**Medical Image Analysis** (1996/7) volume 1, number 3, pp 225–243


# Landmark methods for forms without landmarks: morphometrics of group differences in outline shape

Fred L. Bookstein*

Institute of Gerontology, University of Michigan, 300 North Ingalls Building, Ann Arbor, MI 48109-2007, USA

**Abstract**

Morphometrics, a new branch of statistics, combines tools from geometry, computer graphics and biometrics in techniques for the multivariate analysis of biological shape variation. Although medical image analysts typically prefer to represent scenes by way of curving outlines or surfaces, the most recent developments in this associated statistical methodology have emphasized the domain of landmark data: size and shape of configurations of discrete, named points in two or three dimensions. This paper introduces a combination of Procrustes analysis and thin-plate splines, the two most powerful tools of landmark-based morphometrics, for multivariate analysis of curving outlines in samples of biomedical images. The thin-plate spline is used to assign point-to-point correspondences, called semi-landmarks, between curves of similar but variable shape, while the standard algorithm for Procrustes shape averages and shape coordinates is altered to accord with the ways in which semi-landmarks formally differ from more traditional landmark loci. Subsequent multivariate statistics and visualization proceed mainly as in the landmark-based methods. The combination provides a range of complementary filters, from high pass to low pass, for effects on outline shape in grouped studies. The low-pass version is based on the spectrum of the spline, the high pass, on a familiar special case of Procrustes analysis. This hybrid method is demonstrated in a comparison of the shape of the corpus callosum from mid-sagittal sections of MRI of 25 human brains, 12 normal and 13 with schizophrenia.

*Keywords:* morphometrics, multivariate analysis, outlines, Procrustes analysis, thin-plate spline

*Received January 15, 1996; revised December 12, 1996; accepted January 3, 1997*

## 1. INTRODUCTION

Over the last decade, previously scattered or fragmentary tactics and techniques from medical image analysis, multivariate statistics and computational geometry have been interwoven very effectively in a newly standardized methodology for landmark data. This morphometric synthesis binds together the Riemannian structure of David Kendall's shape space, multivariate statistical maneuvers in the tangent space at the Procrustes average form and graphical approaches for visualizing a wide variety of signals in the resulting data sets. The synthesis brings to the analysis of medical images a biometrical spirit—a concern for optimal description of causes and effects—that was hitherto limited to the more esoteric reaches of evolutionary biology. Recent reviews include Bookstein (1996a, 1997) and the papers in Marcus *et al*. (1996).

The power of the synthesis for scientific description, however, has thus far typically been bound to the very demanding abstraction of landmark point data. This is inconvenient for a number of reasons. Such data cannot at present be supplied automatically with any reliability, but instead require the scientist either to locate these loci herself or continually to correct the erroneous selections produced by automatic algorithms. Landmark data seem unavailable for large extents of scientifically important images, and for others, such as renderings of the human cerebral cortex, landmarks cannot be declared with assurance to correspond across reasonable ranges of normal adult variation (to say nothing of disease states). Most seriously, the discrete structure of landmark data does not particularly suit the semantics of scientific explanations

*Corresponding author
(e-mail: fred@brainmap.med.umich.edu)

of form, which are typically concerned with causes or consequences of variation in surface area or volume or with 'forces' (biomechanics, ultrastructure) associated with image content over extended regions.

The two techniques at the core of contemporary morphometrics are the Procrustes-projection construction of shape coordinates and the visualization of localized shape phenomena by thin-plate splines. (Both will be reviewed in the next section.) There have been several previous attempts to extend these techniques from the somewhat constricted domain of landmarks to the curving form more generally. The thin-plate spline, for instance, has been extended to incorporate arbitrary information about affine derivatives (Bookstein and Green, 1993) and curvature (Little and Mardia, 1996; Mardia *et al*., 1996) and to treat whole extended curves spanning landmarks (Cutting *et al*., 1995). Procrustes analysis is likewise being extended to whole curves (see, for instance, Sampson *et al*., 1996). But in these extensions there has not hitherto been any formal coherence analogous to the statistical geometry that binds together the two approaches for landmarks.

This paper introduces such a combination of these tools for a problem that apparently has not previously been formalized: the task of effectively describing group differences in data from curving forms that, while not featureless, nevertheless need not have any reliably point-like landmarks anywhere along the arcs. We will proceed by construing each component of the synthesis, the spline and the Procrustes fit, as a gently nonlinear filter for regional differences of outline shape. We will find the filters to be related by a directional complementarity of band-pass characteristics, so that the power of the Procrustes-based filter is greatly heightened after a spline-based preprocessing.

The data set used here to exemplify the localization techniques of this paper has been exploited previously to demonstrate landmark-based techniques (Bookstein, 1995a, 1996b). The original images are single thick parasagittal slices, of not particularly high quality, selected from clinical MR brain scans of 12 doctors and 13 patients from the Adult Psychiatry Unit, University of Michigan Hospitals. Previous analyses of the group difference began with a data set of 13 landmarks located by John DeQuardo, MD; the discrimination was sharpened after I included four additional intersects placed around the inner boundary of the corpus callosum. But of the original 13 landmarks, three were already positioned rather arbitrarily along that arc and two others lay with equal arbitrariness along the falx cerebri. My colleague, Bill Green, in preparing to develop the algorithms for averaging here (Green, 1995, 1996), deleted those five landmarks in favor of augmenting the data set by tracings of the two arcs in full. For the complete feature set, the average of all 25 forms, against which each case has been relaxed, is shown in Figure 1. From this scheme

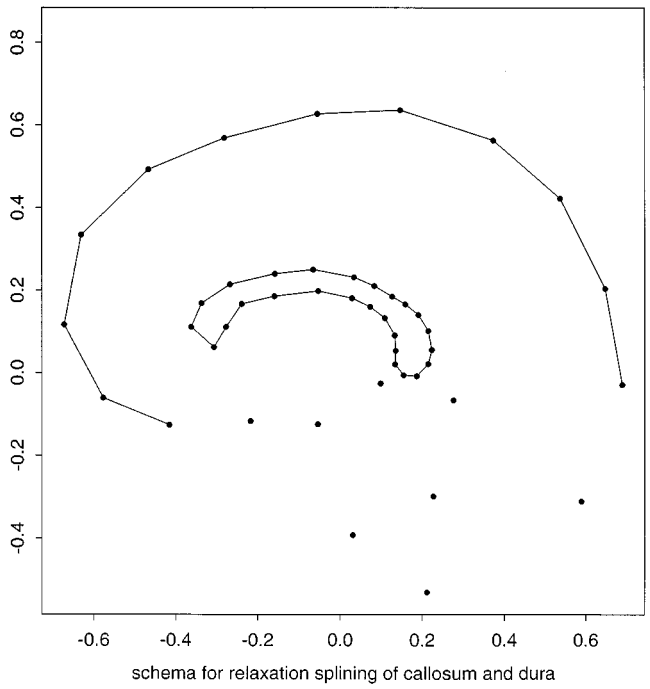


**Figure 1.** Scheme for relaxing information about the curving form of a parasagittal brain section: eight landmark points, a callosal arc of 26 points and a calvarial/dural arc of 12 points. Only the callosal arc is used in this paper, but assignments of homology are based on a spline relaxation that involves the other structures as well. Here the geometry corresponds to the Procrustes mean shape. The head is facing to the left.

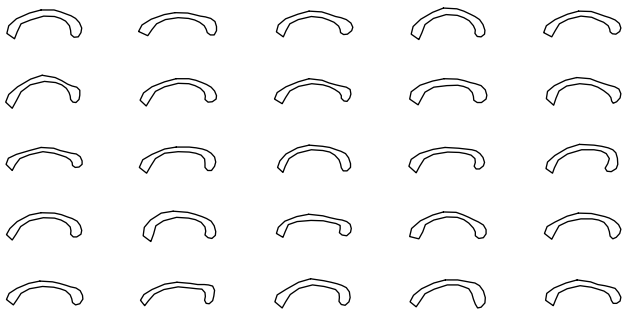

**Figure 2.** Our data set: 25 26-point polygons around the callosal border as traced by W. D. K. Green. Medical staff, first 12 forms; schizophrenics, last 13. Sample and planes selected by John DeQuardo, Adult Psychiatry Unit, University of Michigan. The bulbosity at the rightmost (posterior) end is named 'splenium'; the one at the left (anterior), 'genu'.

we will use only the callosal outline itself, a polygon of 26 semi-landmarks in each of the forms. These 25 polygons are shown in Figure 2. The bulb at the far right (toward the back

of the head) is called splenium; that at the left (anterior), genu ('knee'). The narrowing of the arch near splenium is the isthmus. Owing to limitations in the original images, we chose to intentionally blunt the rostrum of the corpus callosum, which otherwise would appear as a right-facing cusp at their lower left. The first 12 images in this series pertain to the doctors and the last 13 to their patients.

## 2. PRINCIPAL TECHNIQUES OF THE MORPHOMETRIC SYNTHESIS

### 2.1. Procrustes shape distance

In ordinary language the shape of an object is described by words or quantities that do not vary when the object is moved, rotated, enlarged or reduced. The translations, rotations and changes of scale we thereby ignore constitute the similarity group of transformations of the plane. When the 'objects' are finite ordered point sets (either landmarks or the semilandmarks to be introduced presently), it turns out to be useful to say that their shape is simply the set of all point sets that 'have the same shape'. That is, we have formally defined the shape of a set of points as the equivalence class of that point set, within the collection of all point sets of the same cardinality, under the operation of the similarity group.

Almost all of the multivariate statistics to follow will rely on one simple construction: a distance measure for landmark shapes. It proves most natural to define the squared Procrustes distance between two shapes (that is, two equivalence classes) as the minimum of summed squared distances between corresponding points over the similarities that studies of shape are to ignore. As Kendall (1984) explains, this is the only approach consistent with how we want a shape distance to relate to Euclidean distance in the original plane. Once the scale of a landmark set A is fixed, the squared shape distance between A and another landmark set B is the minimum of summed squared Euclidean distances between the landmarks of A and the corresponding landmarks in point set C as C ranges over the whole set of shapes equivalent to B. Under this metric, the set of equivalence classes actually becomes a Riemannian manifold, Kendall shape space, which is a submersion of the original $\mathbf{R}^{2k}$ or $\mathbf{R}^{3k}$.

The Procrustes metric is most easily understood graphically. The top row of Figure 3 shows two quadrilaterals of points presumed to be landmarks. Compute the centroid of each set of four, then rescale so that the sum of squares of the distances shown is fixed at unity (second row). The original root sum of squares, the scaling factor here, is usually called 'centroid size' in the statistical literature; it is the same as the square root of the net moment of the original set of landmarks around their centroid. Superimpose one of the scaled forms over the other at their centroids and spin it (third row).

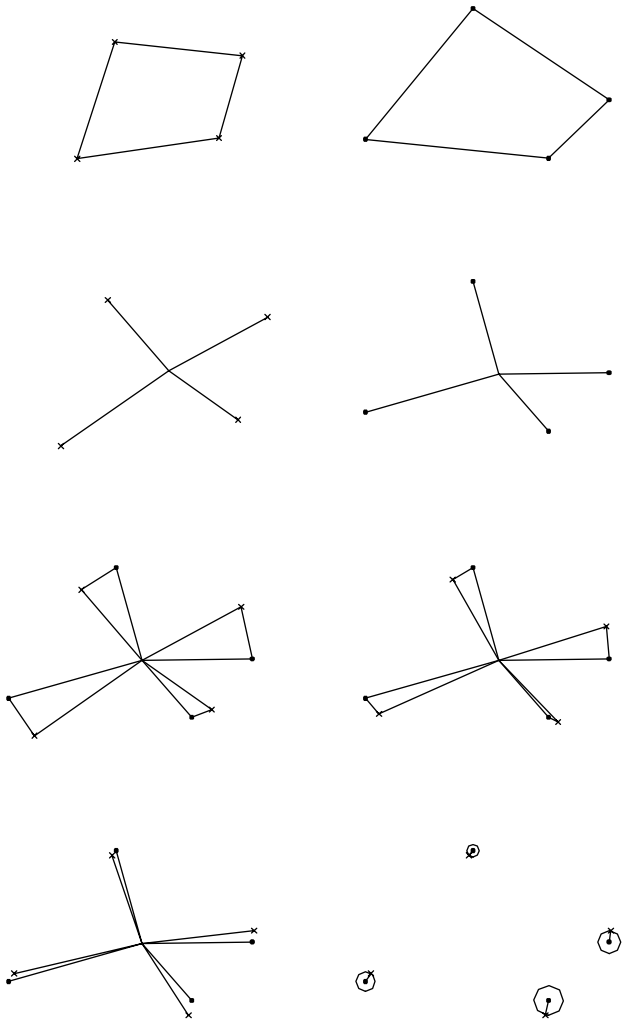

**Figure 3.** Procrustes shape distance for two quadrilaterals of landmarks (top row). Each form is scaled separately (second row) to sum of squares 1 around its own centroid. After the centroids are superimposed (third row left), either form is rotated about the other (right) until the sum of squared distances between matched landmarks is minimized. Fourth row, the squared Procrustes distance between the original quadrilaterals is the minimum sum of squared distances between corresponding points; it is proportional to the sum of the areas of the circles drawn here.

In general, there will be a unique rotation (fourth row, left) that minimizes the sum of squares of the distances between corresponding landmarks. The squared Procrustes distance between the forms is the sum of squares of those residual distances at its minimum. It is proportional to the total area of the circles shown at the lower right.

Formulas for this procedure have been familiar to the multivariate statistical community since Gower's original matrix formulation (see, for instance, Rohlf and Slice, 1990). If $X_1$ and $X_2$ are $k \times p$ matrices ($p = 2$ or 3) for the coordinates of $k$ landmarks after the centering and rescaling steps indicated in the figure, and if the singular-value decomposition of $X_1^{\mathrm{t}} X_2$ is $U D V^{\mathrm{t}}$ with all elements of $D$ positive, then the rotation required to superimpose $X_2$ upon $X_1$ optimally is just the matrix $V U^{\mathrm{t}}$, $2 \times 2$ or $3 \times 3$. In a different notation, for two complex vectors $\boldsymbol{z}_j = (z_{1j}, \ldots, z_{kj})^{\mathrm{t}}$, $j = 1, 2$, with $\sum_i z_{ij} = 0$ and $\sum_i z_{ij}\bar{z}_{ij} = 1$, the superposition of the second form upon the first is approximately

$$\boldsymbol{z}_2 \to \left( \sum_i z_{i1}\bar{z}_{i2} \right) \boldsymbol{z}_2 \qquad (1a)$$

and the Procrustes distance between the two shapes is approximately

$$\mathrm{PD}^2(\boldsymbol{z}_1, \boldsymbol{z}_2) = 1 - \left| \sum_i z_{i1}\bar{z}_{i2} \right|^2. \qquad (1b)$$

This version takes the form of an 'error variance' $1 - R^2$ in a univariate regression (Bookstein, 1991). A formulation according better with the global geometry of shape space, but less explicitly indicating its origin in a least-squares problem, is $\mathrm{PD}(\boldsymbol{z}_1, \boldsymbol{z}_2) = \cos^{-1} |\sum_i z_{i1}\bar{z}_{i2}|$. The rotation in the fourth row of the figure is by an angle $\arg \sum_i z_{i1}\bar{z}_{i2}$.

### 2.2. Averaging shapes and the Procrustes tangent space

The average of an ordinary list of numbers—their sum divided by their count—has a least-squares property: it is the quantity about which they have the least sum of squared differences. This notion goes over directly to the study of shapes (equivalence classes) now that we have a distance measure. (In general, this indirect approach to averaging complex structures in a metric space is called a Fréchet mean.) The necessary minimization is not too difficult (Kent, 1995). Indeed, for two-dimensional data, it can be expressed as a closed-form eigenanalysis. For a sample of complex vectors $\boldsymbol{z}_j^{\mathrm{t}} = (z_{1j}, \ldots, z_{kj})$, $j = 1, \ldots N$ with $\sum_i z_{ij} = 0$ and $\sum_i z_{ij}\bar{z}_{ij} = 1$, all $j$, the Procrustes average is the shape $\boldsymbol{z}$ that is the first eigenvector of the matrix $\sum_j \boldsymbol{z}_j \bar{\boldsymbol{z}}_j^{\mathrm{t}}$, $k \times k$.

In practice, for reasons I will explain in the next paragraph, this average is more often computed by the iterative approach sketched in Figure 4—an alternation between fitting to a tentative average and averaging of the fitted locations landmark by landmark. As in this toy example, beginning from any shape in a sample, typically the algorithm converges to sufficient accuracy by the second iteration. After we have computed the average, we can put each individual shape down over the average using the similarity transformation that made the sum of squares from the average a minimum for that particular case, resulting in the diagram at the lower center in Figure 4.

This tactic, far from being a mere graphical aid, is in fact the most crucial step in the morphometric toolkit. Kendall shape space is of dimension $2k - 4$, $2k$ original Cartesian coordinates decremented by the four degrees of freedom (two for translation, one for rotation, one for scale) of the similarity group of the plane. The tangent space to this manifold is likewise of dimension $2k-4$. The construction at the lower center in Figure 4 actually shows the projection of each shape of our sample onto the tangent space to that manifold at the sample average shape. That is to say, the Riemannian metric (in the small, in the vicinity of the sample average) is preserved—for each pair of specimens, the summed squared distance between the representatives of the cases in Figure 4 is the same as their Procrustes shape distance, while the representation is flattened to a linear space of the correct dimension. The $k$ coordinate pairs of that figure actually serve as a set of $2k$ redundant coordinates for this $(2k - 4)$-dimensional tangent space.

A tangent space can be imagined as a hyperplane touching a hypersurface, but the modern approach construes it instead as a linear space in its own right, the space of all linear germs of scalar functions along curves through the point at which it 'touches' the manifold. To the statistician, these functions are just what we mean by shape variables—algebraic functions of the coordinates that are invariant under changes of position, orientation or scale (Bookstein, 1991). That is, the tangent space construction of Figure 4 provides the setting for all possible linear multivariate analyses of the information in the shapes of the landmark configurations. Furthermore, the usual multivariate sum-of-squares metric of that representation is the underlying Procrustes metric of the manifold, and so the multivariate statistics will be commensurate with the original Procrustes geometry.

For shape variations generated by circular Gaussian variations of landmark location, whatever the means and however large that circular variance, the complete statistical theory of these shape representations is known exactly. Small (1996) is a good introduction to this distribution, the 'Mardia–Dryden distribution', which was first announced in 1989, and Goodall (1991) evaluates the associated multivariate normal approximations. For shapes that are concentrated in a small region of the full shape space (as any within-species sample of the shape of an organ is likely to be), one can carry out most of the ordinary maneuvers of multivariate statistical analysis—tests of group differences, correlations of shape with causes or effects, principal-components analysis with respect to Procrustes distance—directly in terms of this redundant basis for the tangent space. The suitability of these linear approximations is very high (Bookstein, 1991; Kent, 1995). We will

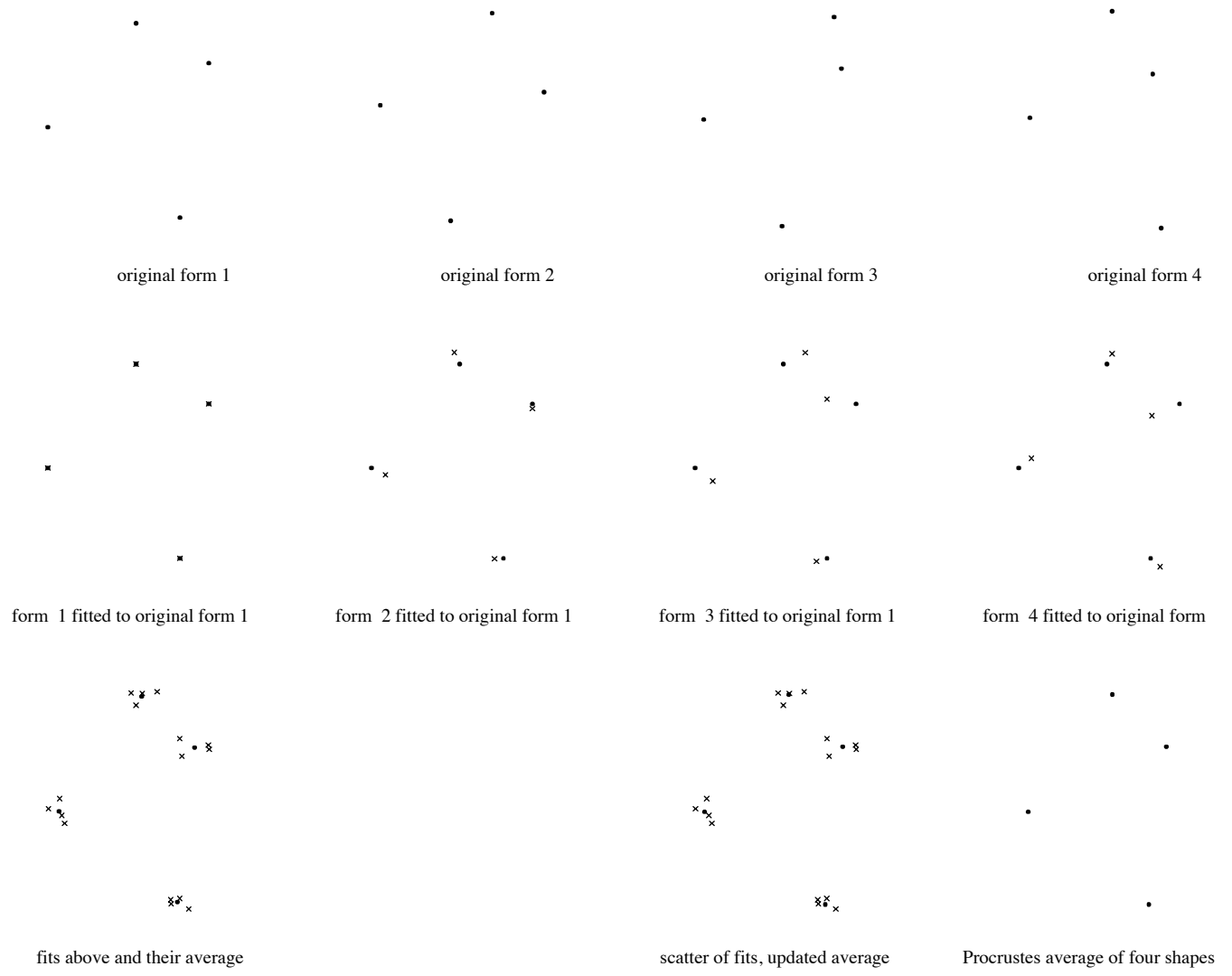


**Figure 4.** Procrustes averaging and Procrustes shape coordinates. Top row, four forms of four landmarks. Middle row, Procrustes fit of each ($\times$s) to an arbitrary starting guess (dots: the first form). Bottom left, the next estimate of the average (dots) is the average of the fitted locations from the previous step. A second round of fits and averages changes it hardly at all—the algorithm seems to have converged already. Bottom right, the Procrustes shape coordinates ($\times$s) are the locations of the landmarks after the fitting step upon the average shape (dots) at the convergence of the algorithm. These actually represent the orthogonal projection of the sample onto the tangent space to Kendall's shape manifold at the shape on the right (see text).

make use of this when we average shapes over groups in Figure 8.

In three dimensions the Kendall manifold lacks the symmetries of the two-dimensional setting and the mathematical statistics of all this is much less elegant; nevertheless three-dimensional versions of all these data-analytic maneuvers are known. For Procrustes distance see Rohlf and Slice (1990); for the tangent spaces and approximations see Goodall (1991).

### 2.3. The thin-plate spline for landmark points and its eigenspaces

The other foundation for the landmark methodology that this paper extends to curves is a very simple and useful visualization of shape difference as deformation: the thin-plate spline interpolant between two sets of landmarks. Let $U$ be the function $U(\boldsymbol{r}) = r^2 \log r$, and consider a reference shape (in practice, a sample Procrustes average) with landmarks

$P_i = (x_i, y_i),\ i = 1, \ldots, k$. Writing $U_{ij} = U(P_i - P_j)$, build up matrices

$$K = \begin{pmatrix} 0 & U_{12} & \ldots & U_{1k} \\ U_{21} & 0 & \ldots & U_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ U_{k1} & U_{k2} & \ldots & 0 \end{pmatrix}, \qquad Q = \begin{pmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ \vdots & \vdots & \vdots \\ 1 & x_k & y_k \end{pmatrix}, \tag{2}$$

and

$$L = \begin{pmatrix} K & Q \\ Q^{\mathrm{t}} & O \end{pmatrix}, \qquad (k+3) \times (k+3),$$

where $O$ is a $3 \times 3$ matrix of zeros. The thin-plate spline $f(P)$ having heights (values) $h_i$ at points $P_i = (x_i, y_i)$, $i = 1, \ldots, k$, is the function

$$f(P) = \sum_{i=1}^{k} w_i U(P - P_i) + a_0 + a_x x + a_y y \tag{3}$$

where

$$W = \left(w_1, \ldots, w_k, a_0, a_x, a_y\right)^{\mathrm{t}} = L^{-1} H \tag{4}$$

with $H = \left(h_1, h_2, \ldots, h_k, 0, 0, 0\right)^{\mathrm{t}}$. Then we have $f(P_i) = h_i$, all $i$: $f$ interpolates the heights $h_i$ at the landmarks $P_i$. Moreover, the function $f$ has the minimum bending energy of all functions that interpolate the heights $h_i$ in that way: the minimum of

$$\iint_{\mathbf{R}^2} \left( \left( \frac{\partial^2 f}{\partial x^2} \right)^2 + 2 \left( \frac{\partial^2 f}{\partial x \partial y} \right)^2 + \left( \frac{\partial^2 f}{\partial y^2} \right)^2 \right) \tag{5}$$

where the integral is taken over the entire picture plane. The value of this bending energy is

$$\frac{1}{8\pi} W^{\mathrm{t}} H = \frac{1}{8\pi} H_k^{\mathrm{t}} L_k^{-1} H_k, \tag{6}$$

where $L_k^{-1}$, the bending energy matrix, is the $k \times k$ upper-left submatrix of $L^{-1}$ and $H_k$ is the corresponding $k$-vector of 'heights' $\left(h_1, h_2, \ldots, h_k\right)$.

A plane-to-plane interpolation is couched as a Cartesian pair $(f_x, f_y)$ of these functions based in the same matrix $L$ and for which $f_x$ uses a vector $H_x$ of $x$-coordinates of a target form and $f_y$ uses a vector $H_y$ of $y$-coordinates. The bending energy being minimized is now the quadratic form $H_x^{\mathrm{t}} L_k^{-1} H_x + H_y^{\mathrm{t}} L_k^{-1} H_y$. Examples of these grids, applied to semi-landmarks instead of landmarks, can be seen in Figures 5, 9 and 10.

The spline helps visualize statistical summaries based on these Procrustes coordinates in a remarkably effective manner. Any vector computed in the course of a multivariate analysis—a mean difference, for example—can be visualized as a deformation in this way. That the integral in Equation (5) is minimized means that the spline fits shape changes with the smallest possible variation of affine derivative (shapes of the little grid cells in those figures). If a given change of affine derivative can be managed over a larger interval, its contribution to the integral of squares is lower; hence the spline tries to represent deformations 'as globally as possible'. In this way it leads the trained eye to a remarkably reproducible assessment of both global and local features of change. Eigenvectors of the bending-energy matrix $L_k^{-1}$ of Equation (6) can be shown (Bookstein, 1995b; Mardia, 1995) to be orthonormal in the Procrustes geometry of the tangent space, not just the Cartesian geometry of the original digitizing space. The eigenvectors emerge paired by any orthogonal pair of Cartesian directions. In general, they emerge in order of localization (Bookstein, 1991), but we will not need to exploit that fact here.

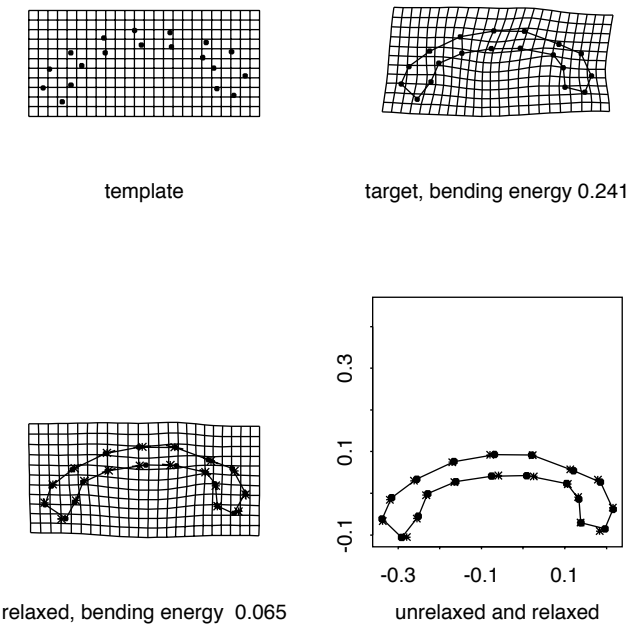


**Figure 5.** The geometry of relaxation using a thin-plate spline. Upper left, a 'template' or starting form of 19 points. Upper right, arbitrary set of pre-assigned homologues, with the ordinary thin-plate spline interpolating these point-pairs. Lower left, landmarks after a relaxation along escribed chords (of which short segments are shown) through starting positions. Lower right, the principal effect of the relaxation is the counter-rotation of the opposing curve segments at the center, as viewed in the regularization of the grid at its center between the upper right and lower left panels. That the bending energy at small scales is so sensitive to digitizing noise is one reason it is not often used as a metric in the morphometric literature (see Bookstein, 1996a, 1997).

Uniform transformations (affine transformations, shears) enter into this series as a 'zeroth pair', corresponding to the the eigenspace of eigenvalue zero. Augmented thereby, the set of eigenvectors of bending energy rotates the redundant basis of $2k$ Procrustes shape coordinates into a complete orthonormal set of $2k - 4$ values. The associated vector decomposition, the components of which are called partial warps, decomposes any observed change uniquely into a sum of fundamental modes of variation based only on the average landmark shape, without any assumptions about material properties in between the landmarks. These and other helpful aspects of the spline are reviewed in Bookstein (1996a, 1997).

The three-dimensional version of this same thin-plate spline has been known for as long as the 2-D version. The kernel $r^2 \log r$ is replaced by $|r|$, so that the spline is no longer smooth at the landmarks. The matrix $Q$ of Equation (2) gains a fourth column of $z$-coordinates, the vectors $H$ and $W$ gain $(k+4)$th terms 0 and $a_z$, respectively, and the integral (5) now has six terms instead of three. The bending energy minimized by the spline is now $-\left(H_x^{\mathrm{t}} L_k^{-1} H_x + H_y^{\mathrm{t}} L_k^{-1} H_y + H_z^{\mathrm{t}} L_k^{-1} H_z\right)$.

## 3. SPLINE RELAXATION ALONG CURVES

In the application to landmark points, although the bending energy (6) for the spline is a global minimum of the integral (5), this property does not drive the calculus of minimization; it is a formal identity. Under other circumstances, however, the characterization can become an algorithm for optimization. In the present application it allows the spline method to be extended so that some of the target landmarks are freed to slide along lines (Bookstein, 1991). Let there be a 'nominal set' of right-hand landmarks $Y_1^0, \ldots, Y_k^0$ collected coordinatewise as the vector $Y^0 = (Y_{1x}, \ldots, Y_{kx}, Y_{1y}, \ldots, Y_{ky})$, the concatenation of the two vectors $H$ of the preceding treatment. We seek the spline of one set of landmarks $X_1 \ldots X_k$ onto another set of landmarks $Y_1 \ldots Y_k$ of which a sublist $Y_{i_1} \ldots Y_{i_m}$ are free to slide away from their nominal positions $Y_{i_j}^0$ along directions $u_j = (u_{jx}, u_{jy})$. It seems entirely natural to proceed by minimizing the corresponding bending energy.

That energy is the quantity $Y_x^{\mathrm{t}} L_k^{-1} Y_x + Y_y^{\mathrm{t}} L_k^{-1} Y_y$ where the landmarks $Y_{i_j}$ of the sublist range over lines $Y_{i_j}^0 + t_j u_j$. To minimize, collect the parameters $t_1, \ldots, t_m$ in a vector $T$ of length $m$. $T$ is the parameter vector over which we will minimize the energy of the corresponding spline. The minimization is easiest to notate if we collect the directional constraints $u_1, \ldots, u_m$ in a matrix of $2k$ rows and $m$ columns in which the $(i_j, j)$th entry is $u_{jx}$ and the $(k + i_j, j)$th entry is $u_{jy}$, otherwise zeros. The task is now to minimize the form

$$Y^{\mathrm{t}} \begin{pmatrix} L_k^{-1} & 0 \\ 0 & L_k^{-1} \end{pmatrix} Y \equiv Y^{\mathrm{t}} \boldsymbol{L}_k^{-1} Y \tag{7}$$

over the hyperplane $Y = Y^0 + UT$. The solution to this familiar generalized or weighted least-squares problem is achieved for the parameter vector

$$T = -\left(U^{\mathrm{t}} \boldsymbol{L}_k^{-1} U\right)^{-1} U^{\mathrm{t}} \boldsymbol{L}_k^{-1} Y^0. \tag{8}$$

Figure 5 demonstrates this computation for a little scheme of 19 'landmarks' slightly simpler than the real average form we will actually be using. The thin-plate spline from the template to the 'raw data' shows considerable local shear in the middle of the arch and at its tips. The relaxation allows points along the upper and lower arcs of this arch to slide in opposite directions from their starting positions (large $\times$s at lower left) to the positions that jointly minimize bending energy and allow the tips to relax to the left or right similarly. In this example, the directions $(u_{j1}, u_{j2})$ of relaxation are set to escribed chords $u_j = Y_{j+1} - Y_{j-1}$. (An escribed chord to a polygon is a line through a vertex parallel to the join of its two nearest neighbors. At convex vertices it lies outside the polygon.) This quick-and-dirty method works adequately except where the polygon is turning rapidly, for which more sophisticated procedures are required. Deviations of the target polygon in a direction normal to the template boundary are not relaxed, as one can see under the 'shoulders' of this form, where most of the remaining bending is to be found. In our example, all the landmarks were allowed to slide. But the representation of Equation (8) does not require $m = k$, so any subset of landmarks could be fixed in advance. We will not need that extension here.

The minimization in Equation (8) accords with the integral formulation (5): the relaxation will try to bring the variation of all first-order derivatives toward zero. Sometimes it can achieve that minimum. For instance, if the method is applied to a hexagon of landmarks, it will always arrive at a completely affine (linear) transformation, as in Figure 6. (The number of parameters that must be fitted is now reduced to six, which is the dimension of the complete space of affine transformations, and the machinery of Equation (8) is linear.) For more complicated data, the minimization will tend to cast changes of affine derivative at the largest spatial scales. It is therefore suited to multivariate statistical procedures that scan down a hierarchy of spatial scales, such as the nested $T^2$ convincing us of the significance of the group difference here (see below).

In three dimensions, the vector $Y^0$ is of length $3k$ and $\boldsymbol{L}_k^{-1} = \mathrm{diag}(L_k^{-1}, L_k^{-1}, L_k^{-1})$. The matrix $U$ has three noncontiguous rows per landmark and one or two columns per sliding landmark, as each is constrained to a curve (one degree of freedom) or to a surface patch (two degrees of freedom). In the former case, the entries $U_{i_j,j}$, $U_{k+i_j,j}$, $U_{2k+i_j,j}$ are the direction cosines of the tangent to the curve; for the constraint

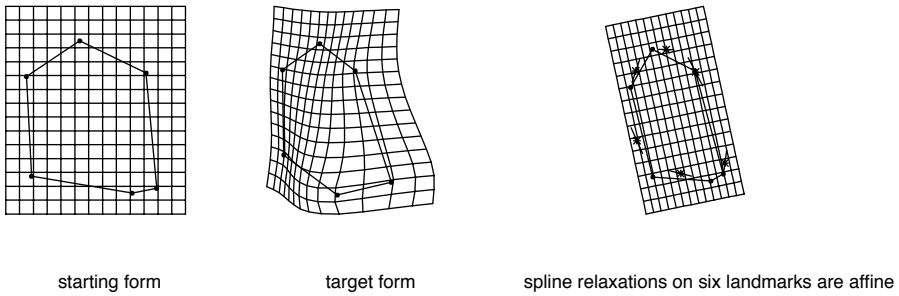


**Figure 6.** Spline relaxation of a hexagon of landmarks will almost always result in an exactly affine transformation: the quadratic form in Equation (7) can be zeroed.

to surfaces, the corresponding two columns of $U$ are any two perpendicular vectors in the (approximate) tangent plane. These features are built into the Green's program package `Edgewarp3D`, release of which is expected early in 1997.

The minimization in Equation (8) will generate a relaxation no matter what the geometry of the constraint vectors $U_j$, as long as the matrix $U^t \boldsymbol{L}_k^{-1} U$ is nonsingular. In Figures 5–7 these vectors were taken as finite differences $Y_{i+1} - Y_{i-1}$ corresponding to escribed chords of the polygon. These can lead to relaxations some distance from the starting outlines when forms are sampled at a spacing too sparse to suit the actual curvature. In Green (1995), they are instead taken as $Y_i - Y_{i\pm1}$, i.e. actual edge segments of the nominal starting polygon. It would be simple to improve this preliminary algorithm in several other ways, for instance, incorporating B-spline representations of the starting polygons instead of this crude $C^1$ construction, then iterating (Green, 1996) to ensure that points are clamped to the original outline representations. However, there is an additional step in the procedure which moots all these second-order optimizations of the single outline—the Procrustes averaging we have already introduced for landmarks, which aggregates over a sample of first-order variations.

### 3.1. Procrustes averaging of curves

For the application to curves, the Procrustes averaging algorithm of Figure 4 is altered by inserting an additional step corresponding to Figure 5 for every form in every cycle. The resulting iterative sequence is as in Figure 7. One must start with all shapes in some approximate registration, such as by centering and rotating to common principal moments. Upon a starting estimate of the average, perhaps the first form in the data set, one carefully assigns a reasonable number of points spaced roughly inverse to the outline curvature, then samples every other outline by a purely local geometric correspondence (such as by closest-edge projection). This initialization could be done automatically (W. D. K. Green, personal communication). The point count is maintained through all subsequent computations. Each outline of the data set is relaxed along itself to that starting set of points and the loci that result from the relaxation are Procrustes-averaged by the usual algorithm of Figure 4. The average that results is the starting form for the next round of relaxations. In practice this algorithm always converges. At convergence, the loci that represent each outline form of the data set by a selection of points from its edges correspond between every pair of individual forms of the data set by virtue of having arisen by spline relaxation from the same point of the final Procrustes average. They thus share some of the criteria that make landmarks useful for scientific analysis; here we call them semi-landmarks. They fail to be true landmarks in that in contrast to the requirements of Bookstein (1991) they cannot be defined on a single image—they exist only in the context of a group average. Of the two coordinates of each semi-landmark, one (the coordinate along the curve) is the variation of the 'definition'—this coordinate is set for all cases at the same time, early in the algorithm of Figure 7. Only the coordinate normal to the outline carries information about differences between specimens or groups.

Green (1996) shows how to extend the algorithm used here to any combination of open or closed curves and landmark points. The data of Figure 2 have already been preprocessed by this algorithm using the landmark-and-arc scheme of Figure 1. When there are reliable 'corners' of the curves, they are matched by the spline regardless of whether they appear in the slip list $i_1, \ldots, i_m$. (In effect, it costs less energy to bend one corner into another corner than to bend a flat spot into a corner while simultaneously flattening another corner nearby.) In other words, features that behave like good landmark points (Bookstein, 1991) are treated like points whether or not the investigator labels them as such in advance (cf. McEachen *et al*., 1994). This landmark–outline composite may already subsume a great many approaches suggested elsewhere for various special cases, including open curves

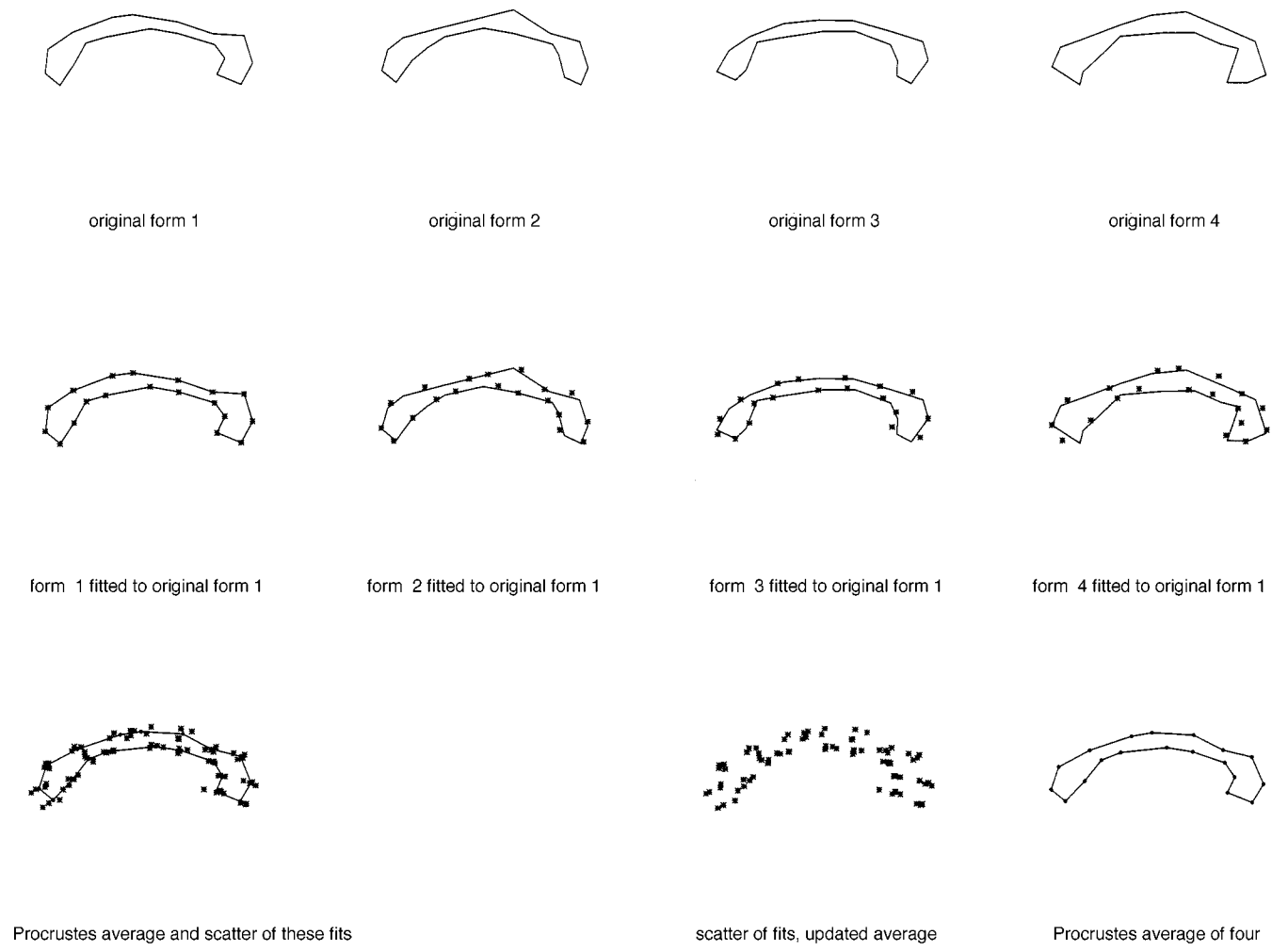


**Figure 7.** New algorithm for combining Procrustes averaging with spline relaxation. Top row, four 'original forms', generated by random variation around the template of Figure 5. Middle, each original outline is relaxed against an initial estimate of the average, in this case, the first original outline. Relaxation is along escribed chords as in the text. Bottom left, the Procrustes average of the relaxed forms from the preceding step becomes the next estimate of the average outline. Bottom right, the algorithm has converged by the end of the second iteration.

between two landmarks (Dean *et al*. 1996a, Sampson *et al*. 1996) and landmarks with directions through them (Bookstein and Green, 1993). Green conjectures that this extended algorithm can be expected to converge to useful representations of sample average shapes and variation around those averages in all of these extended settings.

## 4. PROCRUSTES DESCRIPTION OF GROUP SHAPE DIFFERENCES

The final step in the Procrustes approach to landmark statistics, Figure 4 lower right, represents shape variation by the fitted locations of each landmark when superposed by least-squares upon the sample average. This algorithmic step persists in the hybrid with the spline relaxation (see Figure 7). For this little data set of 26 semi-landmarks for 25 callosal shapes, there results the scatter of Procrustes shape coordinates at the upper left in Figure 8. (Remember that although this is a Procrustes scatter, it is taken around the spline-relaxed average.) The component variations 'at' the separate semi-landmarks are mostly well behaved and only modestly noncircular. The outlying points at the lower right center derive from one callosum of extremely recurved splenium, the form at the right in the third row of Figure 2.

As for landmarks, most multivariate analyses of the shapes of these outlines can be carried out directly on these Procrustes

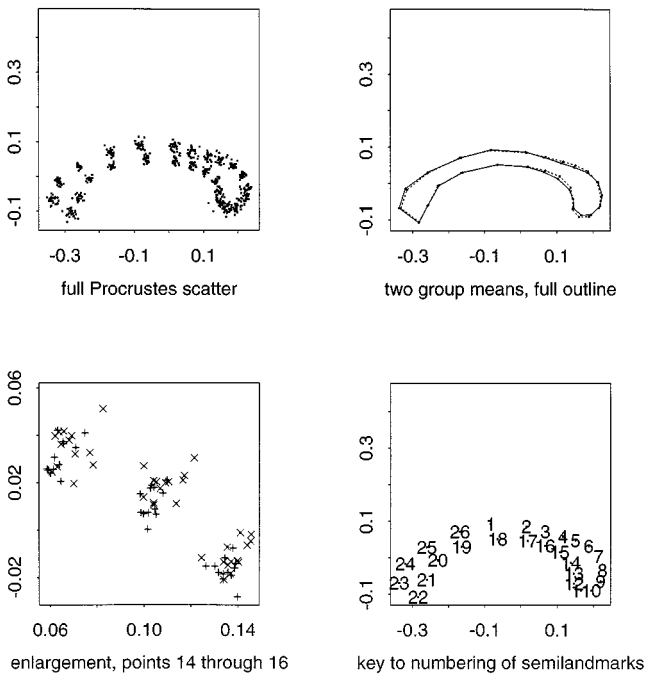


**Figure 8.** Procrustes analysis of the whole callosal form. Upper left, conventional scatter of Procrustes-fit coordinates (least-squares superposition to an average) at convergence, where it is the fixed-point of the slipping algorithm as well. Upper right, group mean contours in this registration. Solid line, normals; dashed line, schizophrenics. Lower left, enlargement of scatters at points 14, 15, and 16. ×, schizophrenics; +, normals. The difference at point 15 is significant at about the 2% level by ordinary $T^2$ test. Lower right, key to semi-landmark numbers referred to in the text.

shape coordinates. For this paper the simplest such analysis will suffice: the computation and comparison of two group average shapes. These averages (simple centroids by group of the points scattered in Figure 8, upper left) are shown in the upper right-hand panel. The largest displacement is at semi-landmark 15. The lower left-hand panel shows scatters by group at semi-landmarks 14, 15 and 16, the subarc that appears most interesting. In this registration, the difference at semi-landmark 15 (×s versus +s) is significant separately (by Hotelling's $T^2$ test on its pair of coordinates) at $p \sim 0.02$.

This is not evidence enough for a claim of statistical significance, of course, since we selected this point as the most interesting out of 26. The morphometric synthesis offers an assortment of global significance tests that do not require any selection of points or features (Bookstein, 1996c). One that is appropriate in this context is a nested series of Hotelling's $T^2$ tests on the partial warps derived from the bending-energy matrix $L_k^{-1}$ above. (The original data would not be amenable to ordinary $T^2$ testing at all: there are at least 26 degrees of freedom for shape, but only 25 specimens in the sample.) The group difference is orthogonally projected (in the Procrustes geometry) onto a series of subspaces of dimensions of the tangent space that do not exceed a certain degree of 'bending'. One first tests the 2-subspace (plane) of uniform (affine) changes, then the span of the uniform and the largest-scale partial warp, and so on until degrees of freedom are exhausted. The most significant of these tests is for the span of the uniform part and the first five warps, corresponding to vector multiples of the eigendeformations in Figure 9 along with uniform shears. Projected onto these six two-dimensional modes of deformation, the group difference is statistically significant by Hotelling's $T^2$ test at $p \sim 0.002$. As we had 11 of these tests to choose from (half the sample size, minus 1), the corrected probability is a little above 2%, which is, indeed, satisfactorily significant. The projected (filtered) mean outlines with a significant difference are shown at the lower right in the figure; compare with the upper right panel of Figure 8.

The same spline visualization that we used to relax individual outlines to the emerging mean outline shape can be used to represent the difference between the two group average shapes of Figure 8. The relation between them is shown in Figure 10 magnified by a factor of three (since the spline is a matrix manipulation, no iteration is required). This version of the finding is easier to report to a scientific audience than the statistical display of the same information in Figure 8. Its main features are quite clear: the average callosum of these schizophrenics differs from that of the normal group mainly in a substantial upward-rightward deviation of the isthmus together with a small diminution of genu size. There seems to be no alteration in size or shape of the main arc of the arch. The assumption that these averaged semi-landmarks can be treated as landmarks, without further slipping, is sustained by the plot of the 'forces' (coefficients $W_x$, $W_y$) of the mapping, shown at the right in the figure. Almost all of these lie normal to the average outline curve. This means that the comparison of averages is relaxed already, as it should be.

## 5. LOCALIZING CHANGES OF OUTLINE SHAPE

Figure 8 shows a displacement of semi-landmark 15 between the groups in the direction normal to the average boundary by 136% of the averaged within-group variance—a signal-to-noise ratio (S/N) of 1.36. But some of this noise can be circumvented. Its denominator (net within-group variation) incorporates displacements from the Procrustes optimization along other parts of the outline. For instance, variations in size of genu, the structure at the far left, will displace the centroid of the entire form in ways that have nothing to do with the shape difference in the vicinity of point 15. But

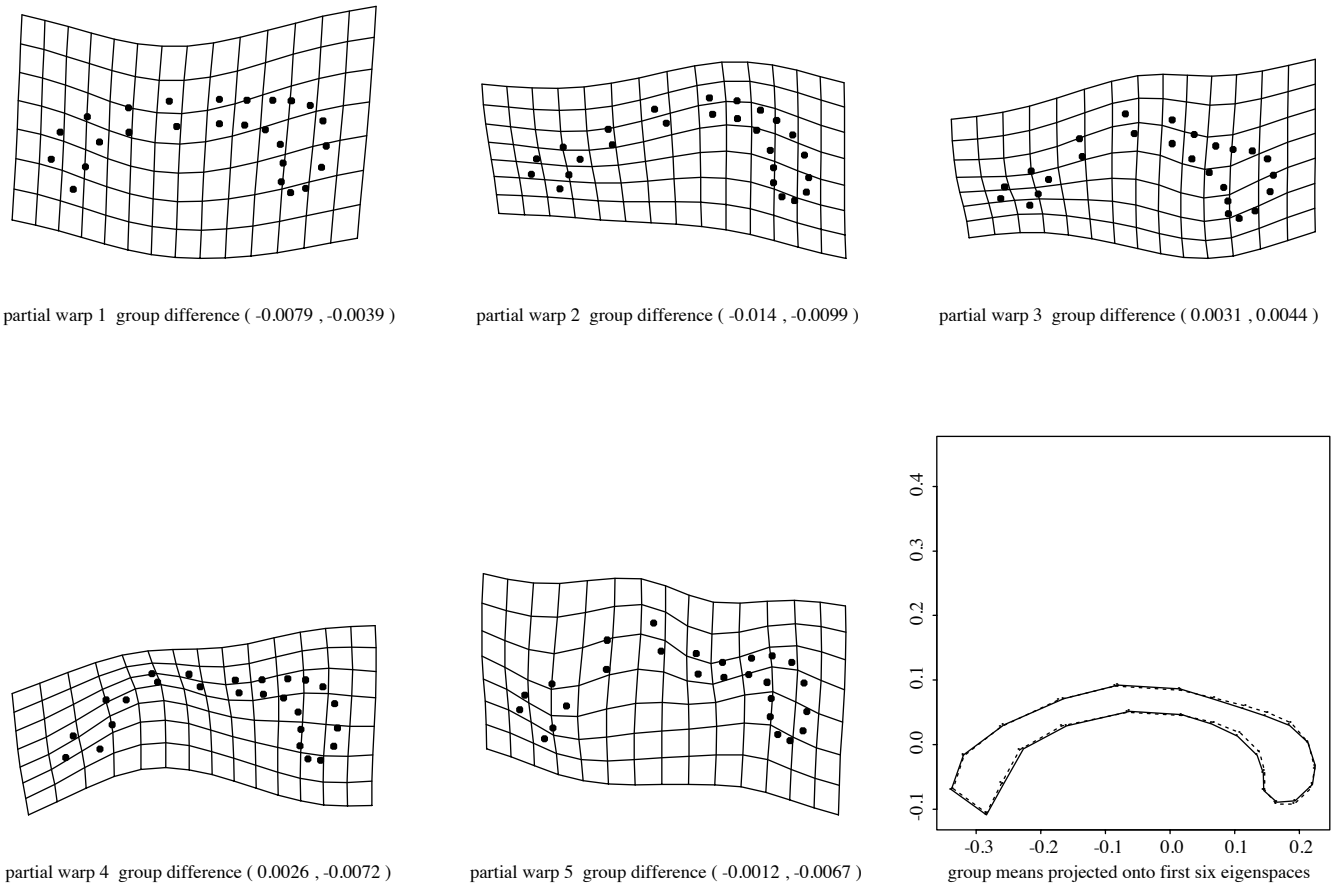


**Figure 9.** First five partial warps of the averaged outline (Figure 8) along with their components in the group difference there and (lower right) the low-pass-filtered difference analysis of this order.

because the Procrustes step recenters the whole form, and because the vector toward genu is oblique to the average outline tangent at point 15, this variation at genu will contribute to the denominator of the S/N ratio we have just inspected. The Procrustes analysis can also attenuate the numerator of the S/N ratio. In Figure 8, the systematic difference at genu has shifted the dashed (schizophrenic) form leftward, diminishing the shift at semi-landmark 15 that we are trying to assess. Since there are such group differences at the other end, it is best to exclude them from our judgment of what is happening at point 15.

We experiment with a different starting Procrustes fit: just half the form, the 'back half', as in Figure 11. (To divide into 'halves' this way is not an arbitrary choice. The susceptibility of Procrustes fits to large-scale changes is largest along the longer principal moment of the mean shape, which is obviously this horizontal direction. Empirically, the largest two principal components of shape are aligned with this contrast as well, so that to whiten the large-scale noise one would subset left from right in preference to any other statistical surgery.) This change does not involve any further spline relaxations. But the individual shifts and rotations to the average and the individual scalings by centroid size have been recomputed *de novo*. The net effect is to alter the superposition of the mean contours and the geometry of variation around them, while leaving their separate shapes pretty much unchanged. For this posterior part of the callosum, the variance of Procrustes shape coordinates is distinctly less than before. The group difference at semi-landmark 15 now has a S/N of 1.74, and the finding has a significance level (by Hotelling's $T^2$) of 0.0015 (which still needs correction by a factor of 26). A plot showing the original polygonal outlines, lower right, confirms this tendency of separation for the locus at which the outlines transect their approximate shared 'normal' near point 15.

Instead of trying harder and harder to find subsets of features onto which to project, we might try another strategy, the

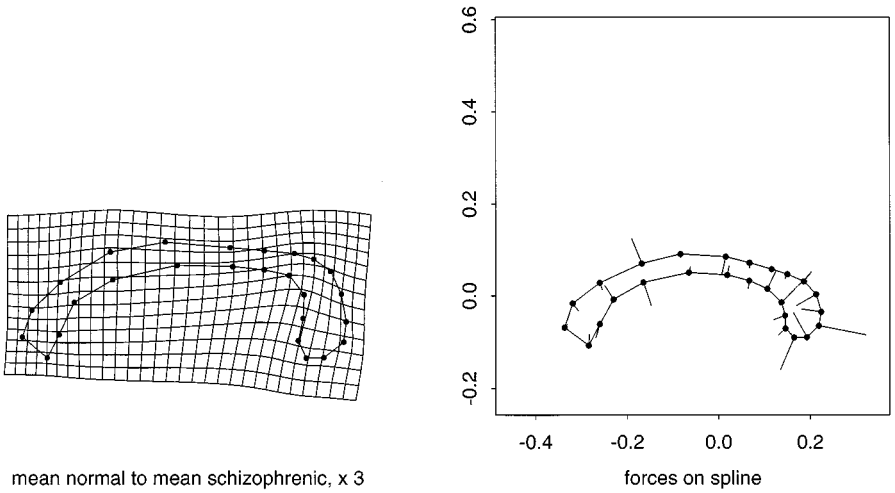


**Figure 10.** Thin-plate spline for the comparison of group mean semi-landmark shapes. Left, grid for the threefold magnification of the change from normal mean to schizophrenic mean. Right, coefficients of the spline, showing perpendicularity to the mean curve at points of low curvature.

omnibus test, that accepts features at all scales simultaneously, but requires that they be combined according to an *a priori* formula (in this case, the usual Procrustes sum of squares). If our data came in the form of the circular isotropic Gaussian variances of the mathematical development, we could use Goodall's (1991) equivalent $F$-distribution. But for data as highly nonlinearly processed as these slipped spline residuals, there is no 'distribution' available. We exploit, instead, a permutation test (Good, 1994; Bookstein, 1997). In a permutation test, one chooses a statistic (in our case, summed squared Procrustes distance normal to the average curve, weighted inversely by the pooled variance of that distance) and computes its distribution when the true group assignment is replaced by many hundreds or thousands of random permutations of case number over group. The permutation process makes no assumptions whatever about the structure of variation across the landmarks within the full pool of 25 cases; it is concerned only with the relevance of the group label to the average shape. For each such permutation, there is a squared Procrustes distance between group means. The appropriate tail probability for the hypothesis of a group mean difference in shape, in the metric of this Procrustes sum of squares, is exactly the fraction of times a random permutation results in a group mean Procrustes distance larger than the one afforded when the group labels were taken at their true values.

For this posterior half of the form, this test reports a significance level of 1.4%. The $p$-value must be corrected for the number of degrees of freedom it represents from the shape space as a whole. Here, the fully-slipped data set bears 26 degrees of freedom (one per landmark) and the 'splenium half' of the form is tested between landmarks 3 and 15, which is 13 of those 26; so the correction is by a factor of 2. (The remaining degrees of freedom are 11 for the other 'half' of the form and 2 for the relation between the two halves.) This results in a significance level of 2.8%, which is in adequate agreement with the implications of that single $T^2$ at semi-landmark 15. The same test applied to the outline as a whole is too badly confounded by large-scale variation to be usable in this omnibus form.

### 5.1. An adaptive high-pass filter for shape differences

What this development hints at can be restated explicitly in a suggestive way. The Procrustes-based $T^2$ at one landmark is working as a signal detection filter at the smallest scale, the scale of one landmark at a time. It thus complements the analysis in Figure 9, the low-pass analysis exploring the spectrum of the group difference from its global end. The signal being sought, a shift of one semi-landmark on a background of a subset of the others presumed unchanging in mean position, is a special case of the 'resistant residual' familiar to the morphometric community (Rohlf and Slice, 1990). Any mean shift signal will be passed through the 'Procrustes projection' (Kent, 1995) like any other quasilinear shape signal. This projection attenuates more or less as the landmark suspected of shift lies near to or far from the centroid of the mean Procrustes form. Thus we should be searching in neighborhoods centered over the target landmark. The preliminary spline-relaxation step has augmented the power of this filter in a very suggestive way. Because of the strong anisotropy of its spectrum, one can assume any local changes along the curve

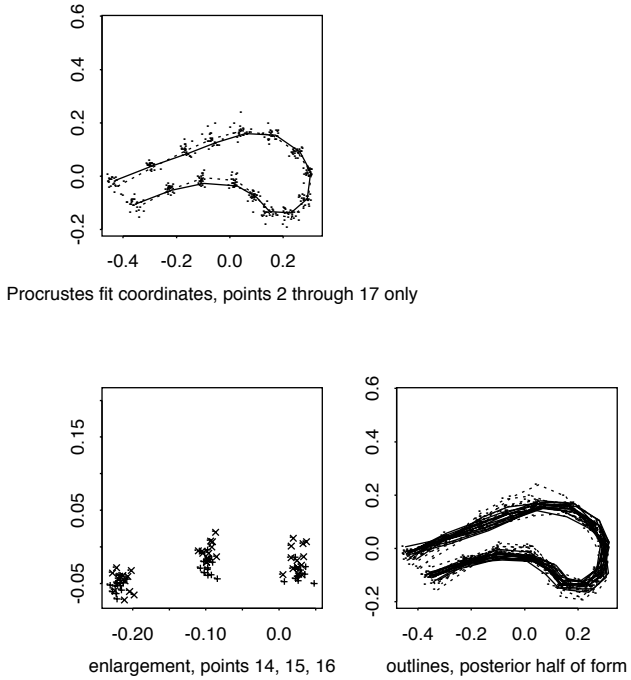


**Figure 11.** Procrustes analysis for the posterior half of the form, rotated to its own Procrustes horizontal. Upper left, means and Procrustes-fit coordinates. Lower left, enlargement for points 14, 15, and 16. As a multiple of its own standard error, the group mean difference at point 15 in the direction normal to the outline is 30% larger than it was in Figure 8. Lower right, superposition of original outline polygons indicates the extent of separation at point 15 normal to the shared tangent direction.

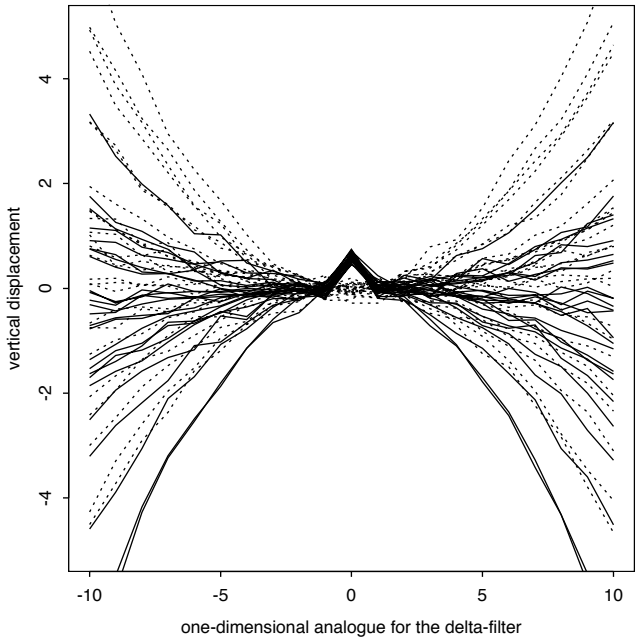


**Figure 12.** One-dimensional analogue for the claim that the $\delta$-filter will be complementary to the large-scale analysis of shape difference provided by the spectrum of the thin-plate spline. The functions here are a superposition of uncorrelated noise of variance 0.01 and multiples of $x^2$ by Gaussians of variance 0.001 over a mean shift of 0.6 at the single argument 0 (obvious sharp peak at the center). The two groups are shown as two different types of line (dotted and solid). The text derives the optimal neighborhood width for detecting the peak at 0. This optimal interval is $(-4, 4)$, smaller than one might expect, owing to the pernicious effect of the long-range order in the quadratic term.

to be negligible by comparison with those across the curve. Hence the Procrustes filter needs to be read only in one single direction, the direction normal to the mean curve.

The complementarity of these two filters can be usefully explored in the one-dimensional equivalent shown in Figure 12. Instead of outlines, consider two samples of functions over a wide interval of integers centered at $x = 0$. In this simulation, the function has an expected value of 0 at every value except $x = 0$; there, one group has an expected value of 0, the other, some nonzero value $\alpha$ (here 0.6). The observations vary around this expected value by the sum of two processes of quite different spatial scales. One is a completely uncorrelated process of Gaussian noise i.i.d. $N(0, \sigma^2)$, at each point of the domain. The other is a process of very long-range order indeed—for each case an independent normal deviate $\epsilon_i$, distributed as $N(0, \tau^2)$, multiplying an exact parabola $y = x^2$ applying throughout the domain. (Here $\sigma^2 = 0.01, \tau^2 = 0.001$.) The figure displays one realization of this composite process for two groups of 25. The small-scale process is intended as the analogue of ordinary edge-location noise, while the long-range order corresponds to large-scale shape differences or signals from other parts of the tissue under study.

By analogy with our Procrustes-based search for a normal deviation, in this set-up we apply a filter that contrasts the value of the function at 0 to its average value across some neighborhood of zero. This is the '$\delta$-filter' with coefficients

$$(2m)^{-1}\big(\overbrace{-1,\ldots,-1}^{m},\ 2m,\ \overbrace{-1,\ldots,-1}^{m}\big)$$

centered at 0. As $m$ varies, the filter will supply an estimate $\hat{\alpha}$ of the mean shift $\alpha$ at 0 whose error variance is a function of $m$ having a proper minimum somewhere. In fact, the variance of the filter output for the single case is $\sigma^2 + (2m)^{-2}\big(2m\sigma^2 + m(m+1)(2m+1)\tau^2/3\big)$, which is minimized for $m = \sqrt{\frac{3}{2}\frac{\sigma^2}{\tau^2} + \frac{1}{2}}$. For the values of $\sigma^2$ and $\tau^2$ here, this

yields $m \sim 4$: an interval less than half the width of the plot around the target point 0.

Back in the domain of morphometrics, this model of quadratic growth (the function $x^2$) for long-range or large-scale effects is qualitatively quite plausible—growth gradients, the largest partial warps, and other realistic candidates have always been modeled by such supralinear scaling (Bookstein, 1991, Sections 7.3 and 7.4). If local shape differences are also present, they might be expected to be found as optima of the two-dimensional version of the $\delta$-filter at some intermediate window size. Furthermore, because the spline relaxation of every specimen curve against the average preferentially filtered out tangential differences of group mean shape within 'small' neighborhoods, it is fair to characterize candidate findings by the statistical signal normal to the average curve, which is to say, by exactly the S/N ratio of the two experiments above.

We are thus led to the analysis shown in Figure 13: a search over $k(k-2)$ potential neighborhoods in search of local maxima of S/N ratios for the group difference signal normal to the mean curve. We hope for a 'corrugated' surface, with maxima that are sharply a function of position on the outline (semi-landmark number) and only very mildly a function of neighborhood size. By analogy with Figure 12, the neighborhoods are centered at the landmarks at which S/N is computed. The list of candidate neighborhoods is then just the union of all the lists of $j$ nearest neighbors of each semi-landmark in turn, $j = 2, 3, \ldots, k-1$. (The hierarchy of neighborhoods is computed from the Procrustes average shape.) For legibility, the figure begins at $j = 5$. The values plotted at the right ends of these traces correspond to the neighborhoods of size 26 (which are all the same 'neighborhood', the whole set of points—the numerators of these ratios are the displacements displayed in Figure 8). The figure suggests another global significance test for the net shape difference: the univariate $t$-test of group difference in the average shift of semi-landmarks 14 and 15 in the original Procrustes registration. That $t$-ratio is 4.0, significant at $p \sim 0.0005$. Multiplied by 26 (the number of such segments), that is a probability of 1.3%, comparable with the figure of 2% which we arrived at by the complementary filter using the hierarchy of eigenspaces of bending energy. It is not necessary to fit all of these windows, as the dependence on scale (Figure 13) is quite smooth, but there seems to be no harm in doing so.

This novel diagram is remarkably informative about aspects of the data omitted from Figures 8 or 10. Clearly the comparison of callosal form between the groups shows just one region of localized shape change, the arc 14–15 we have already noticed in the global splines. The single small region of best S/N is the neighborhood of size eight around semi-landmark 15, for which the S/N is 1.76 and the directional $t$-ratio a satisfactory 4.4, $p \sim 0.0002$. Multiplied by 26 (points) and further by a factor of $\sim$3, the effective dimensionality of the curves in Figure 13, this is about $\sim$2% again, the same range at which we have arrived by two earlier methods.

The Procrustes analysis that produced this S/N, the corresponding superposition of mean outlines and the classification afforded thereby are all collected in Figure 14. Our two complementary filters arrived at the same conclusion about the shape change. It is mainly a rearrangement of the isthmus between splenium and arch. At large scale (Figure 10) this has the appearance of a displacement; here at small scale, the appearance is instead of a sculpting or bending. At larger scales, according to Figure 13, semi-landmark 14 is a bit more informative separately, but the corresponding visualization is already apparent in Figure 10. Although two other semi-landmarks, 2 and 11, seem to embody some small-scale information, clearly the S/N surface is 'corrugated'. Just as we had hoped, it expresses widely spaced local effects. The output of this center-versus-periphery Procrustes filter is very robust to variation of neighborhood size but highly sensitive to position along the curve.

At the lower left in Figure 14 is the classification afforded by this optimal local filter. Classification is clearly improved over that in Figure 8, reproduced at lower right, by attending to the 'resculpting' of the isthmus. This is in spite of the difficulty in digitizing this arc by hand. (In many original images the fornix springs from this arc to the left, requiring a subjective decision as to where the underlying outline should be located. It was estimated, of course, without any knowledge of diagnosis.)

The model underlying the filter here contrasts strongly with the basic distributional model for Procrustes-type analyses of discrete landmarks. Our standard landmark methods (Bookstein, 1995–97) cast a 26-'landmark' data set like this one into a 48-dimensional space of shape phenomena within which a strenuous symmetry is enforced *a priori*. Variation tangential to and normal to outlines must be weighted equally and group differences are modeled as equally plausible regardless of the direction along which they lie in this space. For instance, every set of three landmarks, neighboring or not, is considered an equally likely locus of some potentially crucial aspect of shape discrimination. In this context, the best single overall multivariate test statistic is Goodall's (1991) omnibus $F$-ratio (see also Bookstein, 1996a, 1997), which is a multiple of the ratio of Procrustes distance between the group means to the Procrustes variance within the groups separately. For the data here its value is 1.32 on 48 and 1104 d.f., significant at $p \sim 0.07$ only. That was, in essence, a single test of the displacement between mean curves in Figure 8 with respect to the net degree of variation around those means aggregated over all the semi-landmarks separately. Because the $F$ could not take into

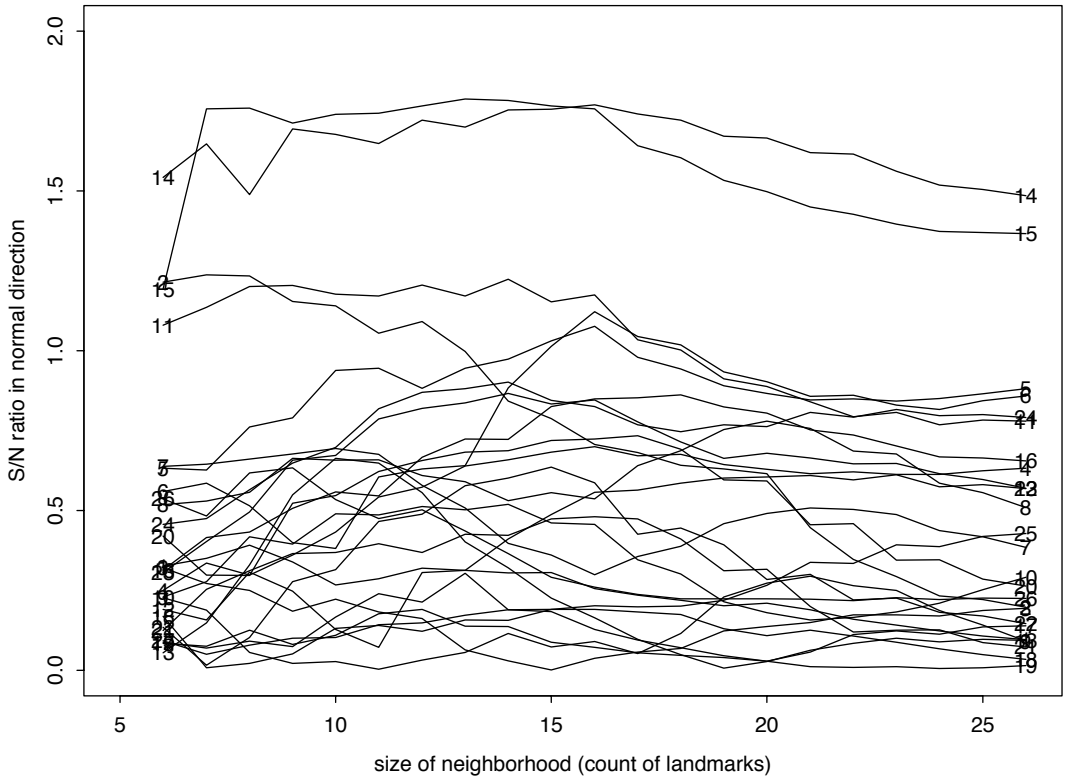


**Figure 13.** Performance of the $\delta$-filter at all semi-landmarks and all neighborhood sizes from 6 through 26. The S/N plotted is in the direction normal to the mean outline. Semi-landmarks are numbered at both ends of the trace of their S/N ratios over the range of neighborhood sizes. The chart argues for a signal at small scale restricted to the vicinity of semi-landmarks 14 and 15 (see text).

account the concentration of the finding in both direction and location, it was blocked from finding anything particularly implausible about the configuration of group differences as a whole.

An earlier attempt to localize the shape difference between these same callosa (Bookstein, 1995a, 1996b) exploited a highly irregular landmark analysis to arrive at the same finding. From a preliminary landmark-based image averaging of the region of callosum, a shift was found between the average lower borders. A new landmark was constructed on the individual border as the intersection of a straight line from 'splenium' to 'genu', where these two notions are now construed as points rather than regions. Analysis of the proportion into which this new landmark divided the diameter of the callosum resulted in a finding about as strong as that in Figure 8. In the alternate method, the equivalent of the $\delta$-filter is a consideration of the shapes formed by this new landmark with respect to the two or three landmarks nearby. Addition of just one more point, on colliculus, resulted in a finding with about the same power as that in Figure 14 here. However, as several colleagues have stubbornly pointed out quite sharply, this hybrid technique corresponds to no formal theory of morphometrics and affords no significance tests.

## 6. RELATED APPROACHES

The adaptive filter optimization here is partly analogous to two otherwise disparate techniques that have been the subject of previous discussions in the literature of medical image analysis. Techniques of anisotropic diffusion (Perona *et al*. 1994; Whitaker and Gerig, 1994) modify images toward the equilibration of picture content within regions without allowing information to diffuse across sharp boundaries. They thus achieve part of the purpose of the processing needed here—an enhancement of the signal normal to the boundary, by flattening gradients elsewhere—without the corresponding relaxation along the boundary that seems the key to the success of our filter. That lacuna is understandable, of course, in that these diffusion techniques deal with analysis of pictures

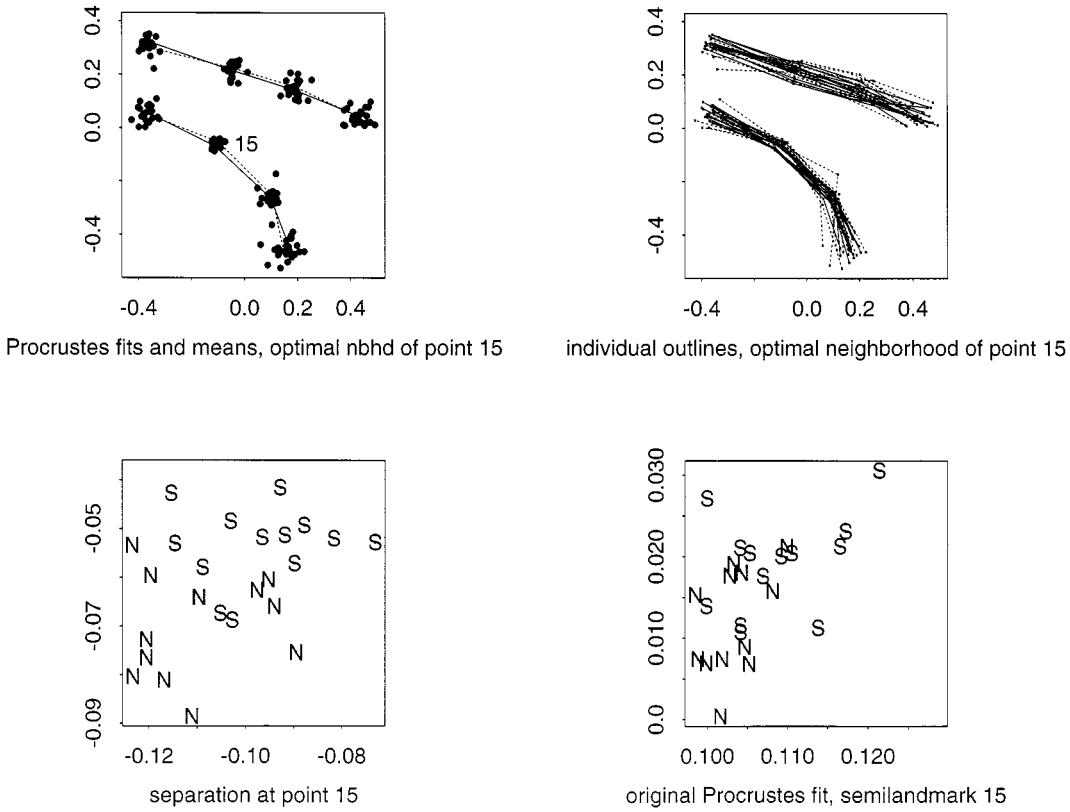


**Figure 14.** Analysis for the optimal neighborhood around point 15. Upper left, group mean outlines and Procrustes scatter. Upper right, individual outlines. Lower left, output of the $\delta$-filter at point 15. Lower right, comparable output of the full Procrustes fit (Figure 8) with which we began. N, normals; S, schizophrenics. All panels are in the orientation of the entire form, Figure 8.

one at a time, whereas what is relaxed by the spline is not a property of any single image, but of a pair or a whole data set of instances in relation to a template or an average.

The $\delta$-filter method can be considered as a limiting case of a contrast-of-annuli filter. In light of the strongly anisotropic nature of the preliminary spline processing, these annuli or Gaussians are better taken not as circular but as severely elongated in the direction of the edge under study. The combined method here can be thought of as an even larger-scale adaptation of such contrasts to the vicissitudes of large-scale curvature at some distance—it is as if the edge were straightened before these parametrically simple filters were applied. The large-scale approach of Figure 9 is thus complementary to this entire family of filters; cf Coggins and Huang (1993). Had we chosen a great many more points on these outlines, perhaps 260 rather than 26, we would experiment with alterations of the inner radius of the Procrustes filter, to allow the consideration of more than one 'center' at a time. For the equivalent in the one-dimensional analogue of Figure 12, see Bookstein (1996d).

But the $\delta$-filter is not at all closely related to other extant methods of multivariate statistical analysis of curving form, which share most of their spectral characteristics with the complementary, large-scale landmark technique introduced in connection with Figure 9. In the hands of other multivariate statisticians, samples of outlines have almost been analyzed by one or another decomposition into orthogonal series [Nastar and Ayache, 1994; Cootes and Taylor, 1996; Sampson *et al*., 1996; see Rohlf (1990) for a review of the classical work]. Sometimes these are geometric orthogonal functions, eigenfunctions of energetics of the mean form (Fourier decompositions, 'normal modes of bending') and sometimes they are empirical orthogonal functions (principal components of covariations of position). In the former approach, features at small scale are required to be orthogonal to those at large scale, so that the last few terms of any decomposition look like high-frequency oscillations around the average outline and thus cannot localize. In the latter approach, small effects on large regions of outline dominate large effects on small regions, as was the case for the

callosal outlines here, and so a local signal that was detected at all would already have been low-pass filtered—it would be visualized at too large a spatial scale to be informative. The finding at semi-landmarks 14–15 here suits none of these function-space approaches. If it is developed further, a wavelet decomposition of curving form such as that suggested by Ackroyd and Mardia (1996) may prove complementary to the large-scale approaches via orthogonal decomposition. But because wavelet decompositions are, in general, not invariant against the Euclidean similarity group, considerable modifications will be required before there can be a wavelet description of any scientific phenomenon like the local boundary shift that has proved so robust a finding here.

Tagare *et al*. (1995) have suggested a curve-to-curve matching protocol optimizing a functional that curiously complements this slipping spline. In effect, the correspondence they compute minimizes the integral along the curve of a sum-of-squares that combines acceleration or deceleration along the curve together with differences of bending in the normal direction. When the weight of this second term (which serves as a regularization) is set suitably, the combination is nearly the same as the first and third terms in the integrand for bending energy, Equation (5). But in the spline method the integral is taken over the entire picture plane and so allows aspects of the global differential geometry of the curves to affect the correspondence. It is the linearization of that global aspect, not the local, that is most conducive to subsequent statistical analyses.

Davatzikos *et al*. (1996) studied callosal outlines similar to ours that were acquired by an automatic active contour method. Outlines of a sample are related to an *a priori* norm by an elastic relaxation. The relation of each individual to the norm is then described by an areal distortion function, in effect the area of the little squares in our spline figures. Groups (in their example, eight male and eight female elderly from Baltimore) are compared by averaging this derived quantity at corresponding points of the normative form or over regions, and thresholding those differences at various effect sizes.

Davatzikos was kind enough to send me his data set for reanalysis by the method of the present manuscript. The permutation test of spline-slipped coordinates for the posterior 'half', this time comprising 50 semi-landmarks (for 16 cases!), finds a shape difference between the sexes that is significant at $p \sim 0.015$ before the correction factor of 2 is applied. The difference is primarily a vertical extension of the splenium. Davatzikos *et al*. (1996) found a change of area in this region, but did not report any directionality and apparently could not test for statistical significance in a sample so small. Indeed, they offered no graphical representation of sample variation at all.

What differences between the methods might account for this difference in power? Both algorithms have a matching step in which points of a candidate callosal outline are assigned to points of a standard outline by minimizing a formulation in terms of energy and both 'energies' are abstract, nonphysical. But in general, description by the areal-ratio scalar, like any other scalar description, uses one parameter (here the determinant) from the three that together describe the tensor field of affine derivatives, and so discards more than half of the information available for discussions of variability. Restoring it in the form of those tensors, furthermore, is not a very good idea (Bookstein, 1991). Statistical analysis works much better using the more linear spaces of shape features reviewed here. If transformations are uniform, specialized Procrustes methods are optimized for their estimation, comparison and testing (see Bookstein, 1996a, 1997). Visualization by areal ratio following elastic matching is tuned to one particular form of shape change: relatively homogeneous increase in area over compact regions that the elastic model treats in a somewhat homogeneous fashion. In the present data set, for instance, changes of area might partially describe a reshaping of splenium, but would not detect the shift of isthmus strongly to the upper right, which is our principal finding.

Of course, the initial steps in the Davatzikos procedure, active contour analysis and elastic relaxation, are familiar aspects of the analysis of individual images. In particular, the active contour method is obviously better than manual digitizing of the outline of the callosum where it blurs into neighboring structures (such as the fornix mentioned above). In this context, standard imaging tools can adequately simulate the performance of a skilled human. They do not extend to the further task of biometric analysis, however, at which unaided humans typically do not do very well either. Our Procrustes tactics and the thin-plate spline, while simpler, algebraically totally explicit, and exactly suited to the specific task of quantitative group comparison, are familiar mainly to the evolutionary biologist. They are not to be found in graduate curricula for imaging or in textbooks and their sources in the refereed literature were, until recently, quite sparse. But they can find scientific signals that are out of reach of familiar methods tuned to other purposes, such as object recognition or image segmentation. There is no particular reason to expect methods suited for one task to be suited for another, any more than the biometric techniques of this paper would bring any particular efficacy to the context of segmentation.

## 7. CONCLUDING REMARK

The finding here is obviously tentative. The callosum is a complex three-dimensional structure not adequately sampled by its midsagittal section, certainly not the crudely positioned

approximations here, and our 13 patients were not of any homogeneous clinical class. Instead this presentation was intended as a prototype for new and potentially powerful methodological possibilities.

To be able to treat curving form by landmark methods without requiring their acquisition in landmark-anchored form is a significant step forward in the morphometrics of biomedical images. For many three-dimensional data sets, for instance, representation of high-contrast surfaces works well by the method of ridge curves or crest lines (Thirion, 1994; Dean *et al*., 1996a), which extracts reliable three-dimensional loci at which the surface is locally most like a sharp edge. Statistical methods for analysis of such data have hitherto either been limited to very simple parametric models—rigid motion, polynomial warping—or have required the extraction of specific landmark points specimen by specimen (typically, curvature maxima along these curves). Such landmarks are typically much noisier than the arcs on which they lie. The combination of the spline relaxation and the associated multiscale multivariate statistics makes it possible to analyze such curves as a whole, by assigning semi-landmarks that describe the relation to a template without requiring any geometric semantics of characterization upon the individual form. In this aspect these new methods are converging with other approaches to form-comparison from McGill, Washington University and elsewhere, in which parametrization of the individual form is subordinated to parametrization of the relation to a template. For landmark data *sensu stricto*, these parametrizations are identical *a priori*: the points of shape space serve equally well as vector specifications of the corresponding splines (Bookstein, 1996a).

The spline relaxation and mixed multivariate methods used here will sustain valid findings to the extent that the multiscale model driving them is an adequate preliminary representation of the information discriminating the image groups under study: discrete group differences of either large or small scale. In this aspect it joins the mainstream of morphometrics to the most active current in analysis of medical images separately. Morphometrics can thus contribute to medical image analysis as a method for multiscale image comparison, a theme absent from many other current applications of the multiscale theme. [It does, however, appear in the McGill approach to atlas matching: see Evans *et al*. (1996).] This approach may, for instance, be helpful in studies of myocardial wall motion, the effect of disease on which has often been modeled as local defect of 'motion' with respect to a globally twisting shape change at quite large scale (e.g., McEachen *et al*., 1994). It may also be helpful in more detailed consideration of the shapes of fluid-filled cavities such as the cerebral ventricles (Dean *et al*., 1996) or other complex shapes having subtle but crucial functional implications, shapes that have hitherto proved difficult for analyses that followed conventional image-by-image processing.

## ACKNOWLEDGEMENTS

Preparation of this contribution was supported by NIH grants DA-09009 and GM-37251 to Fred L. Bookstein. The former grant is jointly supported by the National Institute on Drug Abuse, the National Institute of Mental Health and the National Institute on Aging as part of the Human Brain Project. All the statistical graphics were produced in the Splus package available from MathSoft, Inc., Seattle. The callosal outline data were acquired by Bill Green using an experimental version of his program package `edgewarp` that will eventually be available by FTP from the directory `pub/edgewarp` on `brainmap.med.umich.edu`. I thank Christos Davatzikos for access to the callosal outlines from his 1996 publication.

## REFERENCES

Ackroyd, R. G. and Mardia, K. V. (1996) A MCMC approach to wavelet warping. In Mardia, K. V., Gill, C. A. and Dryden, I. L. (eds), *Proceedings in Image Fusion and Shape Variability Techniques*, pp. 129–140. Leeds University Press, Leeds.

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

Bookstein, F. L. (1995a) How to produce a landmark point: the statistical geometry of incompletely registered images. In Melter, R., Wu, A., Bookstein, F. and Green, W. D. K. (eds), *Vision Geometry IV. SPIE Proc*., Vol. 2573, pp. 266–277.

Bookstein, F. L. (1995b) Metrics and symmetries of the morphometric synthesis. In Mardia, K. V. and Gill, C. A. (eds), *Proceedings in Current Issues in Statistical Shape Analysis*, pp. 139–153. Leeds University Press, Leeds.

Bookstein, F. L. (1996a) Biometrics, biomathematics, and the morphometric synthesis. *Bull. Math. Biol*., 58, 313–365.

Bookstein, F. L. (1996b) Biometrics and brain maps: the promise of the morphometric synthesis. In Koslow, S. and Huerta, M. (eds), *Neuroinformatics: an Overview of the Human Brain Project. Progress in Neuroinformatics,* Vol. 1, pp. 203–254. Lawrence Erlbaum, Hillsdale, NJ.

Bookstein, F. L. (1996c) Combining the tools of geometric morphometrics. In Marcus, L. F., Corti, M., Loy, A., Naylor, G. J. P. and Slice, D. E. (eds), *Advances in Morphometrics. NATO ASI Series A: Life Sciences*, Vol. 284, pp. 131–151. Plenum, New York.

Bookstein, F. L. (1996d) Landmark methods for forms without landmarks: localizing group differences in outline shape. In Amini, A., Bookstein, F. and Wilson, D. (eds), *Proc. Workshop on Mathematical Methods in Biomedical Image Analysis*,

San Francisco, pp. 279–289. IEEE Computer Society Press, Los Alamitos, CA.

Bookstein, F. L. (1997) Shape and the information in medical images: a decade of the morphometric synthesis. *Comp. Vision Image Understanding,* to appear.

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

Coggins, J. M. and Huang, C. (1993) Defining optimal feature sets for segmentation by statistical pattern recognition. In Wilson, D. C. and Wilson, J. N. (eds), *Mathematical Methods in Medical Imaging II*, *SPIE Proc*., Vol. 2035. pp. 80–88.

Cootes, T. F. and Taylor, C. J. (1996) Active shape models: a review of recent work. In Mardia, K. V. and Gill, C. A. (eds), *Proceedings in Current Issues in Statistical Shape Analysis*, pp. 108–114. Leeds University Press, Leeds.

Cutting, C., Dean, D., Bookstein, F., Haddad, B., Khorramabadi, D., Zonnefeld, F. and McCarthy, J. (1995) A three-dimensional smooth surface analysis of untreated Crouzon's disease in the adult. *J. Craniofacial Surgery*, 6, 444–453.

Davatzikos, C., Vaillant, M., Resnick, S. M., Prince, J. L., Letovsky, S. and Bryan, R. N. (1996) A computerized approach for morphological analysis of the corpus callosum. *J. Comp. Assis. Tomogr.*, 20, 88–97.

Dean, D., Marcus, L. and Bookstein, F. (1996a) Chi-square test of biological space curve affinities. In Marcus, L. F., Corti, M., Loy, A., Naylor, G. J. P. and Slice, D. E. (eds), *Advances in Morphometrics*, *NATO ASI Series A: Life Sciences*, Vol. 284, pp. 211–233. Plenum, New York.

Dean, D., Buckley, P., Bookstein, F., Kamath, J., Kwon, D., Friedman, L. and Lys, C. (1996b) Three-dimensional MR-based morphometric comparison of schizophrenic and normal cerebral ventricles. In Hòhne, H. and Kikinis R. (eds), *Visualization in Biomedical Computing*, *Lecture Notes in Computer Science,* Vol. 1131, pp. 363–372. Springer-Verlag, New York.

Evans, A. L., Collins, D. L. and Holmes, C. J. (1996) Computational approaches to quantifying human neuroanatomical variability. In Toga, A. W. and Mazziotta, J. C. (eds), *Brain Mapping: the Methods*, pp. 343–361. Academic Press, San Diego, CA.

Good, P. (1994) *Permutation Tests*. Springer-Verlag, New York.

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

Green, W. D. K. (1995) Spline-based deformable models. In Melter, R., Wu, A., Bookstein, F. and Green, W. D. K. (eds), *Vision Geometry IV*, *SPIE Proc*., Vol. 2573, pp. 290–301.

Green, W. D. K. (1996) The thin-plate spline and images with curving features. In Mardia, K. V., Gill, C. A. and Dryden, I. L. (eds), *Proceedings in Image Fusion and Shape Variability Techniques*, pp. 79–87. Leeds University Press, Leeds.

Kendall, D. G. (1984) Shape-manifolds, Procrustean metrics, and complex projective spaces. *Bull. London Math. Soc*., 16, 81–121.

Kent, J. T. (1995) Current issues for statistical inference in shape analysis. In Mardia, K. V. and Gill, C. A. (eds), *Proceedings in Current Issues in Statistical Shape Analysis*, pp. 167–175. Leeds University Press, Leeds.

Little, J. A. and Mardia, K. (1996) Edgels and tangent planes in image warping. In Marcus, L. F., Corti, M., Loy, A., Naylor, G. J. P. and Slice, D. E. (eds), *Advances in Morphometrics*, *NATO ASI Series A: Life Sciences*, Vol. 284, pp. 263–270. Plenum, New York.

Marcus, L. F., Corti, M., Loy, A., Naylor, G. J. P. and Slice, D. E. (eds) (1996) *Advances in Morphometrics*, *NATO ASI Series A: Life Sciences*, Vol. 284. Plenum, New York.

Mardia, K. V. (1995) Shape advances and future perspectives. In Mardia, K. V. and Gill, C. A. (eds), *Proceedings in Current Issues in Statistical Shape Analysis*, pp. 57–75. Leeds University Press, Leeds.

Mardia, K. V., Kent J., Goodall, C. and Little, J. (1996) Kriging and splines with derivative information. *Biometrika*, 81, 207–221.

McEachen, J. C., II, Nehorai, A. and Duncan, J. (1994) A recursive filter for temporal analysis of cardiac motion. In *Proc. IEEE Workshop on Biomedical Image Analysis*, pp. 124–133. IEEE Computer Society Press, Los Alamitos, CA.

Nastar, C. and Ayache, N. (1994) Classification of nonrigid motion in 3D images using physics-based vibration analysis. In *Proc. IEEE Workshop on Biomedical Image Analysis*, pp. 61–69. IEEE Computer Society Press, Los Alamitos, CA.

Perona, P., Shiota, T. and Malik, J. (1994) Anisotropic diffusion. In ter Haar Romeny, B. M. (ed.), *Geometry-Driven Diffusion in Computer Vision*, pp. 73–92. Kluwer, Dordrecht.

Rohlf, F. J. (1990) Fitting curves to outlines. In Rohlf, F. J. and Bookstein, F. (eds), *Proc. Michigan Morphometrics Workshop*, pp. 167–177. University of Michigan Museums, Michigan, MI.

Rohlf, F. J. and Slice, D. (1990) Extensions of the Procrustes method for the optimal superposition of landmarks. *Systematic Zoology*, 39, 40–59.

Sampson, P. D., Bookstein, F., Sheehan, F. and Bolson, E. (1996) Eigenshape analysis of left ventricular outlines from contrast ventriculograms. In Marcus, L. F., Corti, M., Loy, A., Naylor, G. J. P. and Slice, D. E. (eds), *Advances in Morphometrics*, *NATO ASI Series A: Life Sciences*, Vol. 284, pp. 211–233. Plenum, New York.

Small, C. G. (1996) *The Statistical Theory of Shape*. Springer-Verlag, New York.

Tagare, H. D., O'Shea, D. and Rangarajan, A. (1995) A geometric criterion for shape-based non-rigid correspondence. *Int. Conf. on Computer Vision*, Boston, MA.

Thirion, J.-P. (1994) The extremal mesh and the understanding of 3D surfaces. In *Proc. IEEE Workshop on Biomedical Image Analysis*., pp. 3–12. IEEE Computer Society Press, Los Alamitos, CA.

Whitaker, R. and Gerig, G. (1994) Vector-valued diffusion. In ter Haar Romeny, B. M. (ed.), *Geometry-Driven Diffusion in Computer Vision*, pp. 93–134. Kluwer, Dordrecht.