CONCLUDING ADDRESS

Current Issues for Statistical Inference in Shape Analysis

John T. Kent
Department of Statistics
University of Leeds
Leeds, LS2 9JT, UK.

This paper has appeared in:
Proceedings in Current Issues in Statistical Shape Analysis (1995) edited by Mardia, K V and Gill, C A. Leeds University Press, 167–175.

Abstract

This paper gives a review of current statistical approaches to shape analysis together with some thoughts about their future development. Key points include the statistical objectives from the analysis of the data, the modelling assumptions behind the analysis, the statistical properties of the methods used and the type of numerical data extracted from geometric objects.

1 Introduction

Shape is typically defined as those aspects of a geometric object which remain invariant under changes in location, scale and rotation. In practice shape analysis is most often carried out for objects in 2 and 3 dimensions. In many examples objects contain identifiable landmarks, and it is these landmarks which form the basis of further statistical analysis. When the landmarks lie along the outline of an object there is sometimes extra information available in the form of tangent directions and curvatures. The use of this extra information will be considered in Section 5, but most of the paper is focussed on landmark-based methods.

There are several aspects of shape analysis which merit the attention of statisticians. These include estimation, testing, summarization and visualization. An overview of these aspects is given in Section 2.

A related statistical issue concerns modelling assumptions (Section 3). For example statistical models for shapes may be based on underlying models for the landmarks themselves, or they may be constructed directly within shape space. In some special cases specialized models may be constructed.

To illustrate some of the statistical issues described above, we consider in Section 4 the problem of estimating an "average shape" (and the related problem of estimating an "average form"). Estimation may be based on specific statistical models, or may be motivated by algorithmic procedures such as Procrustes analysis or multidimensional scaling. Statistical questions include consistency, efficiency and robustness of the methods used.

The last topic in the paper concerns the way in which geometric information about objects can be turned into numerical information for statistical analysis. As mentioned above, landmarks will usually form the basic information about objects on which the study of shape is based. However, there is sometimes further information to be considered such as tangents and curvatures of curves through landmarks. Some preliminary attempts to model this type of information are given in Section 5.

2 Statistical Aspects of Shape Analysis

The statistical aspects of shape analysis include estimation, testing, summarization and visualization. These aspects play a role in the study of problems such as average shape, shape variability and shape change.

The simplest problem to describe is that of estimating average shape. Using some of the models described in Section 3, several methods of constructing an average shape are compared in Section 4.

Another statistical objective is to describe and summarize variability in a set of data. This task is greatly facilitated when the data are highly concentrated so that the analysis can be carried out in a tangent space to shape space. Of special interest here is the construction of principal components to summarize and visualize the main aspects of shape variability (e.g. Kent, 1994, in dimension $p = 2$). In particular there is the question of an appropriate metric in which to construct these principal components. Kent (1994) suggested calculating principal components with respect to the Euclidean metric in tangent space. However, in some recent work, Bookstein (1994) has suggested taking account of the geometry of the landmarks in $\mathbb{R}^p$ by taking principal components with respect to the bending energy metric (or its generalized inverse, respectively) from thin plate spline analysis. By this means the dominant principal components will focus on small-scale (or large-scale, respectively) features in the data.

The last statistical objective is to describe changes in shape. For example, there may be two distinct groups in the data, and the objective might be to describe the differences between the two groups. Alternatively, there may be explanatory variables and the objective might be to describe how shape changes as these explanatory variables change, that is, a regression analysis. These problems were originally addressed in the seminal paper of Bookstein (1986) and include all four aspects listed above.

Visualization of shape change is often tackled by constructing a deformation of the plane. The effect of the deformation can then be visualized by plotting an object or an orthogonal grid before and after the deformation. Bookstein has played a key role in the development of deformation-based analysis, including the use of thin-plate splines and their decompositions in terms of principal warps (Bookstein, 1989,1991), and the incorporation of first order derivative information (Bookstein and Green, 1993a,b). A more general framework in which to include derivative constraints in deformations was given by Mardia et al (1993).

3 Modelling Strategies

Before discussing specific models for landmark-based shapes, it is useful to set up some notation to describe configurations of landmarks and their associated shapes. Thus, let

$$X = \begin{bmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_{k+1}^T \end{bmatrix}$$

be a $(k+1)\times p$ matrix of $k+1$ landmarks in $\mathbb{R}^p$. One simple way to eliminate location effects is to look at HX where H is a $k\times(k+1)$ matrix with orthonormal rows, satisfying $H\mathbf{1}_{k+1} = \mathbf{0}_k$, where $\mathbf{1}_{k+1}$ and $\mathbf{0}_k$ are vectors of ones and zeros, respectively. Similarly, a simple way to eliminate scale effects is by restricting attention to the "pre-shape" $HX/\|HX\| = Y$, say, where $\|\cdot\|$ denotes the Euclidean norm. Finally, to eliminate rotation effects, treat Y and YR as equivalent, for all $p\times p$ rotation matrices $R \in SO(p)$.

In $p = 2$ dimensions, this representation can be simplified by using complex coordinates. Thus, let $\mathbf{u} = (u_1, ..., u_{k+1})^T \in \mathbb{C}^{k+1}$ be a vector of landmarks. Size and location effects can be eliminated by looking at the pre-shape $H\mathbf{u}/\|H\mathbf{u}\| = \mathbf{z}$, say, where $\mathbf{z}$ lies on the unit sphere in $\mathbb{C}^k$. Rotation effects can be eliminated by treating $\mathbf{z}$ and $e^{i\theta\mathbf{z}}$ as equivalent for all $\theta \in [0, 2\pi)$. Kendall (1984) and Le and Kendall (1993) give a detailed investigation into the geometry of shape space both in $p = 2$ and higher dimensions.

In many practical applications the shape of a set of data configurations will be concentrated about a fixed shape. Denote a representative pre-shape of this fixed shape by $M(k\times p)$ for $p > 2$ and $\boldsymbol{\mu} \in \mathbb{C}^k, p = 2$. Thus, it is useful to consider tangent approximations of shape space at M or $\boldsymbol{\mu}$, respectively.

Goodall (1991) describes the tangent space to shape space at M. In particular a matrix $T(k\times p)$ can be regarded as lying in the tangent space if

$$\text{tr } M^T T = 0 \tag{3.1}$$

$$\text{tr } M^T T S = 0 \tag{3.2}$$

for all skew symmetric $p\times p$ matrices S.

Given a pre-shape Y, let $\hat{Y}$ denote the "full" Procrustes fit to M (i.e. $\hat{Y} = \hat{\gamma} Y \hat{R}$ where $\hat{\gamma} > 0$ and $\hat{R} \in SO(p)$ are chosen to minimize the sum of squares $\| \gamma Y R - M \|^2$). Note that $\hat{Y}$ already satisfies (3.2) since $\| \hat{\gamma} \hat{Y} R - M \|^2$ is minimized over $R \in SO(p)$ at $R = I$, and since any rotation matrix R can be written in the form $R = \exp(iS)$ where S is skew symmetric. Then a "full Procrustes projection" onto the tangent space at M can be defined by orthogonally projecting $\hat{Y}$ onto the tangent space at M, treating $\hat{Y}$ as a vector in $\mathbb{R}^{pk}$ and the tangent space as a subspace of $\mathbb{R}^{pk}$. A variation of this procedure is a "partial" Procrustes fit and projection, which can be defined by fixing $\gamma = 1$ above. Dryden and Mardia (1993) describe the use of the partial Procrustes projection.

In $p = 2$ dimensions this calculation can be simplified by using complex coordinates. In particular, the partial Procrustes projection of $\mathbf{z}$ takes the form

$$e^{-i\theta}(I - \boldsymbol{\mu}\boldsymbol{\mu}^*)\mathbf{z} = \mathbf{v}, \quad \text{say, where} \quad \theta = \arg \boldsymbol{\mu}^* \mathbf{z} \tag{3.3}$$

(Kent, 1994). In practice the difference between the full and partial projections will be minimal for concentrated data.

Another way to represent shape data is given by Bookstein coordinates. In $p = 2$ dimensions, they are obtained by translating, scaling and rotating the configuration so that landmark 1, say, lies at the origin and landmark 2, say, lies at $(1, 0)$. In complex coordinates they are given by

$$w_j = (u_j - u_1)/(u_2 - u_1), \ j = 3, ..., k+1, \tag{3.4}$$

where $(w_3, ... w_{k+1})$ is unrestricted in $\mathbb{C}^{k-1}$. An extension of these coordinates to $p > 2$ dimensions was given by Goodall and Mardia (1993). Note that the specific coordinates constructed depend nonlinearly on the pair of landmarks chosen as the "baseline".

Thus there are at least 4 general modelling strategies which can be used for shape analysis: (1) models on landmarks themselves giving rise to derived distributions on shapes, (2) models in shape space, (3) models in tangent space to shape space, or (4) models in Bookstein coordinates. We shall discuss each strategy in turn. In this paper we shall limit attention to models for independent identically distributed shapes.

The simplest model for landmarks is to assume that the landmarks follow a multivariate normal distribution about a mean configuration. Various levels of generality can be assumed for the covariance matrix, including (a) $\Sigma = \sigma^2 I_k \otimes I_p$ (isotropy, i.e. the coordinates of all the landmarks are independent with the same variance), (b) $\Sigma = \Sigma_k \otimes I_p$, allowing correlations between landmarks, (c) $\Sigma = \Sigma_k \otimes \Sigma_p$, allowing structured correlations between and within landmarks, and (d) Σ unrestricted. Mardia and Dryden (1989) and Dryden and Mardia (1991) worked out the derived distributions in shape space under this model in $p = 2$ dimensions. Goodall and Mardia (1993) give some expansions for the shape density in higher dimensions $p \geq 3$ when the mean configuration matrix is singular under assumptions (a) and (b) on Σ.

A tractable direct model on shape space in $p = 2$ dimensions is given by the complex Bingham distribution with density proportional to $\exp\{\mathbf{z}^* A \mathbf{z}\}$ where A is $k \times k$ Hermitian (Kent, 1994). A limitation of this model is that the deviations from the modal shape have complex symmetry in distribution, a feature which limits the usefulness of the distribution for describing patterns of variability. Therefore, for concentrated data it is also useful to consider alternative models such as an unrestricted multivariate normal distribution in the tangent space to shape space (Kent, 1994). Of course, before using a tangent space model it is first necessary to estimate a mean shape at which to take the projection.

Another approach, pioneered by Bookstein (1986), is to use an unrestricted multivariate normal distribution in Bookstein coordinates. This approach has the advantage of simplicity, but depends on the baseline chosen.

It is of interest to compare these approaches for concentrated data. Approach (1) involves an approximate linear projection from the $(k+1)p$-dimensional space of landmarks onto the $kp-p(p-1)/2-1$-dimensional shape space. Thus $p+p(p-1)/2+1$ variables will be approximately unidentifiable (the parameters for location, rotation and scale). For estimation it is important to assume that any parameters of Σ present in the model remain well-defined when projecting configurations onto shape space. All the approaches give rise to an approximate multivariate normal model in any suitable coordinates, such as tangent coordinates or Bookstein coordinates. The mapping between these two coordinate systems is approximately linear, but not orthogonal. For example the complex Bingham distribution is approximately isotropic multivariate normal in tangent space coordinates, but there are correlations between the landmarks in Bookstein coordinates. In consequence, Kent (1994) argued that principal component analysis was better carried out in tangent space coordinates. See also the discussion at the end of Section 2.

Special models may be constructed in certain cases where there is additional symmetry. For example, Grenander and Miller (1994) propose a model in which the landmarks are normally distributed with a block circulant covariance matrix. This model is motivated by the situation where the underlying object has no natural landmarks on it, but instead n equally spaced points around the outline are chosen as landmarks. This model is invariant under cyclic relabelling of the landmarks. Further analysis of the symmetries in this model can be found in Kent et al, (1995a).

4 Estimation of Average Shape

In this section we focus on the question of consistency for a specific estimation problem. Given data which arise as the shapes of multivariate normally distributed configurations $X_j(k \times p)$, $j = 1, ..., n$, the objective is to estimate the shape or form of the mean configuration. Recall that the "form" or "size-and-shape" of an object consists of those aspects of an object which remain invariant under changes in location and rotation (but not scale).

Several methods of estimation will be considered.

(i) Procrustes analysis (Goodall, 1991). "Full" or "partial" Procrustes analysis can be used to estimate shapes; "partial" Procrustes analysis can be used to estimate forms. In $p = 2$ dimensions Kent (1994) showed that the full Procrustes estimate of average shape coincides with the maximum likelihood estimate for the complex Bingham distribution.

(ii) Maximum likelihood estimation using the marginal distribution in shape or form space for normally distributed configurations of landmarks. See Mardia and Dryden (1989) and Dryden and Mardia (1991) for the shape density in $p = 2$ dimensions under various assumptions on Σ; see Dryden and Mardia (1992) for the form density in $p = 2$ dimensions; and see Goodall and Mardia (1993) for some expansions for shape and form for $p \geq 3$.

(iii) Naive multidimensional scaling (MDS). This method is outlined in Kent (1994) and provides an estimate of mean form or shape (up to reflection) based on the sample squared inter-landmark distances of $X_j((k+1) \times p)$ or $X_j / \parallel H^T H X_j \parallel$, respectively, in the notation of Section 3. Note that the denominator $\parallel H^T H X_j \parallel = \parallel H X_j \parallel$ scales the centred configuration to have sum of squares equal to 1.

(iv) Refined MDS (Lele, 1993). This refined method, available for forms only, involves adjustments to the above method based on moments to correct for biases in the sample squared inter-landmark distances as estimates of the squared inter-landmark distances of the mean configuration.

(v) Sample mean in Bookstein coordinates (given by (3.4) in $p = 2$ dimensions) to estimate average shape. Similar edge superimposition coordinates (given by $w_j =$ $(u_j = u_1)|u_2 - u_1|/(u_2 - u_1),\ j = 3, ..., k+1$ in $p = 2$ dimensions) can be used to estimate average form.

We shall now discuss the consistency properties of the above methods. However, it should be emphasized that for concentrated data any possible inconsistencies will usually be swamped by the variability in the data. Thus possible inconsistency of these methods is not usually an important statistical problem.

Kent and Mardia (1995) show that full Procrustes estimation in $p = 2$ dimensions is consistent for shape under an isotropic distribution for the landmarks (not necessarily a normal distribution). However, if the errors are not isotropically distributed, then inconsistencies can arise. Goodall (1991) suggests some modifications to Procrustes analysis to deal with non-isotropic errors but it is not clear that these modifications will be effective.

In general, the maximum likelihood estimates of mean shape or form will always be consistent, provided the parameters of Σ are mathematically identifiable. However, as noted in the last section, it is important for practical estimation that Σ should not contain too many parameters.

Next consider estimating mean shape using the sample mean in Bookstein coordinates. It is well-known that the sample mean will not be exactly consistent (Bookstein, 1988). Further, as noted by Mardia and Dryden (1994), if the raw landmarks are normally distributed, then the transformed landmarks in Bookstein coordinates have finite moments of first order but not of higher order. They also work out the bias of the sample mean in Bookstein coordinates for the isotropic case. Note that the lack of finite moments of order ≥ 2 reflects more the limitation of a multivariate normal distribution as a sensible model for landmarks rather than a limitation of Bookstein coordinates. A similar lack of consistency for estimating mean form was noted by Lele (1993).

Overall, for concentrated data, the consistency question for shape can be viewed as follows. The multivariate normal distribution for landmarks asymptotically projects onto a multivariate normal distribution in the tangent space to shape space, centred at the shape of the population mean configuration. Further, the above methods (i) (full and partial), (ii), (iii) and (v) for calculating an average shape are asymptotically equivalent to calculating a sample mean vector in the tangent space. Since the sample mean vector is consistent for the population mean vector of a multivariate normal distribution, it follows that all the above methods will be asymptotically consistent under high concentration for the population mean shape.

Finally, consider estimating a mean form. Kent and Mardia (1995) show that partial Procrustes analysis in $p = 2$ dimensions for estimating form is consistent up to a size factor under isotropic normal errors. Lele's (1993) refined method of multidimensional scaling is consistent, not only under isotropic normal errors, but also under certain sorts of anisotropic normal errors. However, his methodology is very sensitive to the assumption of normality. In particular, his estimators depend on the fourth order moments from the normal distribition. Thus, his method will be very sensitive to departures from normality, for example, if the errors are distributed according to a t-distribution. Also his methodology will not be resistant to outliers.

5 Incorporating Derivative Information

As noted in the introduction, the use of landmarks forms the basis for much of statistical shape analysis. However, when landmarks lie along the outline of an object there is additional information which can be incorporated, namely the tangent and curvature of the outline at that landmark. In this section we discuss some possible statistical models to incorporate this additional information.

To simplify the modelling we work in the case of high concentration in $p = 2$ dimensions. Thus, we consider a fixed tangent space to shape space, where the tangent space is taken at an average shape for the data.

The data consist of the projected landmark vectors $\mathbf{v} \in \mathbb{C}^k$ from (3.3), and tangent angles curvatures at some or all the $k+1$ landmarks (a vector $\boldsymbol{\psi}$ of angles, and a vector $\mathbf{k}$ of real-valued curvatures). By construction $\mathbf{v}$ will typically have mean 0; however, $\boldsymbol{\psi}$ and $\mathbf{k}$ will have general means $\boldsymbol{\psi}_0$ and $\mathbf{k}_0$, say. A possible model is assume the $\mathrm{Re}(\mathbf{v})$, $\mathrm{Im}(\mathbf{v}), \boldsymbol{\psi}$ and $\mathbf{k}$ follow a joint multivariate normal model, treating the elements of $\boldsymbol{\psi} - \boldsymbol{\psi}_0$ as lying in $(-\pi, \pi)$. If the elements of $\boldsymbol{\psi}$ are not sufficiently concentrated to support a normal approximation, the exponential family models of Jupp and Mardia (1980) can be tried. In this model $\{\mathrm{Re}(\mathbf{v}),\ \mathrm{Im}(\mathbf{v}), \mathbf{k}\}$ given $\boldsymbol{\psi}$ is normally distributed and each element of $\boldsymbol{\psi}$ given the remaining variables follows a von Mises distribution. In practice it may be necessary to simplify the parameterization in order to produce tractable models. It remains to be seen how much useful insight such models can give in the analysis of shape data.

One of the problems with landmark-based methods is finding the landmarks in the first place. Landmarks are most easily identified where there is a sharp kink in the data, e.g. at a point of high curvature. Landmarks are more difficult to locate precisely along a stretch of the outline which has roughly constant curvature. One possible use of the tangent information here is to highlight the property that landmarks are likely to be less accurately determined along the tangent direction than in the perpendicular direction.

6 Future Directions

Landmarks are summary statistics about objects and it is understanding the behaviour of the objects which is of prime interest. Thus turning information about landmark behaviour into a description of an object's behaviour is important. Bookstein's (1994) use of the thin-plate spline bending energy metric in the construction of principal components is an important step in moving from landmarks to objects, but this is an area in which more work is needed.

Statistical analysis is more straightforward with fewer landmarks, but outlines can be more faithfully reproduced with larger numbers of landmarks. As the number of landmarks increases, a functional version of shape analysis appears in the limit. For example, Kent et al. (1995b) describe the limiting behaviour of a conditional autoregressive model for outlines due to Grenander et al. (1991). In general, principal component analysis extends quite easily to the functional form of shape analysis; see e.g. Cootes et al. (1992) for an application in machine vision. However, in other contexts, regularization will be needed. For example, Leurgans et al. (1993) give an application of canonical correlation analysis to functional data.

Acknowledgements

I am grateful to Kanti Mardia and Ian Dryden for helpful discussions during the writing of this paper.

References

[1] Bookstein, F.L. (1986) Size and shape spaces for landmark data in two dimensions (with discussion). *Statist. Sci.*, **1**, 181-242.

[2] Bookstein, F.L. (1989) Principal warps: thin plate splines and the decomposition of deformations. *IEEE. Trans. PAMI* **11**, 567-585.

[3] Bookstein, F.L. (1991) *Morphometric Tools for Landmark Data: Geometry and Biology.* Cambridge: Cambridge University Press.

[4] Bookstein, F.L. and Green, W.D.K. (1993) A feature space for edgels in images with landmarks. *J. Math. Imaging and Vision* **3**, 231-261.

[5] Bookstein, F.L. and Green, W.D.K. (1993) A thin-plate spline and the decomposition of deformations. In *Mathematical Methods in Medical Imaging* II, eds. D.C. Wilson and J.N. Wilson. SPIE Proceedings Vol 2035, 14-28.

[6] Bookstein, F.L. (1994) The morphometric synthesis: a statistical praxis for biomedical images with landmarks. manuscript.

[7] Cootes, T.F., Taylor, C.J., Cooper, D.H. and Graham, J. (1992) Training models of shapes from sets of examples. In *Proc. British Machine Vision Conf., Leeds*, 9-18. New York: Springer.

[8] Dryden, I.L. and Mardia, K.V. (1991) General shape distributions in a plane. *Adv. Appl. Probab.*, **23**, 259-276.

[9] Dryden, I.L. and Mardia, K.V. (1992) Size and shape analysis of landmark data. *Biometrika* **79**, 57-68.

[10] Dryden, I.L. and Mardia, K.V. (1993) Multivariate shape analysis. *Sankyā* **55** Series A, 460-480.

[11] Goodall, C.R. (1991) Procrustes methods in the statistical analysis of shape (with discussion). *J.Royal Statist. Soc. B.* **53**, 285-339.

[12] Goodall, C.R. and Mardia, K.V. (1993) Multivariate aspects of shape theory. *Ann. Statist.* **21**, 848-866.

[13] Grenander, U., Chow, Y. and Keenan, D.M. (1991) *Hands: A Pattern Theoretic Study of Biological Shapes.* Research notes in Neural Computing, Vol 2, Springer-Verlag, New York.

[14] Grenander, U. and Miller, M.I. (1994) Representations of knowledge in complex systems (with discussion). *J. Royal Statist. Soc. B* **56**, 549-603.

[15] Jupp, P.E. and Mardia, K.V. (1980) A general correlation coefficient for directional data and related regression problems. *Biometrika* **67**, 163-173.

[16] Kendall, D.G. (1984) Shape-manifolds, procrustean matrices and complex projective spaces. *Bull. Lond. Math. Soc.*, **16**, 81-121.

[17] Kent, J.T. (1982) The Fisher-Bingham distribution on the sphere. *J.Royal Statist. Soc. B* **44**, 71-80.

[18] Kent, J.T. (1992) New directions in shape analysis. In *The Art of Statistical Science*, ed. K.V. Mardia, Chichester: Wiley, 115-127.

[19] Kent, J.T. (1994) The complex Bingham distribution and shape analysis. *J.Royal Statist. Soc. B*, **56**, 285-299.

[20] Kent, J.T. and Mardia, K.V. (1994) The link between kriging and thin plate splines. *Probability, Statistics and Optimization.* ed. F.P. Kelly. New York: Wiley, 324-339.

[21] Kent, J.T. and Mardia, K.V. (1995) Consistency of Procrustes estimators. Research Report STAT 95/05, Department of Statistics, University of Leeds.

[22] Kent, J.T., Anderson, C. and Dryden, I.L. (1995a) Deformations with circulant symmetry. In preparation.

[23] Kent, J.T., Mardia, K.V. and Walder, A.N. (1995b) Conditional cyclic Markov random fields. *Adv. Appl. Probab.*, to appear.

[24] Le, H. and Kendall, D.G. (1993) The Riemannian structure of Euclidean shape spaces: a novel environment for statistics. *Ann. Statist.* **21**, 1225-1271.

[25] Lele, S. (1993) Eudidean distance matrix analysis (EDMA): estimation of mean form and mean form difference. *Math. Geology* **25**, 573-602.

[26] Leurgans, S.E., Moyeed, R.A., and Silverman, B.W. (1993) Canonical correlation analysis when the data are curves. *J.Royal Statist. Soc. B* **55**, 725-740.

[27] Mardia, K.V. and Dryden, I.L. (1989) The statistical analysis of shape data. *Biometrika*, **76**, 271-281.

[28] Mardia, K.V. and Dryden, I.L. (1994) Shape averages and their bias. *Adv. Appl. Probab.* **26**, 334-340.

[29] Mardia, K.V., Kent, J.T., Goodall, C.R., and Little, J. (1993) Kriging and splines with derivative information. Research Report STAT 93/22, Department of Statistics, University of Leeds.