and Optimal Inference. Am. Math. Soc. 1982

30. Amari S-I, Nagaoka H. Methods of Information Geometry. Am. Math. Soc. 2001

31. Myung IJ, Balasubramanian V, Pitt MA. Counting Probability Distributions: Differential Geometry and Model Selection. Proc. Nat'l Academy of Sciences 2000;vol. 97:11 170–11 175.

32. Maybank SJ. The Fisher-Rao Metric for Projective Transformations of the Line. Int'l J. Computer Vision 2005;vol. 63(no. 3):191–206.

33. Mio W, Badlyans D, Liu X. A Computational Approach to Fisher Information Geometry with Applications to Image Analysis. Proc. Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR '05) 2005:18–33.

34. Lenglet C, Rousson M, Deriche R, Faugeras O. Statistics on the Manifold of Multivariate Normal Distributions: Theory and Application to Diffusion Tensor MRI Processing. J. Math. Imaging and Vision 2006;vol. 25(no. 3):423–444.

35. Srivastava, A.; Jermyn, I.; Joshi, S. Riemannian Analysis of Probability Density Functions with Applications in Vision; Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR); 2007. p. 1-8.

36. Courant, R.; Hilbert, D. Methods of Math. Physics. Vol. vol. 2. Wiley Interscience; 1989.

37. Mio, W.; Liu, X. Landmark Representation of Shapes and Fisher-Rao Geometry; Proc. IEEE Int'l Conf. Image Processing (ICIP '06); 2006. p. 2113-2116.

38. Horn, BKP. Robot Vision. MIT Press; 1986.

39. Costa SIR, Santos S, Strapasson JE. Fisher Information Matrix and Hyperbolic Geometry. Proc. IEEE Information Theory Workshop 2005:28–30.

40. Catalog of Fishes. Dept. of Ichthyology, California Academy of Sciences; 2007. http://research.calacademy.org/research/ichthyology/

41. Kendall DG. Shape-Manifolds, Procrustean Metrics and Complex Projective Spaces. Bull. of the London Math. Soc 1984;vol. 16:81–121.

42. Huttenlocher DP, Klanderman GA, Rucklidge WJ. Comparing Images Using the Hausdorff Distance. IEEE Trans. Pattern Analysis and Machine Intelligence 1993 Sept.;vol. 15(no. 9):850–863.

43. Fornefett M, Rohr K, Stiehl HS. Radial Basis Functions with Compact Support for Elastic Registration of Medical Images. Image and Vision Computing 2001 Jan.;vol. 19(no. 1):87–96.

44. Banerjee A, Dhillon I, Ghosh J, Sra S. Clustering on the Unit Hypersphere Using Von Mises-Fisher Distributions. J. Machine Learning Research 2005;vol. 6:1345–1382.

45. Lebanon, G. PhD dissertation. Carnegie Mellon Univ.; 2005. Riemannian Geometry and Statistical Machine Learning.

46. Peter, A.; Rangarajan, A.; Ho, J. Shape L'Âne Rouge: Sliding Wavelets for Indexing and Retrieval; Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR); 2008. p. 1-8.

47. GatorBait 100. Univ. of Florida; 2008. http://www.cise.ufl.edu/~anand/GatorBait_100.tgz

Biographies
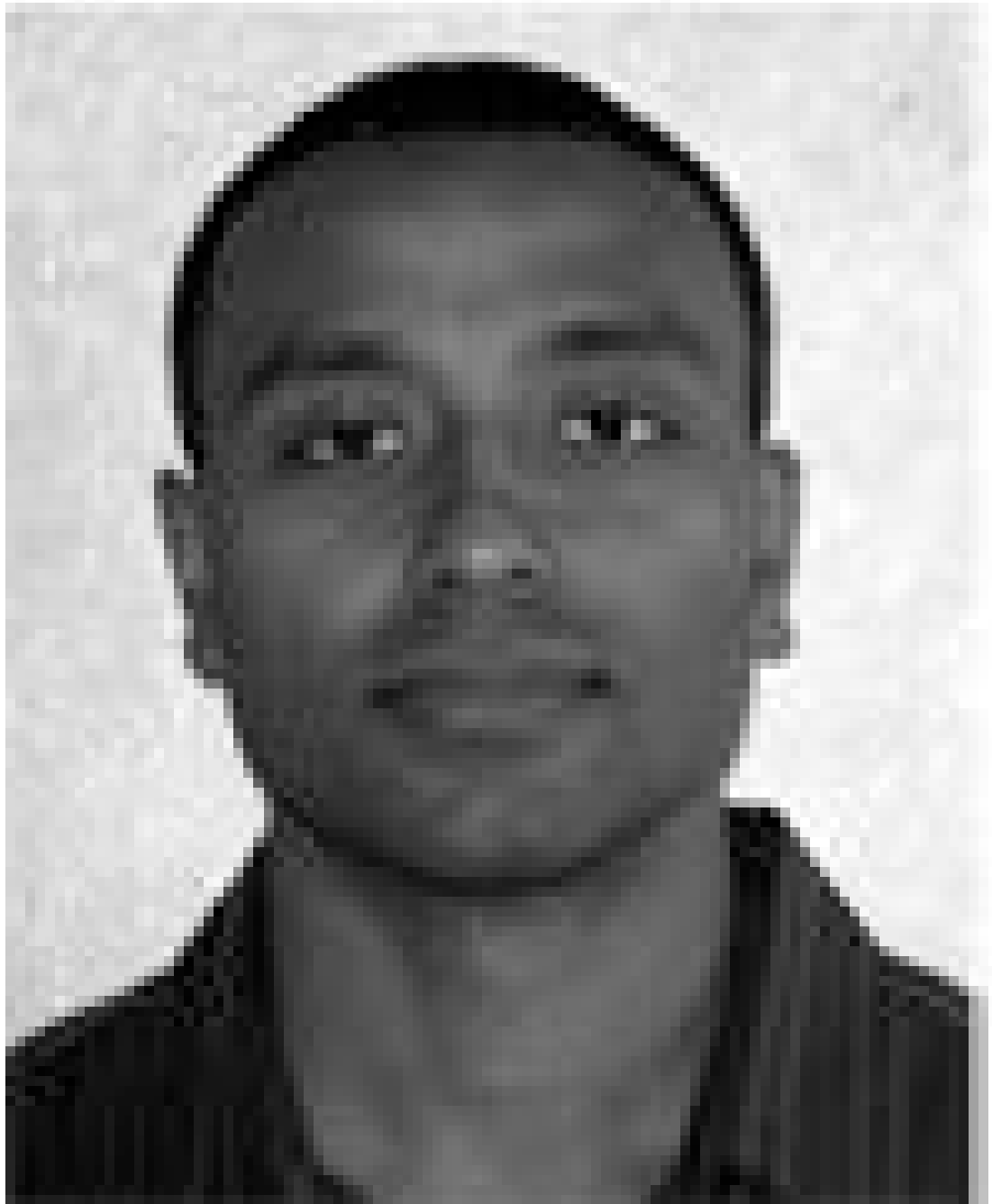
Adrian M. Peter received the BS degree (high honors) in computer engineering and the MS degree in electrical and computer engineering from the University of Florida, Gainesville, in 1999 and 2003, respectively. He returned there in 2005 to complete the PhD degree in electrical and computer engineering, which he completed in 2008. After receiving his BS degree, he worked at the Intel Corporation, where he last held the position of a technology initiatives program manager, responsible for short-range wireless products. After receiving his MS degree, he worked at the Harris Corporation, developing image analysis algorithms for a variety of remote sensing platforms and was a coinventor on eight pending patents. His

research interests are in machine learning and computer vision with a focus on applying information geometry to shape analysis. He is a student member of the IEEE.


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 Departments of Diagnostic Radiology and Computer Science, Yale University. From 1992 to 1995, he held a joint research faculty position in both departments. From 1995 to 2000, he was an assistant professor in the Image Processing and Analysis Group (IPAG), Departments of Diagnostic Radiology and Electrical Engineering,

Yale University. He is currently an associate professor in the Department of Computer and Information Science and Engineering (CISE), University of Florida. His current research interests are best summarized as the application of machine learning to image analysis. He is also interested in the scientific study of consciousness. In 1992, he was the chair of the Neural Information Processing Systems (NIPS) Workshop entitled Deterministic Annealing and Combinatorial Optimization, and in 1995, he was a cochair of a NIPS Workshop entitled Statistical and Structural Models in Network Vision. He has served on the program committees of Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR 1997, 1999, 2001, and 2007), the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2000, 2001, 2003, and 2006), and Information Processing in Medical Imaging (IPMI 2005 and 2007). He has also cochaired the IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA) 2001 and EMMCVPR 2003 and 2005 and was an area chair for the IEEE International Conference on Computer Vision (ICCV) 2007. He has been active in the Toward a Science of Consciousness conferences (1996, 1998, 2000, 2002, and 2008) and the International Congress of Vedanta conferences (2002, 2005, and 2007). He is a member of the IEEE.
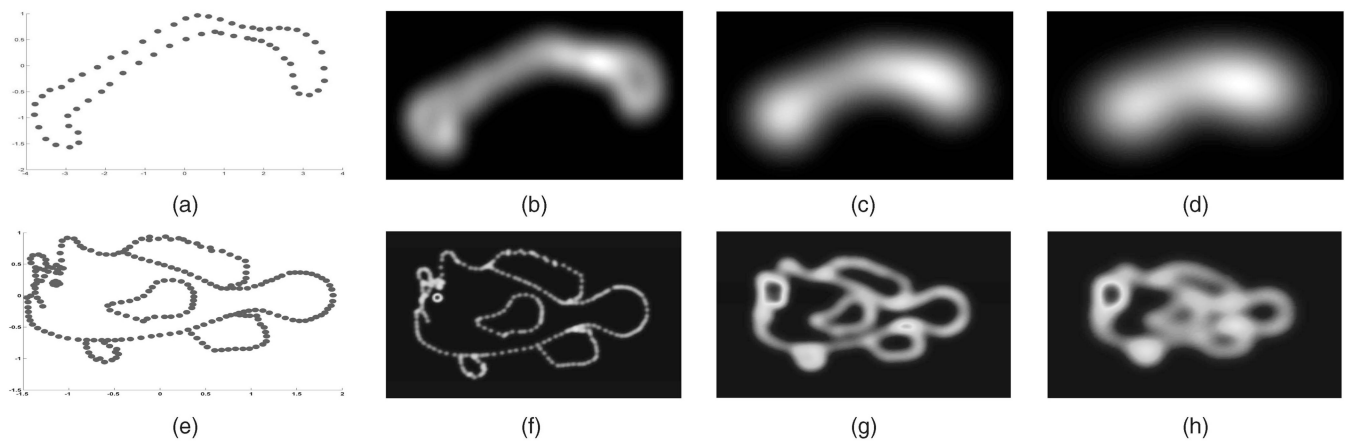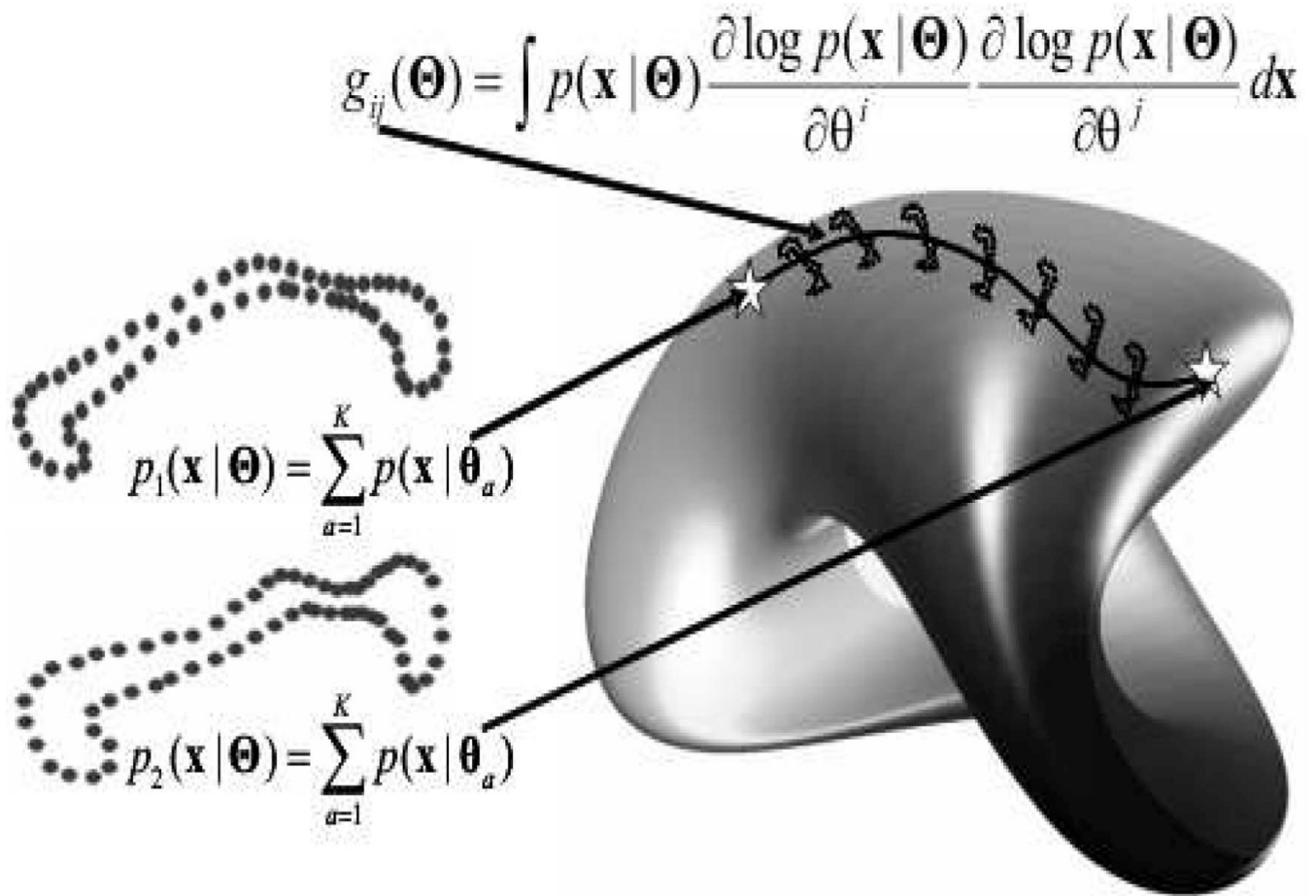
Fig. 1.
Examples of the probabilistic representation model. (a) Original shape consisting of 63 landmarks ($K = 63$). (b), (c), and (d) Overhead view of K-component GMM using $\sigma^2 = 0.1$, $\sigma^2 = 0.5$, and $\sigma^2 = 1.5$, respectively. (e) Original shape consisting of 233 landmarks ($K = 233$). (f), (g), and (h) Overhead view of K-component GMM using $\sigma^2 = 0.001$, $\sigma^2 = 0.01$, and $\sigma^2 = 0.025$, respectively.
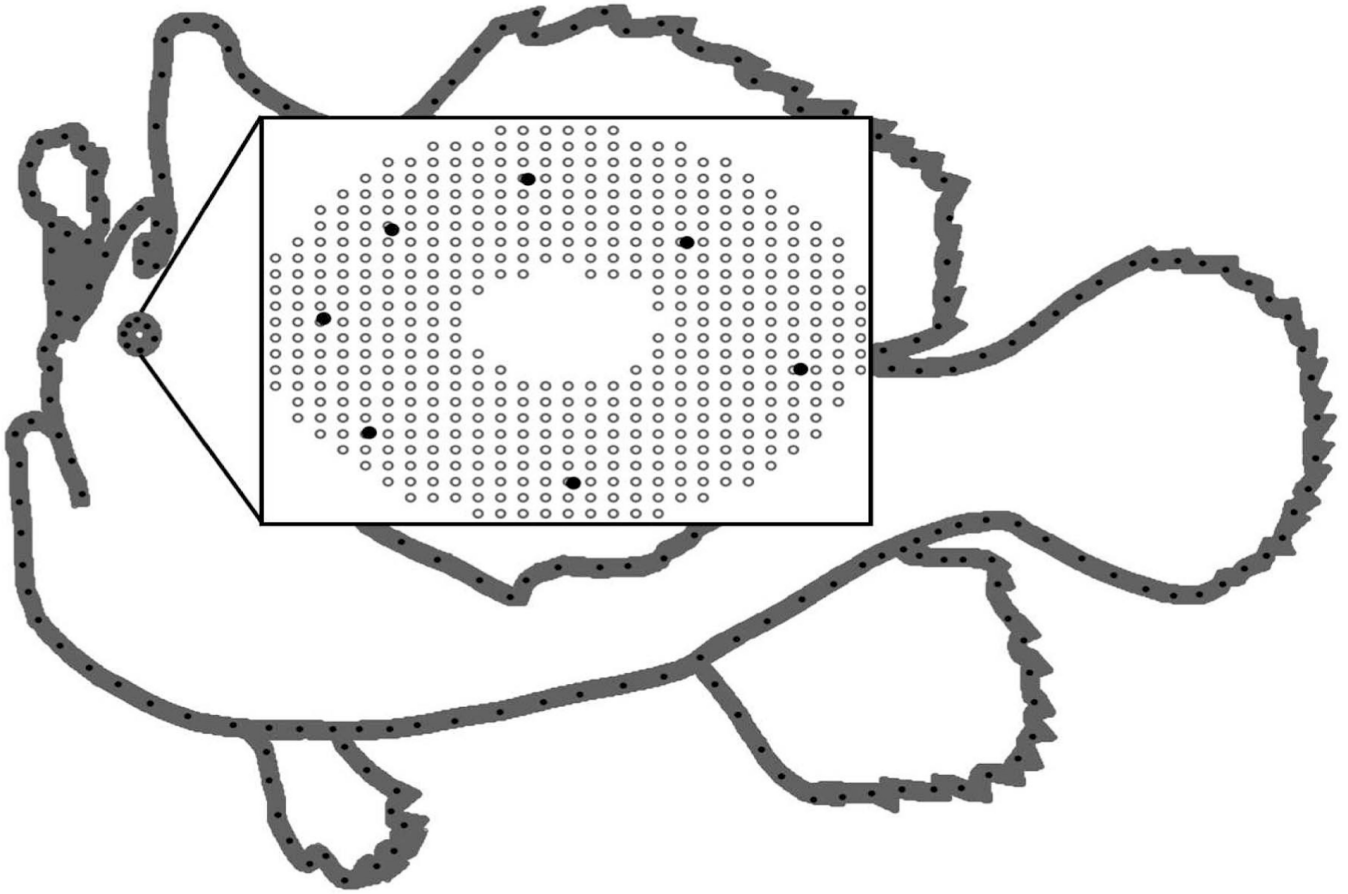
Fig. 2.
Intrinsic versus extrinsic. The original fish data consists of 49K points (due to image resolution, these show up as light gray outline, see zoomed in eye for clearer depiction). The 233 landmarks are illustrated by solid black points. The landmarks are used for intrinsic analysis since they are used as the means of 233 component GMM. See Section 3.5 for the method to move the extrinsic points (surrounding the landmarks) based on the landmark movement.

$$g_{ij}(\Theta) = \int p(\mathbf{x}|\Theta) \frac{\partial \log p(\mathbf{x}|\Theta)}{\partial \theta^i} \frac{\partial \log p(\mathbf{x}|\Theta)}{\partial \theta^j} d\mathbf{x}$$

$$p_1(\mathbf{x}|\Theta) = \sum_{a=1}^{K} p(\mathbf{x}|\theta_a)$$

$$p_2(\mathbf{x}|\Theta) = \sum_{a=1}^{K} p(\mathbf{x}|\theta_a)$$

Fig. 3.
Intrinsic shape matching. Two landmark shapes represented as mixture models end up as two points on the probabilistic manifold. Using the metric tensor $g_{i,j}$, it is possible to obtain a geodesic between the shapes.

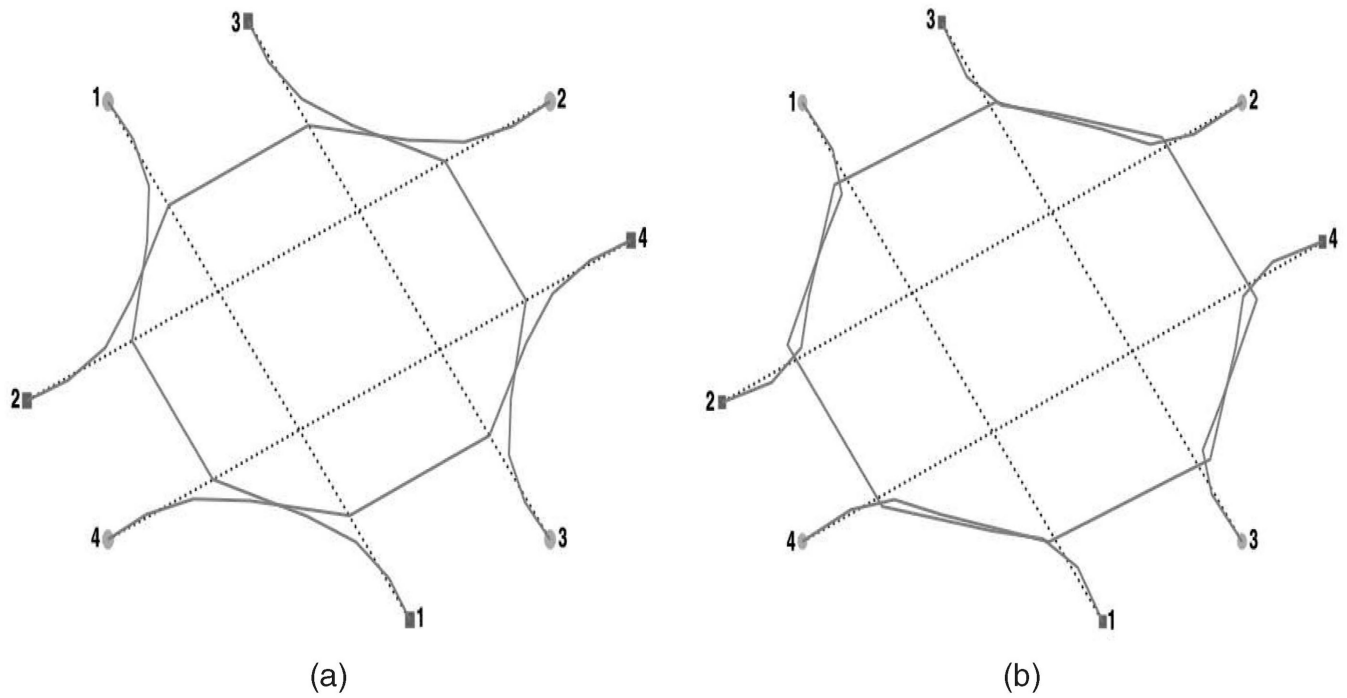(a)   (b)
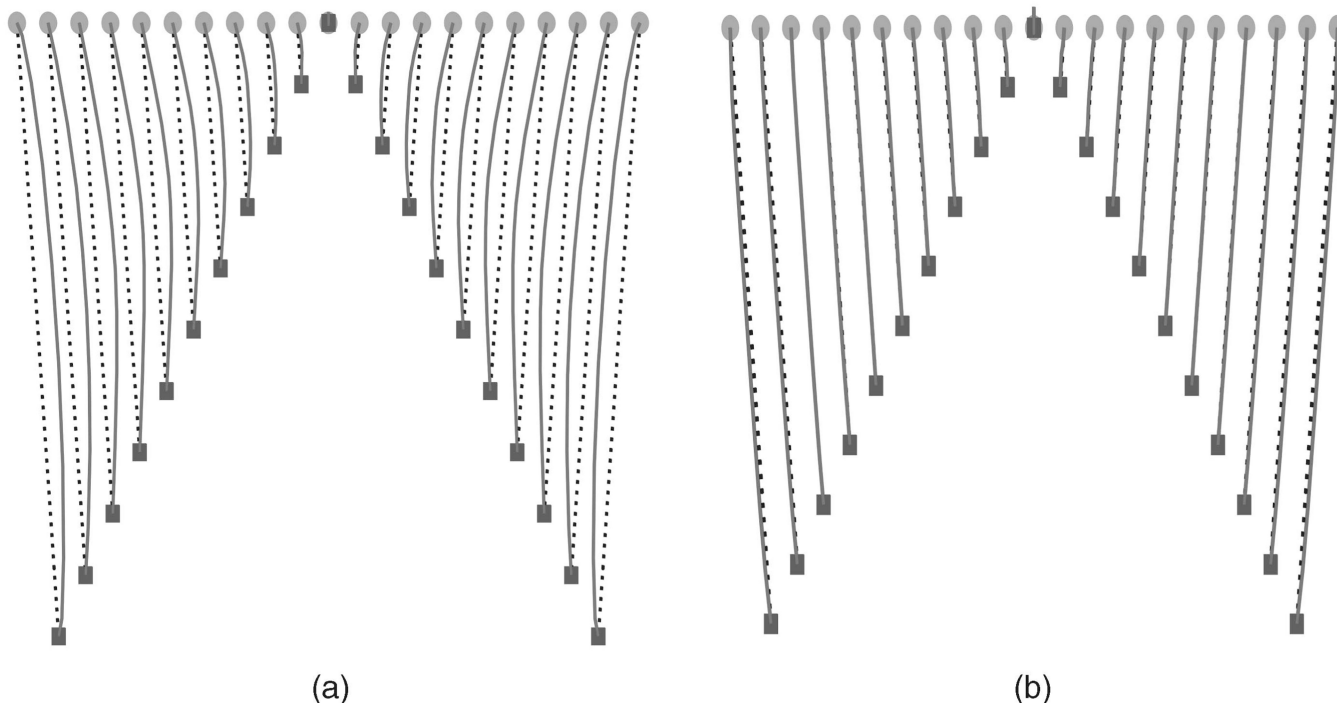
Fig. 4.
Bending of straight line with 21 landmarks. The dashed line is the initialization, and the
solid line is the final geodesic. (a) Curvature of space under the Fisher information metric
evident in final geodesic. (b) The space under α-order entropy metric is not as visually
curved for this transformation.
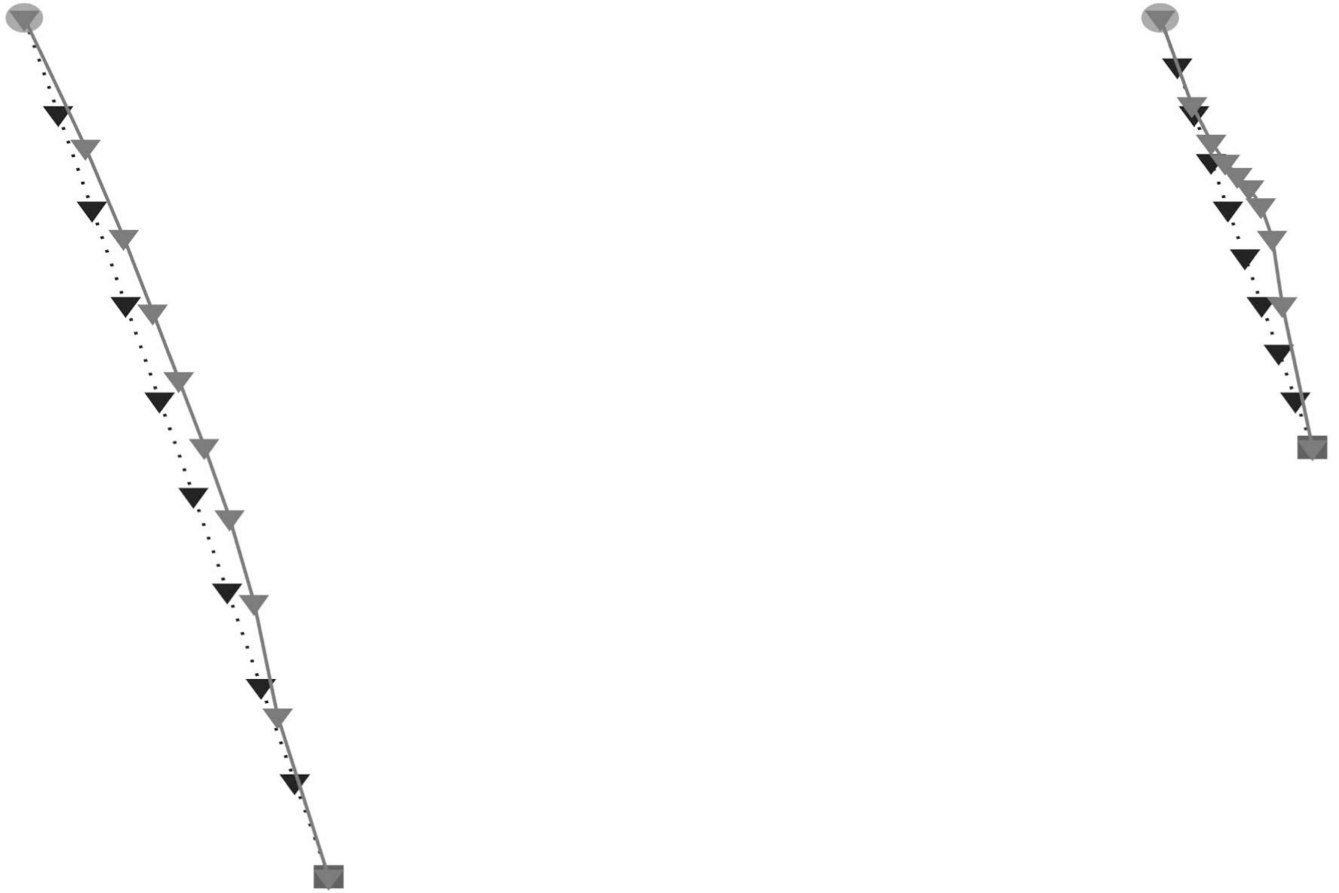
Fig. 5.
Intermediate landmark trajectories under the α-order entropy metric tensor. These are the second and third landmarks from the middle in Fig. 4b. The trajectories show that even though the final geodesic looks similar to the straight line initialization, the intermediate landmark positions have changed, which results in different velocities along the geodesic.

(a)　　　　　　　　　　　　　　　　(b)

Fig. 6.
Rotation of square represented with four landmarks. The dashed line is the initialization, and the solid line is the final geodesic. The circular landmarks are the starting shape, and square landmarks are the rotated shape. (a) Fisher information metric path is curved smoothly. (b) α-entropy metric path has sharp corners.
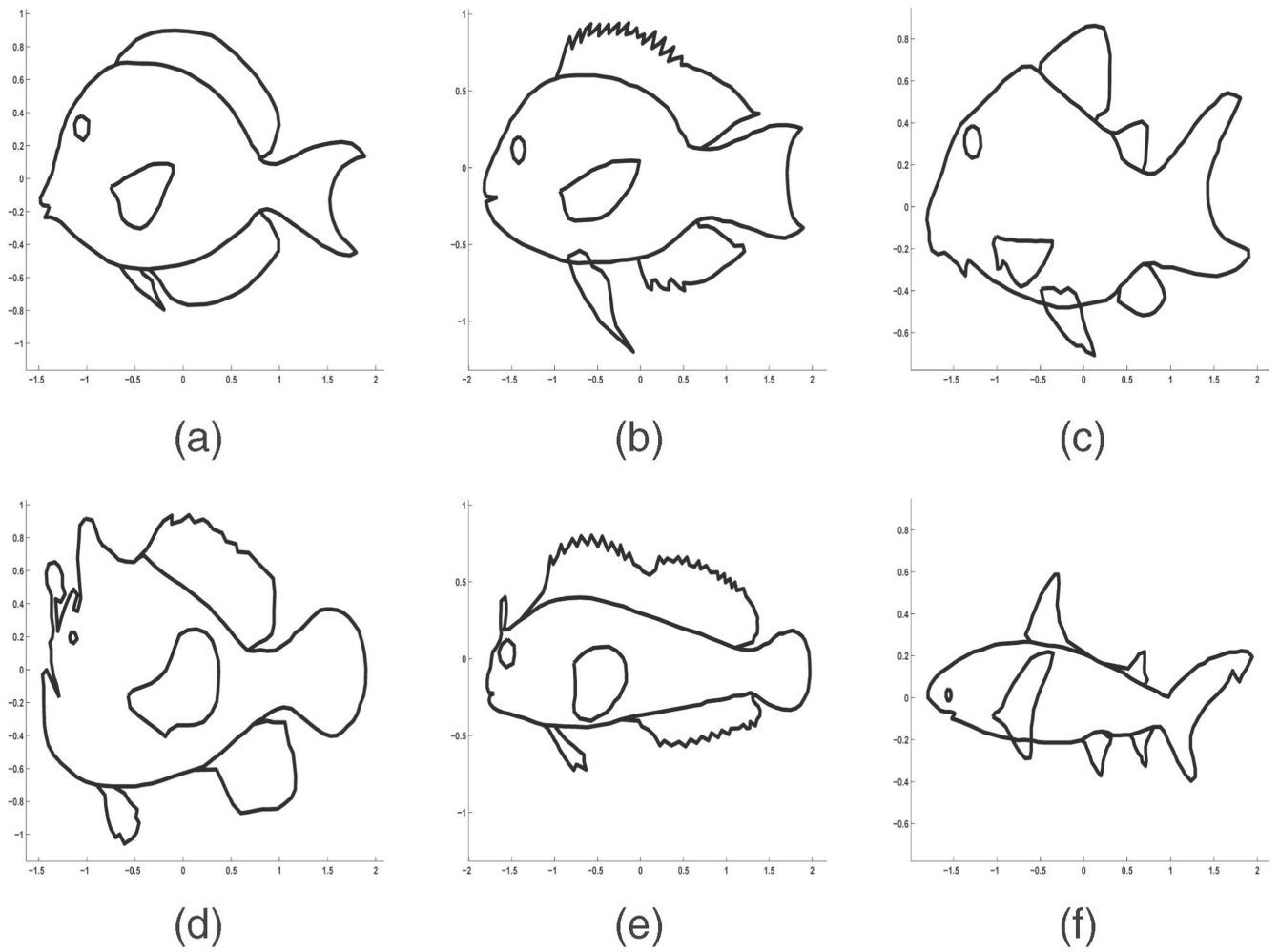
Fig. 7.
Fish shapes with differing topologies. For each vertical pair, we extracted equal number of landmarks: (a) and (d) 233, (b) and (e) 253, and (c) and (d) 214. Source images for these fish were obtained from that in [40].
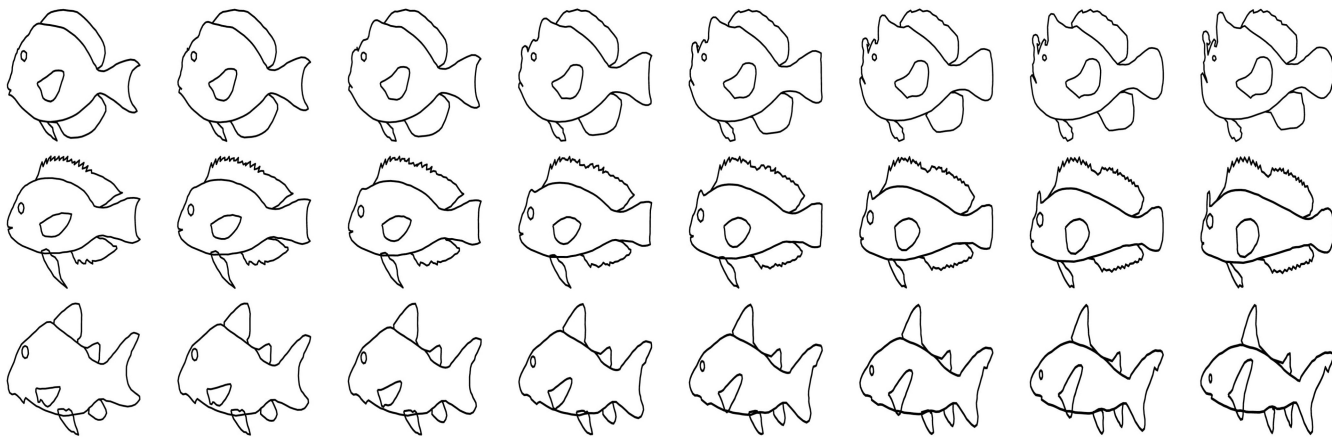
Fig. 8.
Deformation analysis using fish from that in Fig. 7 using α-order entropy metric. Top row shows intermediate warps between (a) and (d), $\sigma^2 = 0.5$. Middle row shows intermediate warps between (b) and (e), $\sigma^2 = 0.25$. Bottom row shows intermediate warps between (c) and (f), $\sigma^2 = 0.25$. The deformations *do not* require a spline model.
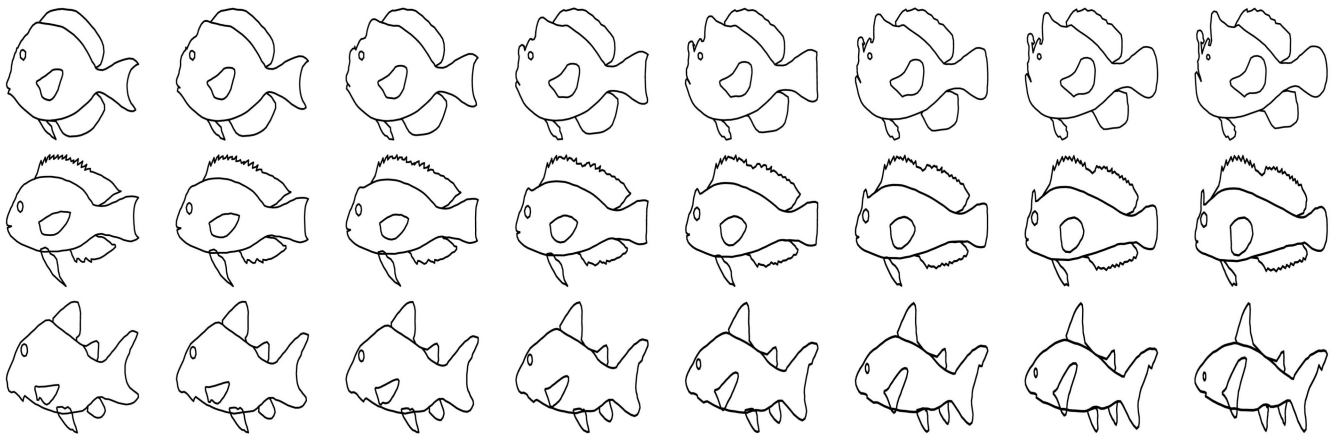
Fig. 9.
Deformation analysis using fish from that in Fig. 7 using landmark diffeomorphisms [8]. All shapes computed with λ = 10. Top row shows intermediate warps between (a) and (d). Middle row shows intermediate warps between (b) and (e). Bottom row shows intermediate warps between (c) and (f). These deformations *require* a spline model.
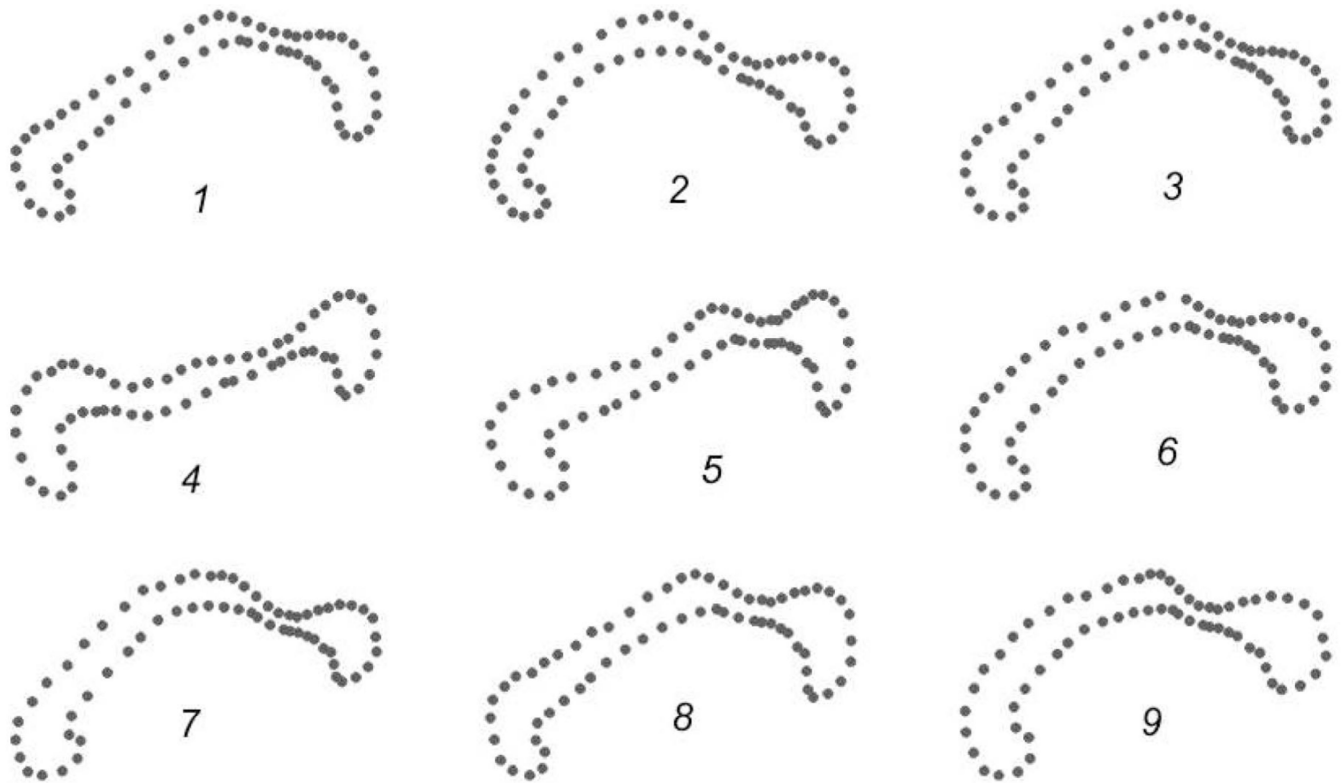
Fig. 10.
Nine corpus callosum shapes used for pairwise matching, 63 landmarks per shape.
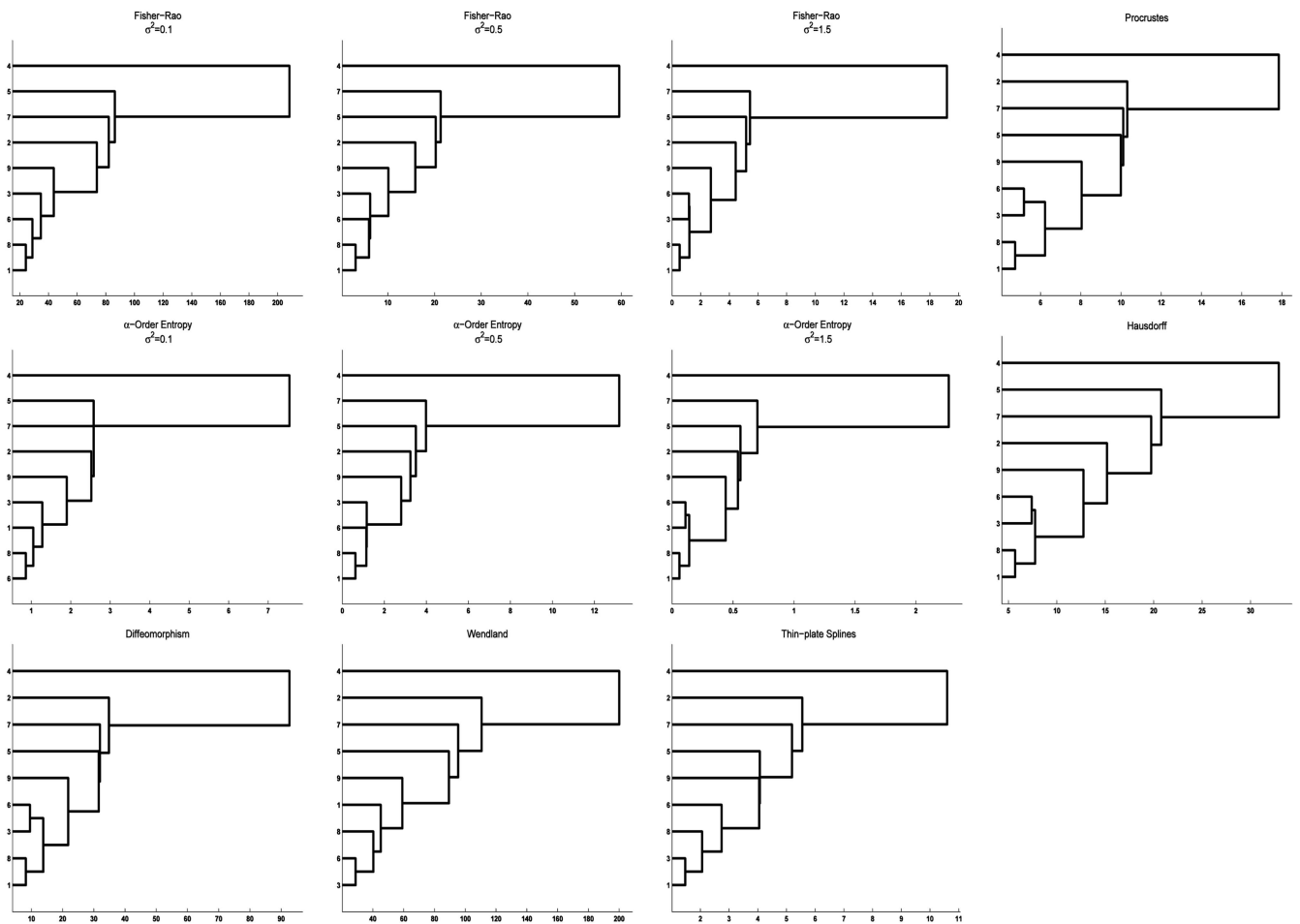
Fig. 11.
Hierarchical clustering with different metrics. Notice that varying σ on the Fisher-Rao and α-Order Entropy metric does not significantly impact global grouping of the shapes (see first three columns of rows one and two). Almost all the metrics agree that shapes 1 and 8 are the best match, while shape 4 is the most dissimilar.

TABLE 1

Pairwise Shape Distances

Pairs	Fishers-Rao (10^{-2})			α-Order Entropy (10^{-3})			Diffeomorphism (10^{-2})	Procrustes (10^{-2})	Hausdorff (10^{-2})	Wendland (10^{-2})	TPS (10^{-2})
	σ²=.1	σ²=.5	σ²=1.5	σ²=.1	σ²=.5	σ²=1.5					
1 vs. 2	142.25	27.17	5.85	4.67	4.64	0.54	45.05	11.73	27.15	128.39	7.72
1 vs. 3	62.22	14.59	3.80	2.06	2.66	0.40	17.72	7.74	11.83	45.08	1.47
1 vs. 4	375.07	87.04	20.31	13.73	16.29	2.27	114.17	18.95	50.29	203.60	10.60
1 vs. 5	119.75	26.72	6.79	4.09	5.07	0.80	42.80	11.49	25.52	131.52	8.28
1 vs. 6	54.15	9.83	2.02	2.15	2.22	0.26	17.97	7.19	13.85	65.04	4.77
1 vs. 7	206.41	52.81	14.76	7.63	10.96	1.88	81.49	16.53	57.06	227.29	13.28
1 vs. 8	24.07	3.08	0.53	1.05	0.62	0.06	8.20	4.73	5.69	50.89	3.05
1 vs. 9	161.57	32.19	7.36	6.65	8.05	1.07	58.49	13.27	26.54	192.29	12.92
2 vs. 3	106.46	20.92	5.86	3.65	3.82	0.65	39.63	11.21	17.32	123.01	6.48
2 vs. 4	571.37	136.56	29.39	19.65	23.83	3.02	182.93	23.54	117.74	351.38	17.62
2 vs. 5	367.50	86.10	21.29	11.00	14.41	2.16	123.99	19.72	72.76	312.08	16.74
2 vs. 6	73.74	15.88	4.44	2.52	3.24	0.55	34.84	10.31	15.19	110.61	5.55
2 vs. 7	150.02	44.22	15.18	5.03	8.86	1.96	80.38	16.47	71.46	254.76	11.72
2 vs. 8	136.85	27.96	6.39	3.95	4.56	0.60	53.75	12.68	23.35	169.20	9.42
2 vs. 9	94.52	20.60	5.02	3.74	5.59	0.87	43.67	11.53	28.81	147.21	10.52
3 vs. 4	610.51	153.60	38.20	21.85	28.07	4.27	201.91	25.17	93.71	348.10	11.13
3 vs. 5	231.03	53.58	12.41	6.92	8.57	1.12	67.43	14.55	33.53	153.21	6.80
3 vs. 6	34.58	6.21	1.18	1.28	1.16	0.11	9.54	5.17	7.41	28.71	2.74
3 vs. 7	92.02	21.34	5.44	3.67	4.68	0.75	39.61	11.28	19.74	100.58	6.74
3 vs. 8	59.26	13.33	3.27	1.86	2.24	0.32	18.32	7.69	12.11	47.59	2.06
3 vs. 9	119.42	22.62	4.71	5.18	5.75	0.69	40.40	10.96	29.79	116.41	9.39
4 vs. 5	208.30	59.56	19.19	7.54	13.18	2.70	92.67	17.85	32.92	200.05	12.84
4 vs. 6	435.13	110.01	27.50	15.85	21.45	3.32	147.27	21.96	64.50	311.83	23.36
4 vs. 7	682.10	193.47	54.20	25.14	37.77	6.59	229.74	28.60	104.18	499.73	34.32
4 vs. 8	325.84	79.77	19.83	11.48	14.97	2.30	105.93	18.71	61.59	224.42	16.66
4 vs. 9	512.94	132.76	33.98	18.76	26.97	4.17	172.52	23.82	72.78	374.71	25.14

Pairs	Fishers-Rao (10^{-2})			α-Order Entropy (10^{-3})			Diffeomorphism (10^{-2})	Procrustes (10^{-2})	Hausdorff (10^{-2})	Wendland (10^{-2})	TPS (10^{-2})
5 vs. 6	163.69	37.42	8.72	4.56	5.68	0.76	56.41	13.01	28.47	157.91	10.74
5 vs. 7	311.52	78.63	19.60	8.88	12.34	1.79	91.71	17.46	74.11	233.17	13.85
5 vs. 8	86.32	20.26	5.17	2.58	3.50	0.56	31.57	9.99	20.78	89.38	4.07
5 vs. 9	270.52	63.21	16.13	7.75	11.11	1.63	81.30	16.31	42.61	224.62	12.78
6 vs. 7	82.06	22.30	6.80	2.58	4.06	0.79	38.31	11.13	23.74	105.01	5.70
6 vs. 8	28.72	5.96	1.21	0.86	1.14	0.14	13.81	6.22	7.76	40.29	3.33
6 vs. 9	43.65	10.11	2.71	1.90	2.80	0.44	21.81	8.04	12.75	59.11	4.05
7 vs. 8	145.55	40.08	11.83	4.62	7.85	1.45	67.70	14.50	38.37	151.59	6.87
7 vs. 9	85.01	21.31	6.45	2.62	3.98	0.70	31.97	10.11	28.22	95.40	5.19
8 vs. 9	103.71	23.95	5.87	3.68	5.62	0.80	47.65	11.84	20.90	126.93	9.24

All of the corpora callosa were matched with each other. Fisher-Rao and α-Order Entropy metrics were computed with three different values of $\sigma^2 = \{0.1, 0.5, 1.5\}$ to assess the impact of the free parameter on shape distance. Shapes 1 and 8 have the smallest distance under almost all the distances, while 4 versus 7 is the worst. (See text for more discussions.)