Sankhyā: The Indian Journal of Statistics
1993, Special Volume 55, Series A, Pt. 3, pp. 460–480
Dedicated to the memory of P. C. Mahalanobis

MULTIVARIATE SHAPE ANALYSIS

By I. L. DRYDEN and K.V.MARDIA
University of Leeds, U. K.

SUMMARY. This paper reviews the present state of the art in shape analysis, as well as introducing a few new ideas. The paper begins with a definition of size and then various coordinate systems for shape are considered - the QR decomposition and Bookstein shape variables in particular. Concepts of distance are reviewed including the Procrustes and Riemannian distances. Gaussian models for configurations are then considered, including perturbation models and principal components models. These models, offset in the size and shape or shape spaces, could then be used for size and shape analysis. However, this approach is complicated and we outline approximations based on tangent spaces - the preshape tangent space and the Procrustes tangent space. Standard multivariate analysis can then be performed, for example Hotelling's T^2 test and principal components analysis. We also consider a distribution free approach. Finally, we illustrate the methods with a practical 3D application in biology.

1. INTRODUCTION

Shape analysis is an area of study arising from a wide variety of applications, including in archaeology, biology, chemistry, geography, image analysis and medicine. It is only within the last decade that satisfactory techniques have emerged for landmark data.

The word "shape" is very commonly used in everyday language, usually referring to the appearance of an object. The definition of shape that we consider is intuitive: **shape** is all the geometrical information that remains when location, scale and rotational effects are removed (D.G. Kendall (1984)). So, an object's shape is invariant under the similarity transformations of translation, scaling and rotation. For example, the shape of a human skull consists of all the geometrical properties of the skull that are unchanged when it is translated, rescaled or rotated in an arbitrary coordinate system, i.e., it is invariant under 'pose'. Two objects have the same shape if we can translate, rescale and rotate them to each other so that they match exactly, i.e., if the objects are similar.

The classical way to analyse shape is to locate a finite number of landmarks on each specimen. A **landmark** is a point of correspondence on each object that matches between and within populations. There are three types of landmarks in most applications: biological, mathematical and pseudo- landmarks.

A **biological landmark** is a point assigned by an expert that corresponds in some biologically meaningful way, eg. the pupil of an eye or the meeting of two sutures on a skull. Biological landmarks designate parts of an organism that correspond in terms of biological function and these parts are called homologous (Jardine and Jardine (1967)). **Mathematical landmarks** are points located on an object according to some geometrical property of the figure, eg.

at a point of high curvature or at an extreme point. The use of mathematical landmarks is particularly useful in automatic recognition and analysis, and for example an algorithm for locating landmarks semi-automatically on the outline of a mouse vertebra as described in Mardia (1989). **Pseudo-landmalrks** are points constructed in between biological or mathematical landmarks. For example, if many equally spaced points are taken on the outline of a bone between biological landmarks then these are classed as pseudo-landmarks.

Our notation will be that there are k landmarks in m dimensions, where in practice we usually have $k \geq 3$ and $m = 2$ or $m = 3$. Point configurations can be labelled or unlabelled. Labelled configurations have each landmark assigned a name or number and the landmark with say label 1 on one specimen corresponds in some meaningful way with landmark 1 on another specimen. The methods considered in this paper are for labelled configurations.

The sprit of this weork follows Mahalanobis in being motivated by practical examples. It is an honour to contribute an article in this volume on this new and challenging area of shape analysis.

2. Size and Shape Coordinates

2.1 *Size.* Before considering shape we should define what we mean by size, so that it can be removed from a configuration. We consider precisely the same notion of size as that defined by Mosimann (1970). Consider p positive length measurements $l = (l_1, \cdots, l_p)', p \geq 2$, which are distances between landmarks or pseudo-landmarks. A standard size variable is any real valued function $g(l)$ such that $g(al) = ag(l)$ for positive $a, l_1, \cdots, l_p$. If X is a $k \times m$ matrix of k landmarks in m dimensions, then a suitable size measurement could be the centroid size

$$S(X) = \| CX \| = \sqrt{\sum_{i=1}^{k} \sum_{j=1}^{m} (X_{ij} - \bar{X}_j)^2} \tag{1}$$

where $\bar{X}_j = \frac{1}{k} \sum_{i=1}^{k} X_{ij}$,

$$C = I_k - \frac{1}{k} 1_k 1'_k,$$

and $\| X \| = tr(X'X)^{\frac{1}{2}}$ is the Euclidean norm, I_k is the $k \times k$ identity matrix and 1_k is the $k \times 1$ vector of ones. Note that C is the centering matrix. The centroid size is the most commonly used in geometrical shape analysis (e.g., Bookstein (1986), Kendall (1984), (1989), Goodall (1991), Dryden and Mardia (1992)). Geometrically $S(X)$ is the square root of the sum of squared Euclidean distances from each landmark to the centroid.

An alternative could be to use the 'baseline size', i.e. the length between landmarks 1 and 2 (Bookstein, (1986)).

$$D = \| (X)_2 - (X)_1 \| \tag{2}$$

where $(X)_i$ is the ith row of $X, i = 1, \cdots, k$. Other alternative size measures include the square root of area for planar configurations or the cube root of volume for 3D configurations.

Mosimann (1970) defines the shape vector to be $l/g(l)$ and this differs fundamentally from our approach to shape. In our formalism shape is not constructed from positive lengths but directly from the Cartesian coordinates after removing the similarity transformations.

2.2 *Shape : QR decomposition.* Let us consider k landmarks in m dimensions for general m with $k > m$. Given a $k \times m$ matrix of landmark coordinates X, location can be removed by premultiplying X by a suitable matrix H, where H is the Helmert matrix without the first row, defined by having its first row of elements equal to $1/\sqrt{k}$ and the remaining rows are orthogonal to the first row. Explicitly, the jth row of H is given by,

$$(d_j, \cdots, d_j, -jd_j, 0, \cdots, 0),$$

where $d_j = -\{j(j+1)\}^{-\frac{1}{2}}$ is repeated j times and $j = 1, \cdots, k-1$.
A QR decomposition of Y then leads to

$$Y = T\Gamma, \qquad \Gamma \varepsilon SO(m),$$

where the $(k-1) \times m$ lower triangular matrix T contains size-&-shape coordinates.

Note that T is invariant under the original location and rotation of the configuration. The matrix T has zero entries above the leading diagonal and therefore $(k-1)m - m(m-1)/2$ size-&-shape coordinates.

To obtain the shape coordinates we divide by the scale : $W = T/\parallel T \parallel$ and so there are $(k-1-m)m + (m-1)(m-2)/2$ shape coordinates. Note that $S(X) = \parallel T \parallel = \parallel Y \parallel$ is the centroid size. This coordinate system was introduced in Goodall and Mardia (1992).

A slight variant of this coordinate system includes using a different translation matrix from H such as B where the jth row of B is $(-1, 1, 0, \cdots, 0)$ if $j = 1$ and

$$(-\frac{1}{2}, \frac{1}{2}, 0, \cdots, 0, 1, 0, \cdots, 0) \qquad 1 < j \leq k-1,$$

and the 1 is in the $(j+1)$th column (see Dryden and Mardia (1991)). Another alternative is to use a different scale such as the $(1,1)$th element of T. Note that we could have removed the similarity transformations in a different order. After translating we could have rescaled to give,

$$Z = HX/\parallel HX \parallel$$

and this is known as the **preshape** of X and we write $Z \varepsilon S_m^k$. The preshape space $S_m^k = S^{(k-m)m-1}$ is thus a hypersphere of unit radius in $(k-1)m$ dimensions. The preshape space is invariant under the translation and scale but not the rotation of the original configuration. Finally rotation is removed by a QR decomposition to give the shape coordinates W.

The QR decomposition coordinates are closely related to Bookstein (1986) shape coordinates. Let $(x_j, y_j)'$, $j = 1, \cdots, k$ be $k \geq 3$ landmarks in a plane $(m = 2)$. Bookstein (1986) suggested removing the similarity transformations by translating rotating, and rescaling such that landmarks 1 and 2 are sent to a fixed position. If landmark 1 is sent to $(-\frac{1}{2}, 0)$ and landmark 2 is sent to $(\frac{1}{2}, 0)$ then Bookstein's shape variables are the coordinates of the remaining $k-2$ coordinates after these operations $(u_j, v_j)'$, $j = 3, \cdots, k$ and so

$$u_j = \{(x_2 - x_1)(x_j - x_1) + (y_2 - y_1)(y_j - y_1)\}/d^2 - \frac{1}{2}, \tag{3}$$

$$v_j = \{(x_2 - x_1)(y_j - y_1) - (y_2 - y_1)(x_j - x_1)\}/d^2, \tag{4}$$

where $j = 3, \cdots, k$ and $d^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2 > 0$. Note that $-\infty < u_j, v_j < \infty$. Bookstein (1986) actually sends the baseline to $(0,0)$ and $(1,0)$ and so does not have the $-\frac{1}{2}$ in equation (3). The coordinates have been used widely in shape analysis for planar data, e.g., Bookstein and Sampson(1990) and Mardia and Dryden (1989).

For the 3D case, consider k landmarks $X_i \in \mathcal{R}^3, i = 1, \cdots, k$. The number of shape variables is $3k - 7$ since we have $3k$ landmark coordinates but must remove 3 location, 1 scale and 3 rotation parameters. The Bookstein shape variables can be taken as

$$u_j = \frac{1}{\| X_2 - X_1 \|} A(X_1, X_2, X_3)(X_j - X_1) - (\frac{1}{2}, 0, 0)', \qquad j = 4, \cdots, k$$

where $A(X_1, X_2, X_3)$ is a 3×3 rotation matrix and $X_1 \to (-\frac{1}{2}, 0, 0)', X_2 \to (\frac{1}{2}, 0, 0)'$ and $X_3 \to (u_3, v_3, 0)', v_3 \geq 0$.

Geometrically the figure has been translated, rotated and rescaled so that the baseline (landmarks 1 to 2) is of unit length, has midpoint at the origin and is rotated to lie along the x-axis. The figure is further rotated so that landmark 3 lies in the x-y plane, with $y \geq 0$. The remaining non-fixed coordinates are the shape variables. It is necessary to restrict $y \geq 0$ in order to fully remove rotation; if $y < 0$ then one rotates the configuration by π radians around $x = 0$ to obtain $y > 0$. If $y = 0$ then there is a singularity, but in practice this will not occur. The explicit form for $A(X_1, X_2, X_3)$ is somewhat complicated.

In both the $m = 2$ and $m = 3$ cases it can be shown that the coordinates are simply related to a QR decomposition where the size measure used is the baseline length of equation (2).

2.3 *Shape distances.* A concept of distance between two shapes is required to fully define the non-Euclidean shape space. Given two figures of unit size X_1, X_2 the Procrustes distance $d(X_1, X_2)$ (e.g., Goodall (1991), Ziezold (1993)) is defined by

$$d^2(X_1, X_2) = \inf_{\gamma, \Gamma} \| X_2 - X_1\Gamma - 1_k\gamma' \|^2, \tag{5}$$

where $\Gamma \varepsilon SO(m)$ (rotation matrix) and $\gamma \in \mathcal{R}^m$ (location parameter). Alternatively, the Riemannian distance $\rho(X_1, X_2)$ of Kendall (1984) is

$$\rho(X_1, X_2) = 2 \arcsin (d/2)$$

Geometrically, we can think of d as the closest chordal distance on the preshape sphere between the two rotated versions of X_1 and X_2. Alternatively, ρ is the closest great circle distance. For $m > 2$, there are singularities in the shape space when the k points lie in an $m-2$ dimensional subspace (see Kendall (1989)). Here we are assuming that our shapes are away from any such singularities. For $m = 2$ there is no such problem. It can be shown (Kendall (1984)) that

$$\rho(X_1, X_2) = \arccos \left(\sup_{\Gamma \varepsilon SO(m)} \operatorname{tr} \left(\frac{X_1' H' H X_2 \Gamma}{\| HX_1 \| \| HX_2 \|} \right) \right)$$

and it follows that

$$\rho(X_1, X_2) = \arccos \left(\operatorname{tr} \frac{(X_2' H' H X_1 X_1' H' H X_2)^{\frac{1}{2}}}{\| HX_1 \| \| HX_2 \|} - 2\alpha s \right),$$

where α is the smallest eigenvalue of

$$(X_2' H' H X_1 X_1' H' H X_2)^{\frac{1}{2}} / (\| HX_1 \| \| HX_2 \|),$$

and

$$s = \left\{ \begin{array}{ll} 0 & \text{if } \det (X_2' H' H X_1) \geq 0 \\ 1 & \text{if } \det (X_2' H' H X_1) < 0 \end{array} \right\}$$

The Riemannian distance is particularly simple to calculate for $m = 2$ and we use complex notation. Consider configurations $z = (z_1, \cdots, z_k)', w = (w_1, \cdots, w_k)'$ with centroids z_c, w_c. Then the Riemannian distance between the shapes of z and w is given by,

$$\cos\rho = \frac{|\sum (z_i - z_c)(\bar{w}_i - \bar{w}_c|}{\sqrt{\sum |z_i - z_c|^2} \sqrt{\sum |w_i - w_c|^2}}$$

where $\bar{w}$ means the complex conjugate of w. For the triangle case the Riemannian distance between shapes with Bookstein shape variables (U, V) and (θ, ϕ) is given by

$$cos2\rho = 1 - \frac{\frac{2}{3}\{(u-\theta)^2 + (v-\phi)^2\}}{(\frac{1}{2} + \frac{2}{3}u^2 + \frac{2}{3}v^2)(\frac{1}{2} + \frac{2}{3}\theta^2 + \frac{2}{3}\phi^2)} \tag{6}$$

Discussion of these distances and other deep geometrical properties of the shape space is given by Kendall (1984, 1989) and Kendall and Le (1993).
The above Procrustes distance given by equation (5) involves translating and rotating one configuration to another so as to minimize the sum of square distances. If scaling were also introduced and X_1, X_2 are not necessarily of unit size then the ordinary Procrustes sum of squares is given by

$$OSS(X_1, X_2) = \inf_{\beta,\gamma,\Gamma} \| X_2 - \beta X_1 \Gamma - 1_k \gamma' \|^2,$$

where $\Gamma \in SO(m)$ (rotation),$\gamma \in \mathcal{R}^m$ (translation) and $\beta > 0$ (scale). This sum of squares is a measure of discrepancey, although note that $OSS(X_1, X_2) \neq OSS(X_2, X_1)$. However, if X_1 and X_2 are of unit size then

$$\sqrt{OSS(X_1, X_2)} \tag{7}$$

is a measure of distance between X_1, X_2.

2.4 *Least squares.* If a random sample $X_1, \cdots, X_l$ of $l \geq 2$ figures is available we often need to obtain an estimate of the population mean shape. A least squares approach is that of Generalized Procrustes Analysis (GPA). In GPA each figure is translated, rescaled and rotated relatively to each other so as to minimize the sum of squares (Gower, (1975), Goodall (1991))

$$\sum_{i<j} \| (\beta_i X_i \Gamma_i + 1_k \gamma_i') - (\beta_j X_j \Gamma_j + 1_k \gamma_j') \|^2$$

$$\text{subject to } \sum_{i=1}^{l} S^2(\beta_i \Gamma_i X_i + 1_k \gamma_i') = \sum_{i=1}^{l} S^2(X_i)$$

where $\Gamma_i \in SO(m)$, $\beta_i > 0$, $\|X\| = (\text{tr}(X'X))^{1/2}$ and $S(X)$ is the centroid size given by (1). The Procustes superimposition of the X_i is given by

$$X_i^p = \hat{\beta}_i X_i \hat{\Gamma}_i + 1_k \hat{\gamma}_i', \; i = 1, \ldots, l$$

where $\Gamma_i \in SO(m)$ (rotation matrix), $\hat{\beta}_i > 0$ (scale paparmeter), $\hat{\gamma}_i'$ (location parameter), $i = 1, \cdots, l$. An algorithm to estimate the transformation parameters $(\gamma_i, \beta_i, \Gamma_i)$ was given by Gower (1975) and modified by Ten Berge (1977). The Procrustes estimate of population mean shape is given by,

$$\hat{\mu} = \frac{1}{l}\sum_{i=1}^{l} X_i^P, \tag{8}$$

An alternative approach could be to use model based estimation.

3. SHAPE MODELS

3.1 *Perturbation models* Goodall and Mardia (1993) and Dryden and Mardia (1991, 1992) consider the models

$$x = vec(X) = \mu + \varepsilon \tag{9}$$

where $\varepsilon \sim N(0, \sum)$. For $m = 2$ Dryden and Mardia (1991, 1992) have given the joint marginal size-and-shape and marginal shape distributions, by integrating out the similarity transformations. The distributions are particularly simple if $\sum$ is of complex normal form. We consider the particular case here when $k = 3$ and $m + 2$ and $\sum = \sigma^2 I$. The joint p.d.f. of Bookstein's shape variables (U, V) is given by (Mardia and Dryden, 1989a,b),

$$\frac{1}{3\pi(\frac{1}{2} + \frac{2}{3}u^2 + \frac{2}{3}v^2)^2}\{1 + \kappa(1 + cos2\rho)\}exp\{-\kappa(1 - cos2\rho)\},$$

where

$$\kappa = S^2(\mu) / (4\sigma^2),$$

is a concentration parameter and ρ is the Riemannian distance between the observed shape (u, v) and the population shape (θ, ϕ) of μ, given in equation (6). This has been referred to in the literature as the 'Mardia-Dryden' distribution (Kendall, (1991)).

For $m > 2$ Goodall and Mardia (1993) have investigated size and shape distributions for factored covariances, although the densities are not explicitly known if μ is of full rank.

These models can be used for large sample likelihood based inference and implementation is fairly straightforward for $m = 2$ providing simple covariance structures are adopted. Inference for general $\sum$ or $m \geq 3$ is complicated.

Goodall (1991) considers an alternative approach with factor covariance structures. The Procrustes approach that he considers (see brief details in Section 2.4) involves estimation of the similarity transformations rather than integrating them out.

Kent (1994) approaches modelling through the complex Bingham distribution on the preshape sphere with p.d.f. given by,

$$f(z) = c(A)exp(trz\bar{z}'A), \tag{10}$$

and this fits in well with the Procrustes approach. It is suitable for modelling shape becuase the distribution is rotationally invariant, $f(e^{i\theta}z) = f(z)$.

3.2 *Principal components models.* Cootes et al.(1992) consider the point distribution model which is a principal component model for shape. Let $x = vec(X)$; then we formulate their principal component model for configurations as

$$x = \mu + \sum_{i=1}^{p} y_i \gamma_i + \varepsilon, \tag{11}$$

where

$$y_i \sim N(0, \lambda_i), \qquad \varepsilon \sim N(0, \sigma^2),$$

independently and

$$\mu' \gamma_i = 0, \quad \gamma_i' \gamma_i = 1, \gamma_i' \gamma_j = 0, \quad (i \neq j),$$

and $\lambda_1 \geq \lambda_2 \ldots, \geq \lambda_p$. Here $p \leq k$ and p is preferably taken to be quite small.

Note that, this approach allows us to give a model for flexible varying shapes, with often interpretable principal components. The interpretation of each component can be seen by varying y_i in equation (11) while keeping the other $y_j, j \neq i$, equal to zero. In practice the population parameters must be estimated from a random sample. Let $\hat{\mu}$ and $\hat{\gamma}_i$ be the estimates of μ and γ_i. Then $x = \hat{\mu} + y_i \hat{\gamma}_i$ for different y_i (but fixed i) captures different apsects of shape variation, e.g. relative length, relative width, bending, movement of a subset of landmarks relative to the rest and so on. A practical example will be considered in Section 6.

Of course the principal component model is a special case of the model of equation (9) and the covariance matrix is

$$\Sigma = \sigma^2 I_{mk} + \sum_{i=1}^{p} \lambda_i \gamma_i \gamma_i' \tag{12}$$

Hence, details of the size and shape distributions are available for $m = 2$ in Dryden and Mardia (1991, 1992) and what is known for $m \geq 3$ is given in Goodall and Mardia (1993). Thereafter, inference for large samples could proceed using likelihood based methods.

However, this is complicated in general (particularly for $m \geq 3$) and so we consider various approximations for the situation where the variability in the dataset is small.

4. Approximations

4.1 *Local shape coordinates : tangent spaces.* If variations in a dataset are small then one could consider a local coordinate system. For example one

could choose a linearized space, a tangent plane to the shape space, after Kent (1994).

Let us first look at the $m = 2$ case with complex notation. Consider complex landmarks $z = (z_1, \cdots, z_k)'$ with $w = (w_1, \cdots, w_{k-1})' = Hz/ \parallel Hz \parallel$. Further let γ be a complex pole on the complex preshape sphere. which can for example be chosen as the 'population mean shape'. In practice this can be estimated from the data using a suitable average shape such as the dominant eigenvector of the complex sums of squares and products matrix of $(w_1, \cdots, w_{k-1})'$, which is the maximum likelihood estimate of shape under the model (10). This estimator turns out to be identical to the Procrustes with scaling mean shape, given in equation (8) (Kent (1994)).

The configuration is then rotated to be as close as possible to the pole and then projected onto the tangent plane at γ. Let $\theta = \arg \bar{\gamma}' w$ and then

$$v = \left[I - \gamma\hat{\gamma}'\right] we^{-i\theta}, v \in C^{k-1} \tag{13}$$

are the Kendall tangent plane coordinates for shape (Kent (1994)). Note that $v'\gamma = 0$ and so the complex constraint means we can regard the tangent space as a real subspaces of $\Re^{2k-2}$ of dimension $2k - 4$. Note the inverse projection from v to w is given by,

$$w = \left[(1 - \bar{v}'v)^{\frac{1}{2}}\gamma + v\right] e^{i\theta}. \tag{14}$$

The Kendall tangent plane coordinates can be extended simply into $m \geq 3$ dimensions. Again a pole $\gamma(k - 1 \times m)$ is chosen on the preshape sphere, for example the population mean shape. In practice we again replace this with its sample estimate (such as the Procrustes mean defined in equation (8)). Given a preshape w this is rotated to the pole to be as close as possible (according to the distance used) and we write this as $w\Gamma, \Gamma \in SO(m)$. This is then projected onto a tangent plane at γ,

$$v = (I_{km-m} - vec(\gamma)vec(\gamma)')vec(w\Gamma). \tag{15}$$

Note that $v'vec(\gamma) = 0$ and the tangent space is a $(k-1)m - m(m-1)/2 - 1$ dimensional subspace of $\mathcal{R}^{mk-m}$. The inverse projection is given by.

$$w = mat\left[(1 - v'v)^{\frac{1}{2}}vec(\gamma) + v\right]\Gamma' \tag{16}$$

where $mat()$ is the reverse of vectorizing. Note the equavalent roles of the rotations θ and Γ in the above equation (13-16). Also, note that for $m = 2$ Kent's complex coordinates of equation (13) are not the same as those in equation (15).

An alternative tangent space is the Procrustes tangent space (Goodall(1991)). Again a pole is chosen and this is taken to be the population mean shape γ. Given a new configuratioin X the Procrustes tangent coordinates are obtained by translating, rotating and rescaling X by Procrustes analysis to be as close

as possible to the pole. If the transformed coordinates are X^P, the Procrustes tangent coordinates are given by,

$$X^P - \gamma$$

In practice the pole γ is chosen to be the Procrustes mean shape $\hat{\mu}$ given by (8).

4.2 *Inference in tangent spaces.* Kent (1994) suggest the use of the tangent space coordinates if the data are concentrated and then performs standard multivariate analysis (e.g. Mardia, Kent and Bibby, 1979) in this linear space. This is an approximation to inference using the general model of equation (9). Kent's procedure is easily extended to $m \geq 3$ using the above tangentspace coordinates of equation (15). For example consider the case where we wish to test for mean shape difference between two independent populations. Given independent random samples $X_1, \cdots, X_{n_1}, Y_1, \cdots, Y_{n_2}$ we consider the pole as the overall Procrustes mean obtained by GPA on all $n_1 + n_2$ observations. A test for differences in covariances is carried out (e.g., Box's M test). If there is not a significant difference in the covariances then a Hotelling's T^2 test is carried out on the tangent space coordinates. Note that the space is singular and so a convenient approach is to retain $(k-1)m - m(m-1)/2 - 1$ of the coordinates in the analysis.

4.3 *PCA in tangent spaces.* In order to estimate the parameters of the principal components model with covariance matrix (12) instead of exact marginal MLE mentioned in Section 3.2 we could consider the Procrustes tangent space, providing variations are small.

Given l independent configurations $X_1, \cdots, X_l$, estimates of the above parameters can be obtained by first registering the shapes by GPA and then performing a principal component analysis. Cootes et al. (1992) use a type of weighted Procrustes analysis to obtain registered figures $X_1^P, \cdots, X_l^P$. The estimate of μ is taken to be the Procrustes mean of equation (8) and $\hat{\gamma}_i, i = 1, \cdots, min(l, k)$ are the eigenvectors of

$$\frac{1}{l}\sum_{i=1}^{l}(vec(X_i) - \hat{\mu})(vec(X_i) - \hat{\mu})'$$

with corresponding decreasing eigenvalues $\hat{\lambda}_i$. Given a training set of data this approach allows us to give a model for flexible varying shapes, as described in Section 3.2. Note that the Cootes et al (1992) approach is effectively PCA in the Procrustes tangent space.

Alternatively Kent (1994) has suggested PCA in the tangent space defined in equation (10), which is simply extended here to the situation of equation (15) for general m. This involves a model specified in the tangent space coordinates

and is thus different from the Cootes et al. (1992) approach. Consider l shapes with tangent space coordinates (with pole μ) given by $v_1, \cdots, v_l$. The principal component model for shape could possibly be formulated as

$$v = \mu \sum_{i=1}^{p} z_i \xi_i + \varepsilon,$$

where

$$z_i \sim N(0, \nu_i),\ \varepsilon \sim N(0, \sigma^2)$$

independently, with

$$\mu' \xi^i = 0, \xi_i' \xi_i = 1, \xi_i' \xi_j = 0, (i \neq j),$$

and $\nu_1 \geq \nu_2 \geq \cdots \geq \nu_p$. Here μ is the pole and estimates of the other parameters are again obtained by PCA, but using the sample covariance of the tangent space coordinates v.

In order to visualize shapes generated from the model it is most convenient to project from the tangent space coordinates back into configuration space. Either one could use the inverse transformation of equation (16) or possibly project back into Bookstein coordinates in equations (3), (4).

Note that a third approach to principal components might have been to consider PCA of QR decomposition shape coordinates. However, in this case, there can be problems with spurious correlations. In Bookstein's coordinates (related to QR) registration on a particular edge induces correlations into the remaining vertices. Hence, even in the situation where all landmarks are independent and isotropic, the PCA of Bookstein's coordinates will in general imply that there are strong directions of shape variability. This is clearly undersirable for interpretation. Note that the first two methods (using Procrustes rotations) do not suffer from this problem.

5. Alternatives : distribution free approach

Ziezold (1993) considers a distribution free test for shape analysis. His approach is to use generalized Procrustes analysis (section 2.4) without the scaling step to obtain mean shapes.

We outline his two independent sample test. Given two groups $X_1, \cdots, X_{n_1}$, and $Y_1, \cdots, Y_{n_2}$ of figures, the mean shapes $\hat{\mu}_x, \hat{\mu}_y$ are obtained from GPA (without scaling). Then the Mann-Whitney U-statistics U_1 of

$$dist(X_1, \hat{\mu}_x), \cdots, dist(X_{n_1}, \hat{\mu}_x), dist(Y_1, \hat{\mu}_x), \cdots, dist(Y_{n_2}, \hat{\mu}_x)$$

and U_2 of

$$dist(X_1, \hat{\mu}_y), \cdots, dist(X_{n_1}, \hat{\mu}_y), dist(Y_1, \hat{\mu}_y), \cdots, dist(Y_{n_2}, \hat{\mu}_y)$$

are found , where $dist()$ is a suitable shape distance from Section 2.3. If R_x is the sum of the ranks of the above distances involving the X_i, then the Mann-Whitney U-statistics is $min\{R_x - \frac{1}{2}n_1(n_1+1), n_1n_2 - R_x + \frac{1}{2}n_1(n_1+1)\}$. Ziezold (1993) uses the Procrustes distance of equation (5) but other sensible choices from Section 2.3 could be considered. We then wish to examine whether U_1 and U_2 are small. There are four scenarios:

1. U_1 and U_2 small: $\mu_x \neq \mu_y$
2. U_1 small, U_2 not small: $\mu_x \approx \mu_y$, var(shape(X)) < var(shape(Y)).
3. U_2 small, U_1 not small: $\mu_x \approx \mu_y$, var(shape(X)) > var(shape(Y)).
4. U_1 not small, U_2 not small: $\mu_x \approx \mu_y$, var(shape(X)) $\approx$ var(shape(Y)).

Thus if both U_1 and U_2 are small this indicates 'unequal population mean shapes'. In order to assess 'small' one could use Monte Carlo sampling . For example one could generate a large number N samples without replacement from the dataset in groups of size n_1 and n_2. Then $U_{1i}, U_{2i}.i = 1, \cdots, N$ are worked out and the observed values of U_1, U_2 are included in the ordered sequences. Finding

$$r_1 = rank(U_1),\ r_2 = rank(U_2) \tag{17}$$

gives an indication of whether the observed values are small or not.

It is clear that many other multivariate distribution - free methods can be simply adapted to shape analysis by replacing the Euclidean norm with a shape distance.

6. Analysis of data

In an investigation into sex differences in the crania of a species of macaque *Macaca fascicularis* (a type of monkey) random samples of 9 male and 9 female skulls were obtained. On one side of each skull $k = 7$ biological landmarks were located and the three dimensional coordinates of each point were recorded. The names of the landmarks are:

1. *prosthion,* 2. *nasion,* 3. *bregma,*
4. *opisthion,* 5. *asterion,* 6. *interfrontomalare,*
7. *midpoint of zyg/temp suture,*

The locations of the landmarks on an artists' impression of the x-y (top left), x-z (bottom left) and y-z (top right) projections of a skull are given in Figure 1. Definitions of these landmarks can be found in Bass (1971). We wish to test whether the sexes are different in mean shape, although note that the sample sizes are particularly small in this data set.

The configurations are superimposed by generalized Procrustes analysis of Section 2.4 and projections of the superimposed males and females are given in Figs. 2, 3. The projections used are the same as those in Figure 1. We see that

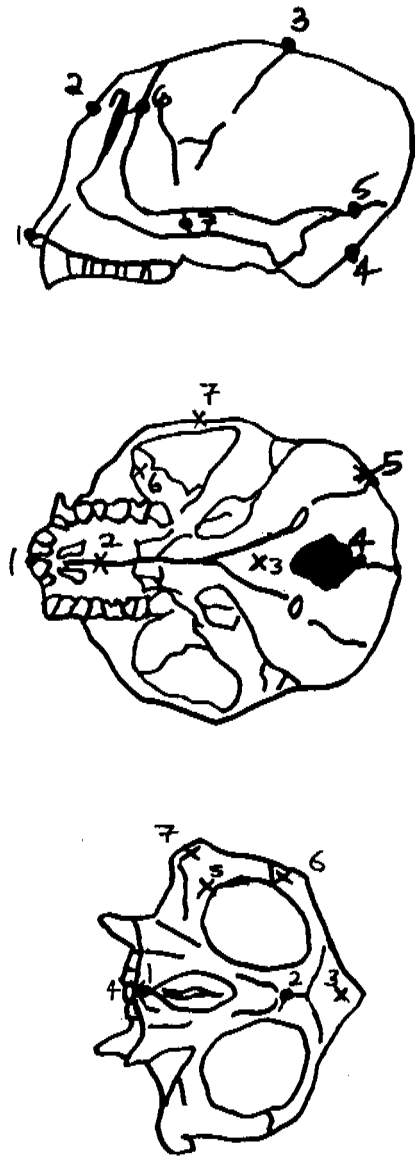

Fig. 1. An 'artists's impression of a skull, indicating the rough position of the landmarks, in the a) x-y, b) x-z and c) y-z projections. The landmarks (• if on show, × if hidden from view) are named in the text.

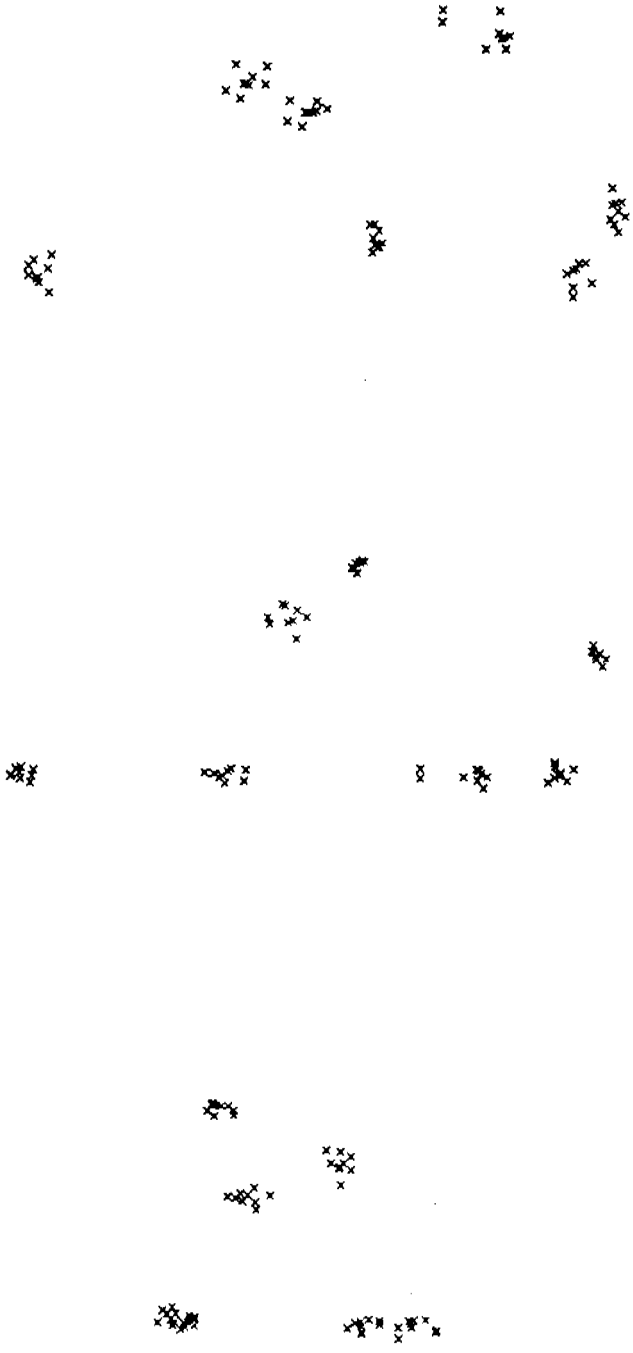

Fig. 2. The 3D male skull data superimposed by general Procustes analysis. The three figures show projections into the a) x-y, b) x-z and c) y-z planes.

Fig. 3. The female skull data superimposed by general Procustes analysis. The three figures show projections into the a) x-y, b) x-z and c) y-z planes. The same scale is used in Fig. 2.

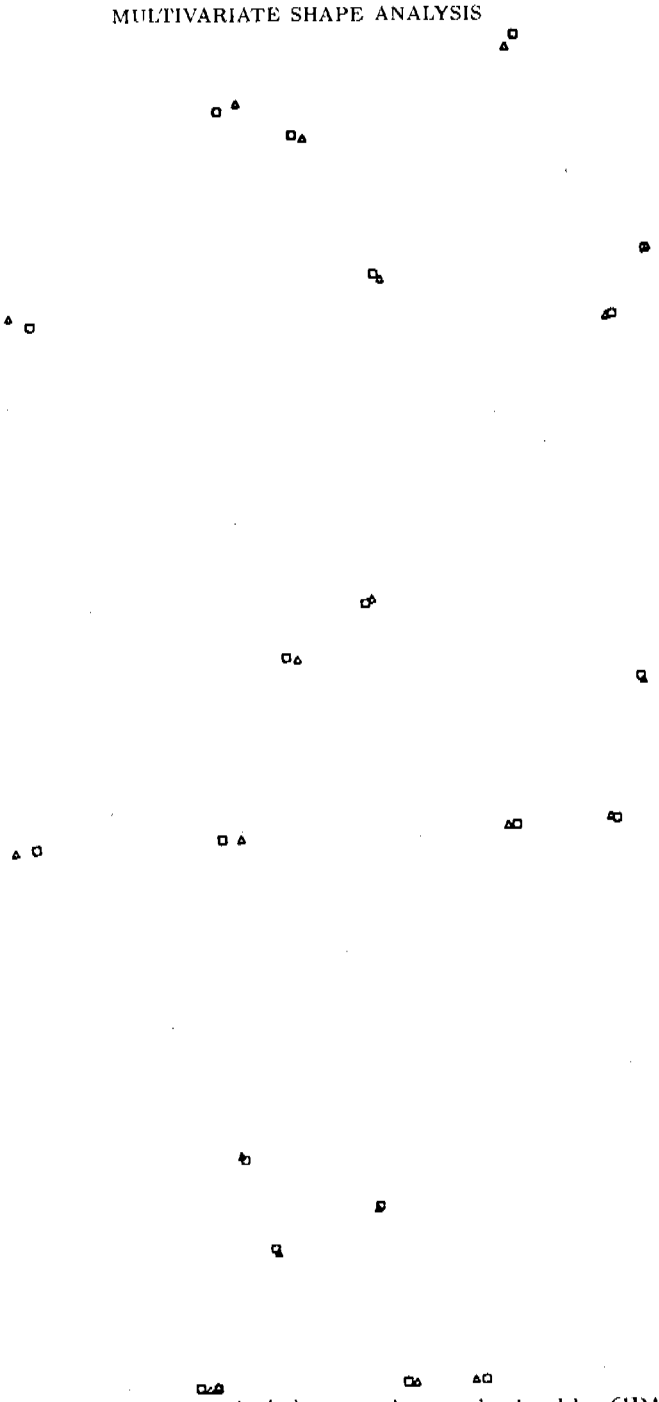

Fig. 4. The male (•) and female (+) mean shapes obtained by GPA and scaled to unit size, projected into the x-y (above left), x-z (below left), y-z (above right) planes.

there is fairly low variability in the samples, particularly in the female group. The mean Procrustes shape estimates are given in Fig. 4 and the male and female mean shapes are fairly similar. As variations seem to be fairly small we use the tangent plane coordinates of equation (15) and then work with standard multivariate analysis. The pole is taken to be overall pooled Procrustes mean and the dimension of this space is $14\{7\times 3-3(location)-3(rotation)-1(scale)\}$. A test for the equality of the population covariance matrices of the sexes would be appropriate (such as Box's M stest) but sample sizes are very small here. Assuming equal covariances, Mahalanobis D^2 between the two groups is 28.90 and so there is no evidence for a shape difference between the sexes using a Hotelling's T^2 test - $\{P(D^2 > 28.90) = 0.37\}$. A similar analysis of two groups of second Thoracic mouse vertebrae data in 2 dimensions was carried out by Kent and Mardia (1992).

However, the assumption of equal shape variances is doubtful here as seen below, and so a permutation test was also carried out. The data were randomly split into two groups and D_i^2 calculated for the ith permutation, $i = 1, \cdots, 99$. The rank r of the observed D^2 in the ordered sequence of D_i^2 was 61 and so the Monte-Carlo p-value is $1 - \frac{r}{100} = 0.39$. Hence, again there is no evidence for a shape difference between the sexes.

The distribution free test of Ziezold (1993) from Section 5 was also performed using the distance of equation (7). This resulted in values of $r_1 = 1$ (female) and $r_2 = 86$ (male). This situation would arise for the case when the mean shapes are not different but the variance in male shape is larger than the female shape variance.

Note that a measure of sample shape standard deviation is

$$SDS = \sqrt{\sum_{i=1}^{l} dist^2(X_i, \hat{\mu})/l},$$

and in this case we see $SDS(female) = 0.015$ and $SDS(male) = 0.024$. Thus the sample male shape variability is indeed larger than for the female group. Note that $dist(\hat{\mu}_{female}, \hat{\mu}_{male}) = 0.054$ and so it is clear that any sensible confidence regions for μ_{female}, and μ_{male} would overlap, and hence there is no evidence for a shape difference between the sexes.

For illustration we consider a principal component analysis of the male shape data. In Fig. 5 we see the first principal component of the tangent plane coordinates with the male Procrustes mean as the pole. The figures are of mean shape $\hat{\mu}$ ($\bullet$) with displacement vectors

$$\hat{\mu} + z_{\alpha/2}\hat{\lambda}_1^{\frac{1}{2}} l_1 \tag{18}$$

where l_1 is the dominant eigenvector of the sample covariance matrix of the tangent plane coordinates with associated eigenvalue $\hat{\mu}_1$ and $z_{\alpha/2}$ is the $100\alpha/2$ percentage point of the standard normal distribution. Note that the mean shape and the vectors have been calculated in the tangent plane and then projected

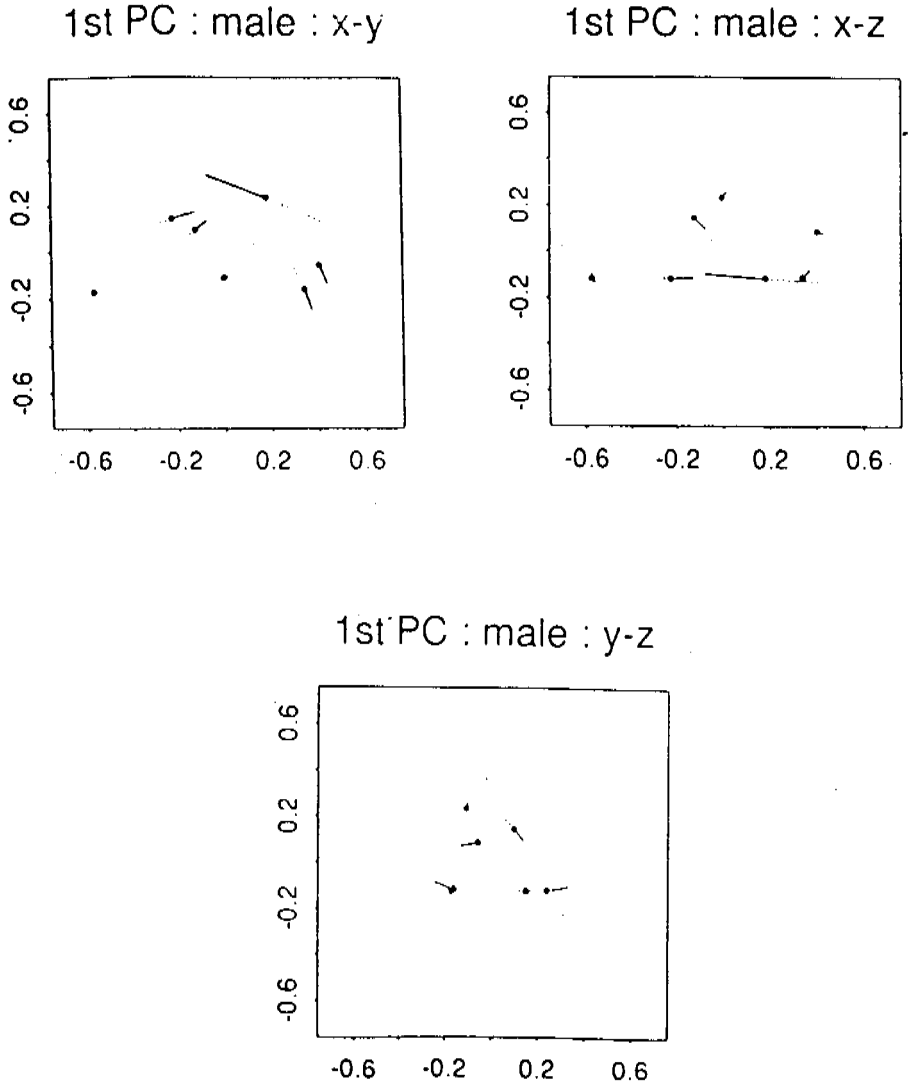

Fig. 5. The first principal component for the males. The figures are projections of the mean shape $\hat{\mu}(\bullet)$ with displacement vectors from +/- 3 standard deviations form the mean (details in the text). The vectors show $z_{\alpha/2}$ in equation (18) varying from 0 to 3 (– –) and from -3 to 0 (——), and they are magnified 3 times for ease of visualization.

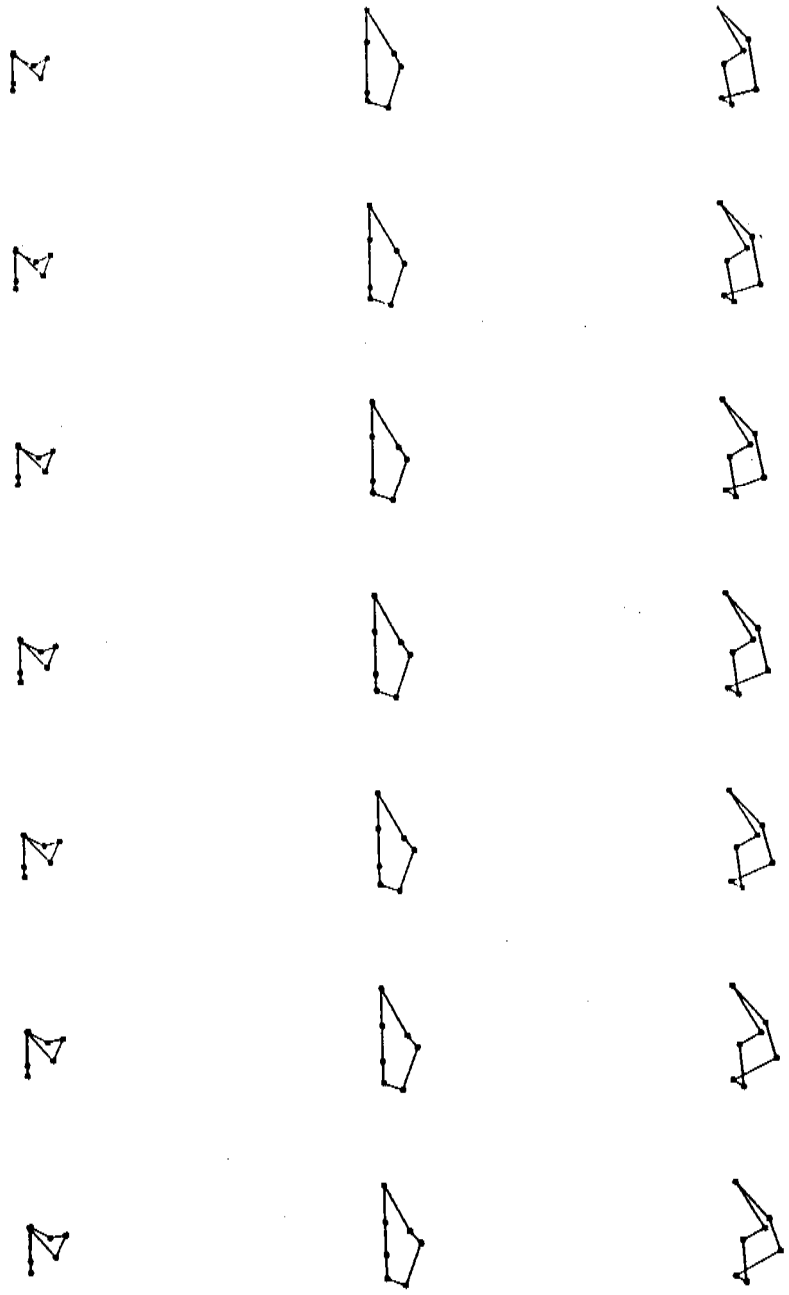

Fig. 6. The first principal component for the males. The figures are a) x-y, b) x-z and c) y-z projections of the figures evaluated along the first principal component. The plots show figures evaluated from equation (18) with $z_{\alpha/2} = -3$ (far left), -2, -1, 0, 1, 2, 3 (far right).

back into the configuration space for visualization using equation (16). The vectors are curved (as the inverse projection is not linear) but the vectors are approximately linear in the close vicinity of the mean shape. On Figure 5 the vectors have been magnified 3 times for ease of visualization.

The first principal component captures 47.4% of the shape variability. Most of the movement appears to be in the x-direction in landmark 3: bregma, and referring to Fig. 2 there appear to be two males with unusual bregma landmarks. A more detailed analysis is required to investigate these possible outliers and their effect on the interpretations.

An alternative way to view principal components is to animate the sequance of shapes that lie along the direction of each principal component. Thus, in Fig. 6, equation (19) is used to evaluate shapes at values of $z_\alpha/2 = -3, -2, -1, 0, 1, 2, 3$ along the first principal component. These are tranformed back to the configuration space using (16), and plotted in series for each of the three projections. Again we see that the main movement is in bregma in the x-direction. When animated these sequences are particularly illuminating and of course a full 3D representation is preferred. Note that variations are quite small here; in datasets with more variability (e.g., the resistors of Cootes et al., 1992) the pictures are more dramatic.

Note that there are actually $k = 24$ landmarks for these data and we have just considered a subset of $k = 7$ landmarks for illustrative purposes. As there are so few skulls in each group, inference is problematic due to the high dimensionality of the space. Further analysis will be considered elsewhere but the high dimensionality could be tackled by considering the first few principal components.

Acknowledgement. We would like to thank Dr. N. Jablonksi of the University of Western Australia for providing the data, which was collected with the support of the Australian Research Council. We would also like to thank Dr.P. O'Higgins for his assistance with the landmark data.

REFERENCES

BASS, N. (1971). *Human Osteology, A Laboratory Manual Of The Human Skeleton.* Special publication, Missouri Archaeological Society, Columbia.

BOOKSTEIN, F.L. (1986). Size and shape spaces for landmark data in two dimensions (with discussion). *Statist. Sci.*, **4**, 181–242.

BOOKSTEIN, F.L. and SAMPSON, P.D. (1990). Statistical models for geometric components of shape change. *Comm. Statist. – Theo. Meth.*, **19**, 1939–1972.

COOTES, T.F., TAYLOR, C.J., COOPER, D.H. and GRAHAM, J. (1992). Training models of shape from sets of examples. In *Proc. British Machine Vision Conference*, Springer–Verlag, pp. 266–275.

DRYDEN, I.L. and MARDIA, K.V. (1991). General shape distributions in a plane. *Adv. Appl. Prob.*, **23**, 259–276.

——— (1992). Size and shape analysis of landmark data. *Biometrika*, **79**, 57–68.

GOODALL, C.R. (1991). Procrustes methods in the statistical analysis of shape. *J. R. Statist. Soc., Ser. B*, **53**, 285–339.

GOODALL, C. R. and MARDIA, K.V. (1992). The noncentral Bartlett decompositions and shape densities. *J. Multivariate Analysis*, **40**, 94–108.

——— (1993). Multivariate aspects of shape theory. *Annals of Statistics*, **21**, 848–866.

GOWER, J. C. (1975). Generalized Procrustes analysis. *Psychometrika*, **40**, 33–50.

JARDINE, N. AND JARDINE, C. J. (1967). Numerical homology. *Nature*, **216**, 301.

KENDALL, D. G. (1984). Shape manifolds, procrustean metrics and complex projective spaces. *Bull. Lond. Math. Soc.*, **16**, 81–121.

——— (1989). A survey of the statistical theory of shape. *Statist. Sci.*, **4**, 87–120.

——— (1991). The Mardia-Dryden shape distribution for triangles : a stochasitc calculus approach. *J. Appl. Probab.*, **28**, 225-230.

KENT, J. T. (1994). The complex Bingham distribution and shape analysis. *J. Roy Statist. Soc. B*, To appear

KENT, J. T. and MARDIA, K.V. (1992). Statistical shape methodology in image analysis. *Tech. Report STAT 92/09C, Dept. Stat., Univ. of Leeds.*.

LE, H. and KENDALL, D. G. (1993). The Riemannian structure of Euclidean shape spaces : a novel environment for statistics. *Annals Statist.* To appear.

MARDIA, K.V. (1989). In discussion of Kendall (1989), *Statist. Sci.*, **4**, 108–111.

MARDIA, K.V. and DRYDEN, I. L. (1989a). Shape distributions for landmark data. *Adv. Appl. Probab*, **21**, 742–755.

——— (1989b). The statistical analysis of shape data. *Biometrika*, **76**, 271–281.

MARDIA, K.V., KENT, J. T. and BIBBY, J. (1979). *Multivariate Analysis*. Academic Press, London.

MOSIMANN, J.E (1970). Size allometry : size and shape variables with characterizations of the log-normal and generalized gamma distributions. *J. Am. Statist. Ass.*, **65**, 930–945.

TEN BERGE, J. M. F. (1987). Orthogonal procrustes rotation for two or more matrices. *Psychometrika*, **42**, 267–276.

ZIEZOLD, H. (1993). Mean figures and mean shapes applied to biological figure and shape distributions in the plane. *Biometrical J.* To appear.

DEPT. OF STAT.
UNIV. OF LEEDS
LEEDS, LS2 9JT
U. K.