Learning Metrics for Shape Classification and Discrimination*

Yu Fan[1], David Houle[2], Washington Mio[1]
[1]*Department of Mathematics,* [2]*Department of Biological Sciences*
Florida State University, Tallahassee, FL 32306, USA
yfan@math.fsu.edu, dhoule@bio.fsu.edu, mio@math.fsu.edu

Abstract

We propose a family of shape metrics that generalize the classical Procrustes distance by attributing weights to general linear combinations of landmarks. We develop an algorithm to learn a metric that is optimally suited to a given shape classification problem. Shape discrimination experiments are carried out with phantom data, as well as landmark data representing the shape of the wing of different species of fruit flies.

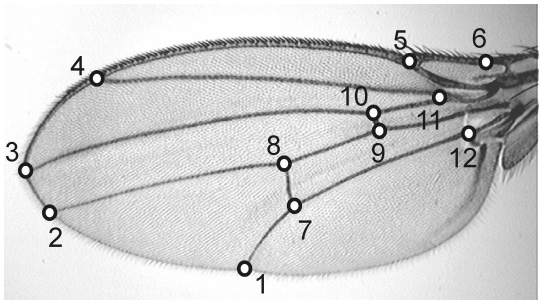

Figure 1. Twelve landmark points on the wing of a fruit fly.

1. Introduction

We introduce a family of shape metrics that generalize the classical Procrustes distance [5, 6] and investigate criteria to select a metric that is best suited to a given shape classification problem. One of the applications that motivate this study of shape is the problem of uncovering morphological characteristics of the wings of fruit flies that can reliably distinguish different species. This is one of a broader set of problems that arise in the investigation of how genetic changes affect phenotype. As indicated in Figure 1, the points where the veins of a wing meet yield a natural collection of 12 landmark points to base the shape analysis upon. This immediately suggests the use of classical Procrustes analysis, which relies on shape representation by indexed collections of landmarks.

If a shape in k-dimensional space $\mathbb{R}^k$ is represented by n landmark points $p_1, \ldots, p_n \in \mathbb{R}^k$, let

$$P = [p_1 \ldots p_n] \tag{1}$$

be the $k \times n$ matrix whose jth column records the coordinates of p_j. If P and Q are two such configurations, after normalization to make the representation invariant to translation and scale, and Procrustes alignment to account for relative orientation, the distance $\|P - Q\|$ is frequently used to quantify shape dissimilarity, where $\|\cdot\|$ is the (Frobenius) norm associated with the inner product

$$\langle P, Q\rangle = \sum_{i=1}^{n} p_i \cdot q_i = \text{trace}\,(PQ^T)\,. \tag{2}$$

Oftentimes, however, the main characteristics that differentiate populations of shapes are confined to specific regions. As such, metrics that attribute equal importance to all landmarks may dilute the most salient differences and that can have an adverse effect on their discriminative power. Motivated by this observation, we propose shape metrics that are able to emphasize particular regions of a configuration. One possible approach is to simply modify the metric using weights for the landmarks, but this is not entirely satisfactory because an important shape feature may depend on a combination of two or more landmarks. Thus, we propose to analyze shape with metrics derived from more general inner products of the form

$$\langle P, Q\rangle_M = \langle PM, Q\rangle = \text{trace}\,(PMQ^T)\,, \tag{3}$$

where M is an $n \times n$ positive-definite, symmetric ma-

*This research was supported in part by NSF grant DMS-0713012 and NIH Roadmap for Medical Research grant U54 RR021813.

1051-4651/10 $26.00 © 2010 IEEE
DOI 10.1109/ICPR.2010.650

trix. The associated distance function is

$$\|P - Q\|_M = \sqrt{\langle P - Q, P - Q\rangle_M}\,. \tag{4}$$

If M is the identity matrix, we ssometimes drop the subscript and use the usual notation $\langle\,,\rangle$ and $\|\cdot\|$. If M is a diagonal matrix with diagonal entries $\lambda_i > 0$, then

$$\langle P, Q\rangle_M = \sum_{i=1}^{n} \lambda_i (p_i \cdot q_i)\,, \tag{5}$$

so that the metric simply assigns weights to the landmarks. However, in general, the M-metric involves more intricate linear combinations of the landmarks. This construction yields a whole family of shape metrics and raises the question of how to choose one that is most effective for a given shape classification problem. We propose a selection criterion based on a principle analogous to that adopted in Linear Discriminant Analysis (LDA) [2]. In other words, given training shapes representing various classes of shapes, we develop a method to estimate M so that the associated shape metric maximizes class separation while minimizing within-class spread. However, unlike LDA, this criterion will be used in the non-linear realm of shapes. Since the measurements of class spread and separation involve mean shapes (cf. [1]), we also address the estimation of means with respect to these generalized shape metrics. We carry out various experiments starting with a preliminary illustration with phantom data. We apply the method to the construction of a metric that discriminates species of fruit flies based on the shape of their wings. The method also allows us to identify the specific features that contribute the most to the discrimination of the species. We compare the results with those obtained with the usual Procrustes distance.

2. Generalized Shape Models

Let P be a $k \times n$ matrix representing an ordered collection of n landmarks in $\mathbb{R}^k$ as in (1). Since configurations that differ by translations and scale represent the same shape, as usual, we normalize P to have its centroid at the origin and scale the matrix to have unit Frobenius norm; that is, $\|P\| = 1$. Here, we assume that the degenerate cases where all landmark points are the same are excluded, which is equivalent to saying that $P \neq 0$ after centering. The space formed by all $k \times n$ matrices representing normalized configurations is known as the *pre-shape space*, which we denote $P(k, n)$. The condition on the centroid restricts P to a subspace of $\mathbb{R}^{k\times n}$ of dimension $(n-1)k$, while the normalization of scale places P on the unit sphere. Thus, $P(k, n)$ is a unit sphere of dimension $(n-1)k-1$.

Next, we consider the effect of orthogonal transformations on P. Let $O(k)$ be the group of $k \times k$ orthogonal matrices. If $U \in O(k)$ and $P \in P(k, n)$, then UP represents the same shape as P. The orbit of a pre-shape P under the action of the orthogonal group $O(k)$ is $[P] = \{UP : U \in O(k)\} \subset P(k, n)$. Kendall's shape space $\Sigma(k, n)$ is the orbit space of pre-shapes under the action of $O(k)$ [5]. Thus, a shape may be thought of as an orbit of pre-shapes. Given a positive definite $n \times n$ matrix M, we define the shape metric

$$\begin{aligned} d_M([P], [Q]) &= \min_{U,V\in O(k)} \|VP - UQ\|_M \\ &= \min_{U\in O(k)} \|P - UQ\|_M\,, \end{aligned} \tag{6}$$

where the last equality follows from the fact that orthogonal transformations preserve the M-norm. If M is the identity matrix, the metric coincides with the one used in classical Procrustes analysis. Since

$$\begin{aligned} \|P - UQ\|_M^2 &= \langle P - UQ, P - UQ\rangle_M \\ &= \|P\|_M^2 + \|Q\|_M^2 - 2\,\langle P, UQ\rangle_M\,, \end{aligned} \tag{7}$$

minimizing $\|P - UQ\|_M^2$ is the same as maximizing $\langle P, UQ\rangle_M = \langle PM, UQ\rangle$. This last problem is well known [5] and the solution may be expressed as follows. Using a singular value decomposition, write $PMQ^T = V_1 \Sigma V_2^T$, with $V_1, V_2 \in O(k)$ and Σ diagonal with nonnegative eigenvalues. Then, $\tilde{U} = V_1 V_2^T$ maximizes $\langle PM, UQ\rangle$. Hence, the M-distance is given by

$$d_M([P], [Q]) = \|P - V_1 V_2^T Q\|_M. \tag{8}$$

We conclude this section with an interpretation of the M-metric. Decompose M as $M = U\Lambda U^T$, with Λ diagonal and U orthogonal. The diagonal entries λ_i are the positive eigenvalues of M and the corresponding columns of U are the associated unit eigenvectors. Under the change of coordinates $Q = PU$, the inner product associated with M becomes

$$\begin{aligned} \left\langle P, P^\dagger\right\rangle_M &= \left\langle PM, P^\dagger\right\rangle = \left\langle PU\Lambda U^T, P^\dagger\right\rangle \\ &= \left\langle PU\Lambda, P^\dagger U\right\rangle = \left\langle Q, Q^\dagger\right\rangle_\Lambda\,. \end{aligned} \tag{9}$$

This means that if we replace the ith landmark p_i with q_i, the effect of the M-metric is simply to attribute the weight λ_i to q_i. We order the eigenvalues to satisfy $\lambda_i \geqslant \lambda_{i+1}$, so that the "importance" of q_i decreases as the index i increases.

3. Mean Shape

Given shapes $[P_i]$, $1 \leqslant i \leqslant m$, a sample Fréchet mean with respect to the M-metric is a shape $[P]$ such

that P minimizes the M-variance function

$$V_M(P) = \frac{1}{m-1}\sum_{i=1}^{m} \|P - U_i P_i\|_M^2 \,, \qquad (10)$$

where U_i optimally aligns P_i with P. We omit the details of the minimization of V_M on the pre-shape sphere because the strategy is similar to that of the mean-shape algorithm of [4]. Experiments demonstrated fast convergence if the samples form a fairly compact cluster in shape space. This is often the case in applications, for example, if the samples represent the same anatomical structure of different individuals such as the wings of several specimens of fruit flies.

4. Learning Shape Metrics

Given training data representing K different shape classes, our goal is to learn a shape metric that best discriminates them. In implementations, it may become difficult to enforce the positivity of the matrix M, so we use a logarithmic representation by writing $M = \exp A$, with A symmetric.

For each $1 \leqslant i \leqslant K$, let the training shapes in the ith class be represented by the pre-shapes P_{ij}, $1 \leqslant j \leqslant n_i$. Given a symmetric matrix A, let $\mu_i(A)$ be a pre-shape that represents the mean of the training shapes in the ith class with respect to $M = \exp(A)$. Define the total within-class scatter with respect to A as

$$S_W(A) = \sum_{i=1}^{K}\sum_{j=1}^{n_i} d_M^2([P_{ij}], [\mu_i]) \,. \qquad (11)$$

Similarly, letting μ be a pre-shape representative of the mean shape of the entire training set, define the total between-class scatter as

$$S_B(A) = \sum_{i=1}^{K} n_i d_M^2([\mu_i], [\mu]) \,. \qquad (12)$$

The proposed criterion for metric selection will be essentially based on the minimization of the ratio $S_W(A)/S_B(A)$. Note that scaling M does not affect the discriminative ability of the shape metric, which is reflected in the fact that $S_W(A)/S_B(A)$ remains the same. Scale can be normalized in several different ways. We adopt the normalization $\det M = 1$ because in logarithmic representation it becomes $\operatorname{trace} A = 0$, a linear constraint that can be easily enforced. Hence, our goal is to minimize $S_W(A)/S_B(A)$ on the subspace of trace zero symmetric matrices. However, this minimization problem may not be well posed because the ratio may keep decreasing as A moves off to infinity; that is, as one or more eigenvalues of $M = \exp A$ become large as others decay so as to keep the determinant unitary. Thus, we add a quadratic regularization term and propose the cost function

$$F(A) = \frac{S_W(A)}{S_B(A)} + \frac{\alpha}{2}\|A\|^2 \,, \qquad (13)$$

where $\alpha > 0$ is a constant.

We employ a gradient descent algorithm to estimate the minimum of F on the linear trace-zero subspace of symmetric matrices. A calculation of the gradient $\nabla F(A)$ yields

$$\begin{aligned} &\frac{1}{S_B}\sum_{i=1}^{K}\sum_{j=1}^{n_i}\int_0^1 e^{(1-s)A}(\widehat{P}_{ij} - \mu_i)^T(\widehat{P}_{ij} - \mu_i)e^{sA}ds \\ &-\frac{S_W}{S_B^2}\sum_{i=1}^{m} n_i \int_0^1 e^{(1-s)A}(\hat{\mu}_i - \mu)^T(\hat{\mu}_i - \mu)e^{sA}ds \\ &\qquad + \alpha A = \nabla F(A), \end{aligned} \qquad (14)$$

where $\hat{}$ indicates optimal orthogonal alignment. This expression shows that, at each step of the optimization algorithm, we need to calculate the total scatter functions $S_W(A)$ and $S_B(A)$, the mean of each cluster, the global mean and various generalized Procrustes alignments, all with respect to the current matrix $M = \exp A$.

5. Experimental Results

A preliminary experiment with phantom data illustrates that the matrix M which the algorithm produces is consistent with our intuition. We generated two classes of shapes, with 4 landmarks each, as follows. For all shapes, landmark 1 is represented by the same point, while landmark 2 for both classes is randomly sampled from a common Gaussian distribution. For landmarks 3 and 4, we used Gaussians with slightly different means, but same variance. Therefore, as indicated in Figure 2, we expect landmarks 3 and 4 to be the most informative for shape discrimination. We generated a total of 1,000 test samples (500 for each class), in addition to training samples. We carried out 4 experiments with T training samples for each class, with $T = 15, 25, 35, 45$. We use the algorithm to learn a matrix M and used the shape metric with the nearest neighbor classifier to discriminate the test samples. Table 1 compares the rate of correct classification with our matrix model and those obtained with the classical Procrustes metric. The symmetric matrix M, learned with $T = 15$, has eigenvalues $\lambda_1 = 0.73$, $\lambda_2 = 0.5$, $\lambda_3 = 0.46$ and $\lambda_4 = 0.086$. The

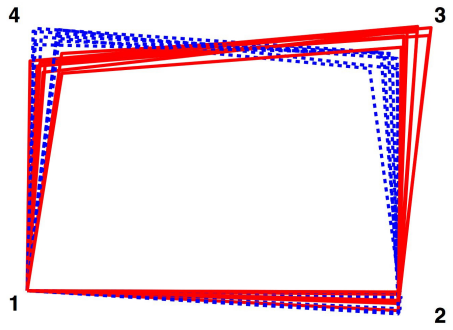

Figure 2. Phantom data with 4 landmarks.

T	Procrustes Metric	Matrix Model
15	92.1 %	96.6%
25	92.6%	97.7%
35	94.2%	98.7%
45	97.0%	99.3%

Table 1. Classification rate with T training samples from each class.

corresponding unit-length eigenvectors are the columns of the matrix

$$U = \begin{bmatrix} 0.63 & 0.35 & 0.50 & 0.48 \\ 0.15 & 0.17 & 0.50 & 0.84 \\ 0.02 & 0.86 & 0.50 & 0.12 \\ 0.76 & 0.35 & 0.50 & 0.23 \end{bmatrix} \quad (15)$$

The first eigenvector u_1 emphasizes the contributions of the landmarks p_1 and p_4, consistent with the facts that the first landmark should be well aligned and the 4th landmark is highly discriminatory. The eigenvector u_2 is dominated by the 3rd landmark, which also discriminates the classes well. The only eigenvector that emphasizes the 2nd landmark over the others is associated with the small eigenvalue λ_4.

We carried out a similar experiment with 12-landmark fly wing data acquired with the software WingMachine [3] for *Drosophila melanogaster* and *Drosophila mauritiana*. We used 233 *D. melanogaster* and 180 *D. mauritiana* test samples, respectively. The classification results obtained with T training samples for each species are reported in Table 2. To visualize the discrimination of the species achieved with the learned shape metric, Figure 3 shows the first landmark q_1 for several samples of *D. melanogaster* (green squares) and *D. mauritiana* (red circles), under the change of coordinates described in Section 2, using the matrix M learned with $T = 15$.

T	Procrustes Metric	Matrix Model
15	90 %	95%
25	91%	96%
40	92%	97%

Table 2. Discrimination accuracy of D. Melanogaster and D. Mauritiana.

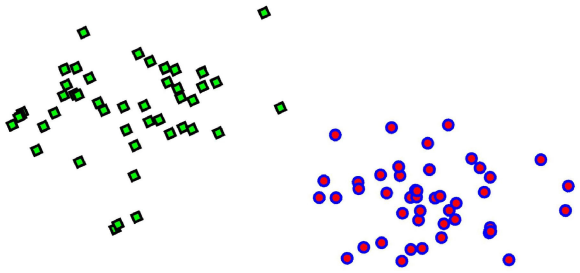

Figure 3. D. melanogaster (□) and D. mauritiana (○).

6. Summary and Discussion

We extended classical Procrustes shape analysis to a family of models that can emphasize particular combinations of landmarks. We developed a learning algorithm to select a model that is best suited to a shape classification and discrimination problem. Experiments were carried out with both phantom data and wing data for different species of fruit flies. Other applications and more extensive statistical analysis of shape using these models will be the subject of future investigation.

References

[1] I. L. Dryden and K. V. Mardia. *Statistical Shape Analysis*. John Wiley & Son, 1998.

[2] R. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7:179–188, 1936.

[3] D. Houle, J. Mezey, P. Galpern, and A. Carter. Automated measurement of Drosophila wings. *BMC Evolutionary Biology*, 3:25, 2003.

[4] S. Huckemann and H. Ziezold. Principal component analysis for Riemannian manifolds, with an application to triangular shape spaces. *Advances in Applied Probability*, 38:299–319, 2006.

[5] D. G. Kendall. Shape manifolds, Procrustean metrics and complex projective spaces. *Bulletin of London Mathematical Society*, 16:81–121, 1984.

[6] D. G. Kendall, D. Barden, T. K. Carne, and H. Le. *Shape and Shape Theory*. Wiley, 1999.