

A Minimum Description Length Approach to Statistical Shape Modeling

Rhodri H. Davies*, Carole J. Twining, Tim F. Cootes, John C. Waterton, and Chris. J. Taylor.

Abstract—We describe a method for automatically building statistical shape models from a training set of example boundaries/surfaces. These models show considerable promise as a basis for segmenting and interpreting images. One of the drawbacks of the approach is, however, the need to establish a set of dense correspondences between all members of a set of training shapes. Often this is achieved by locating a set of “landmarks” manually on each training image, which is time consuming and subjective in two dimensions and almost impossible in three dimensions. We describe how shape models can be built automatically by posing the correspondence problem as one of finding the parameterization for each shape in the training set. We select the set of parameterizations that build the “best” model. We define “best” as that which minimizes the description length of the training set, arguing that this leads to models with good compactness, specificity and generalization ability. We show how a set of shape parameterizations can be represented and manipulated in order to build a minimum description length model. Results are given for several different training sets of two-dimensional boundaries, showing that the proposed method constructs better models than other approaches including manual landmarking—the current gold standard. We also show that the method can be extended straightforwardly to three dimensions.

Index Terms—Active shape models, automatic landmarking, correspondence problem, minimum description length (MDL), point distribution models, statistical shape modeling.

I. INTRODUCTION

STATISTICAL models of shape show considerable promise as a basis for segmenting and interpreting images [1]. The basic idea is to establish, from a training set, the pattern of “legal” variation in the shapes and spatial relationships of structures for a given class of images. Statistical analysis is used to give an efficient parameterization of this variability, providing a compact representation of shape and allowing shape constraints to be applied effectively during image interpretation [2]. One of the main drawbacks of the approach is, however, the need to establish dense correspondence between shape boundaries over a reasonably large set of training images. It is important to establish the “correct” correspondences: if the points that are cor-

responded are not anatomically equivalent, the apparent shape variability can be exaggerated and the application of shape constraints during interpretation becomes less effective (see Fig. 1). In practice, correspondence has often been established using manually defined “landmarks” this is both time consuming and subjective. The problems are exacerbated when the approach is applied to three-dimensional images.

Several previous attempts have been made to automate model building [3]–[9]. The problem of establishing dense correspondence over a set of training boundaries can be posed as that of defining a parameterization for each of the training set, leading to a dense correspondence between equivalently parameterized boundary points. Arbitrary parameterizations of the training boundaries have been proposed [3], [6], but these fail to address the issue of optimality. Shape “features” (e.g., regions of high curvature) have been used to establish point correspondences, with boundary length interpolation between these points [9]–[12]. Although this approach corresponds with human intuition, it is still not clear that it is in any sense optimal. A third approach and that followed in this paper, is to treat finding the correct parameterization of the training shape boundaries as an explicit optimization problem.

The optimization approach has been described by several authors [4], [7], [13] and is discussed in more detail in Section III. The basic idea is to find the parameterizations of the training shapes that yield, in some sense, the “best” model. Kotcheff and Taylor [7] describe an approach in which the best model is defined in terms of “compactness,” as measured by the determinant of its covariance matrix. The parameterization of each of a set of training shapes was explicitly represented and a genetic algorithm search was used to optimize the model with respect to the parameterizations. Although this work showed promise and laid out much important theoretical groundwork, there were several problems: the objective function, although reasonably intuitive, could not be rigorously justified, the method was described for two-dimensional (2-D) shapes and could not easily be extended to three dimensions and the optimization often failed to converge to an acceptable solution.

In this paper, we define a new objective function with a rigorous theoretical basis and describe a new representation of correspondence/parameterization that extends to three dimensions and also results in improved convergence. Our objective function is defined in an information theoretic framework. The key insight is that the “best” model is that which describes the *entire training set* as efficiently as possible, thus, we adopt a minimum description length (MDL) criterion.

In Section II, we describe statistical shape models and outline the model-building problem. Section III reviews previous

Manuscript received October 31, 2001; revised February 25, 2002. The work of Rh. H. Davies was supported by the BBSRC and AstraZeneca, Alderley Park, Cheshire, U.K. The work of C. J. Twining was supported by the EPSRC and MRC. Asterisk indicates corresponding author.

*Rh. H. Davies is with the Division of Imaging Science and Biomedical Engineering, The Stopford Building, Oxford Road, University of Manchester, M13 9PT Manchester, U.K. (e-mail: rhodri.h.davies@stud.man.ac.uk).

C. J. Twining, T. F. Cootes, and C. J. Taylor are with the Division of Imaging Science and Biomedical Engineering, University of Manchester, M13 9PT Manchester, U.K.

J. C. Waterton is with AstraZeneca, SK10 4TG Cheshire, U.K.

Publisher Item Identifier S 0278-0062(02)05537-4.

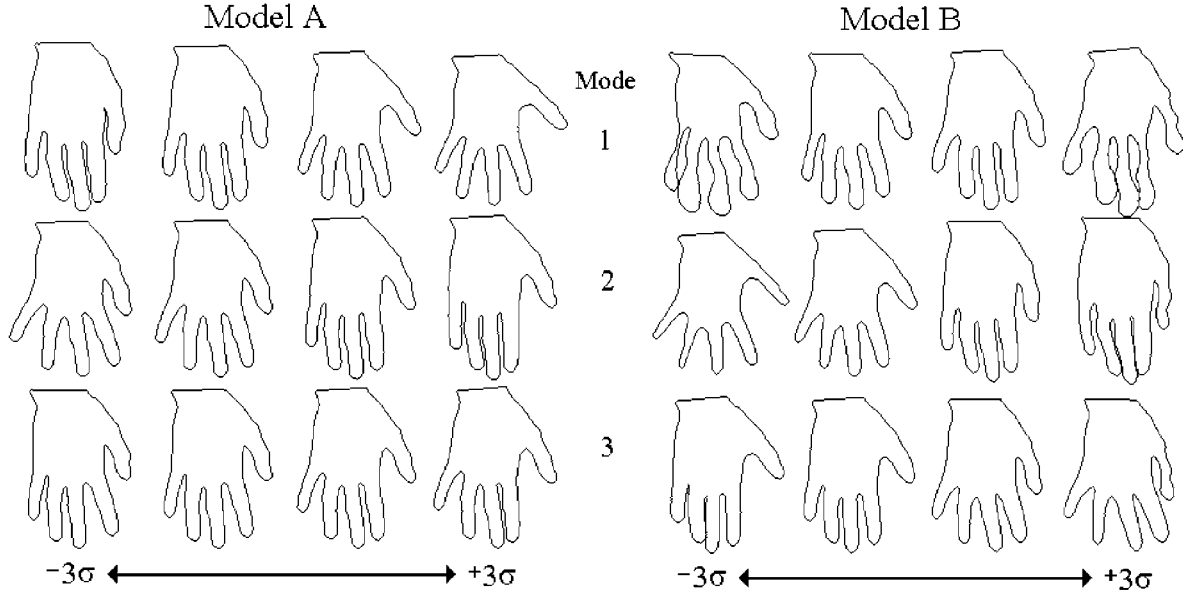


Fig. 1. The first three modes of variation ($\pm 3\sigma$) of two shape models built from the training set of hand outlines but parameterized differently. Model A was parameterized using manual “landmarks” and model B was parameterized using arc-length parameterization. The figure demonstrates that model B can represent invalid shape instances.

attempts to automate the model-building process. Section IV provides a detailed derivation of the MDL objective function. In Section V, we show how shape parameterizations can be represented explicitly and manipulated to build the best model. Section VI presents experimental results of applying the method to several training sets of object outlines.

II. STATISTICAL SHAPE MODELS

A 2-D statistical shape model is built from a training set of example outlines, aligned to a common coordinate frame. Each shape, S_i , ($i = 1 \dots n_s$), can (without loss of generality) be represented by a set of n points sampled along the boundary at equal intervals, as defined by some parameterization ϕ_i of the boundary path. This allows each shape S_i to be represented by an n_p -dimensional (n_p -D) shape vector \mathbf{x}_i , formed by concatenating the coordinates of its sample points. Using principal component analysis, each shape vector can be expressed using a linear model of the form

$$\mathbf{x}_i = \bar{\mathbf{x}} + \mathbf{P}\mathbf{b}_i = \bar{\mathbf{x}} + \sum_m \mathbf{p}^m b_i^m \quad (1)$$

where $\bar{\mathbf{x}}$ is the mean shape vector, $\mathbf{P} = \{\mathbf{p}^m\}$ are the eigenvectors of the covariance matrix (with corresponding eigenvalues $\{\lambda^m\}$) that describe a set of orthogonal modes of shape variation, and $\mathbf{b} = \{b^m\}$ are shape parameters that control the modes of variation.

Since our training shapes are continuous, we are interested in the limit $n_p \rightarrow \infty$. This leads to an infinitely large covariance matrix, but we note that there can only be, at most, $n_s - 1$ eigenvalues that are not identically zero (although they may be computationally zero). This means that in the summation above, the index m only takes values in the range one to $n_s - 1$.

To calculate the nonzero eigenvalues, we consider the $n_p \times n_s$ data matrix \mathbf{W} constructed from the set of vectors $\{(\mathbf{x}_i - \bar{\mathbf{x}}) :$

$i = 1, \dots, n_s\}$. The $n_p \times n_p$ covariance matrix is given by $\mathbf{D} = 1/n_p n_s \mathbf{W}\mathbf{W}^T$ with eigenvectors and eigenvalues $\{\mathbf{p}^m, \lambda^m\}$ thus

$$\mathbf{D}\mathbf{p}^m = \lambda^m \mathbf{p}^m. \quad (2)$$

If we define $\{\mathbf{p}'^m, \lambda'^m\}$ to be the eigenvectors and eigenvalues of the $n_s \times n_s$ matrix, $\mathbf{D}' = 1/n_p n_s \mathbf{W}^T \mathbf{W}$ then

$$\begin{aligned} \mathbf{D}'\mathbf{p}'^m &= \lambda'^m \mathbf{p}'^m \\ \text{From (2) : } \mathbf{D}\mathbf{p}^m &= \lambda^m \mathbf{p}^m \\ \Rightarrow \frac{1}{n_p n_s} \mathbf{W}\mathbf{W}^T \mathbf{p}^m &= \lambda^m \mathbf{p}^m \end{aligned} \quad (3)$$

premultiplying by \mathbf{W}^T

$$\begin{aligned} \Rightarrow \mathbf{D}'(\mathbf{W}^T \mathbf{p}^m) &= \lambda^m (\mathbf{W}^T \mathbf{p}^m) \\ \text{Similarly: } \mathbf{D}(\mathbf{W}\mathbf{p}'^m) &= \lambda'^m (\mathbf{W}\mathbf{p}'^m). \end{aligned} \quad (4)$$

Therefore, for all $\lambda^m \neq 0$, we can assign indices such that

$$\lambda^m = \lambda'^m \text{ and } \mathbf{p}^m = \mathbf{W}\mathbf{p}'^m. \quad (5)$$

Thus, the $n_s - 1$ eigenvalues of \mathbf{D} , which are not identically zero, can be obtained directly from \mathbf{D}' and the eigenvectors are a weighted sum of the training shapes. As shown in [7], in the limit $n_p \rightarrow \infty$ the ij th element of \mathbf{D}' is given by the inner product of shapes i and j

$$\mathbf{D}'_{ij} = \int dt (S_i(\phi_i(t)) - \bar{S}(t)) \cdot (S_j(\phi_j(t)) - \bar{S}(t)) \quad (6)$$

where $\bar{S} = 1/n_s \sum_{i=1}^{n_s} S_i$ is the mean shape and $S_i(\phi_i)$ is a continuous representation of S_i parameterized by ϕ_i . The integral can be evaluated numerically.

New examples of the class of shapes can be generated by choosing values of $\{b^m\}$ within the range found in the training

set. The utility of the linear model of shape shown in (1) depends on the appropriateness of the set of boundary parameterizations $\{\phi_i\}$ that are chosen. An inappropriate choice can result in the need for a large set of modes (and corresponding shape parameters) to approximate the training shapes to a given accuracy and may lead to “legal” values of $\{b^m\}$ generating “illegal” shape instances. For example, consider two models generated from a set of 17 hand outlines. Model A uses a set of parameterizations of the outlines that cause “natural” landmarks such as the tips of the fingers to correspond. Model B uses one such correspondence but then uses a simple arc-length parameterization to position the other sample points. Some “corresponding” points from each parameterization are shown in Fig. 2. The variance of the three most significant modes of models A and B are (1.06, 0.58, 0.30) and (2.19, 0.78, 0.54), respectively. This suggests that model A is more compact than model B. All the example shapes generated by model A using values of $\{b^m\}$ within the range found in the training set are “legal” examples of hands, whilst model B generates implausible examples, as can be seen in Fig. 1.

The set of parameterizations used for model A were obtained by marking the “natural” landmarks manually on each training example, then using simple arc-length parameterization to sample a fixed number of equally spaced points between them. This manual mark-up is a time-consuming and subjective process. In principle, the modeling approach extends to three dimensions, but in practice, manual landmarking becomes impractical.

III. AUTOMATIC MODEL-BUILDING

Various attempts have been made to automate the construction of statistical shape models from sets of training outlines. The simplest approach is to select a starting point and equally space landmarks along the boundary of each shape, but as we have shown in the previous section, this can result in poor models. A similar scheme is advocated by Baumberg and Hogg [3] who equally space spline control points around shape contours. Kelemen *et al.* [6] use spherical harmonic descriptors to parameterize their training shapes but resulting models are not in any obvious sense optimal.

Rueckert *et al.* [8] use a method of nonrigid registration to maximize the normalized mutual information of a set of biomedical images. The nonrigid registration is performed by manipulating a grid of B-spline control points. Principal component analysis is then performed on the resulting deformation field to build a statistical shape model. This tends to minimize the variance of the model but the correspondences are essentially arbitrary.

Tagare [12], Benayoun *et al.* [10], Kambhamettu and Goldgof [11] and Wang *et al.* [9] use shape features to select landmark points. It is not, however, clear that corresponding points will always lie on regions that have similar curvature. Also, since these methods consider only *pairwise* correspondences, they may not find the best global solution.

A more robust approach to automatic model building is to treat the task as an optimization problem. Hill and Taylor [4] attempt this by minimizing the total variance of a shape model,

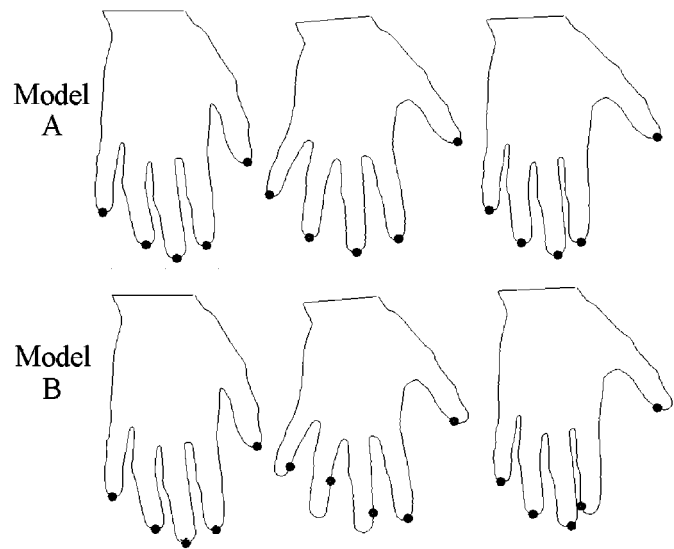


Fig. 2. Some examples of the training set and some “corresponding” points used to construct models A and B. The points “correspond” according to their parameterization.

as measured by the sum of the eigenvalues of the covariance matrix. They choose an iterative local optimization scheme, re-building the model at each stage. This makes the approach prone to becoming trapped in local minima and consequently depends on a good initial estimate of the correct landmark positions. Rangarajan *et al.* [14] describe a method of shape correspondence that also minimizes the total model variance by simultaneously determining a set of correspondences and the similarity transformation required to register pairs of contours.

Bookstein [13] describes an algorithm for landmarking sets of continuous contours represented as polygons. Points are allowed to move along the contours to minimize a bending energy term. Again, it is not obvious that optimizing an energy term will lead to good statistical shape models.

Kotcheff and Taylor [7] describe an objective function based on the determinant of the model covariance. This favors compact models with a small number of significant modes of variation, though no rigorous theoretical justification for this formulation is offered. They use an explicit representation of the set of shape parameterizations $\{\phi_i\}$ and optimize the model directly with respect to $\{\phi_i\}$ using genetic algorithm search. Their representation of $\{\phi_i\}$ is, however, problematic and does not guarantee a diffeomorphic mapping. They correct the problem when it arises by reordering correspondences, which is workable for 2-D shapes but does not extend obviously to three dimensions. Although some of the results produced by their method are better than hand-generated models, the algorithm did not always converge.

IV. AN INFORMATION THEORETIC OBJECTIVE FUNCTION

We wish to define a criterion for selecting the set of parameterizations $\{\phi_i\}$ that are used to construct a statistical shape model from a set of training boundaries $\{S_i\}$. Our aim is to choose $\{\phi_i\}$ so as to obtain the “best possible” model. In Section III, we reviewed the following possible objective functions but none of these guarantee a shape model with ideal properties.

Generalization Ability: The model can describe any instance of the object—not just those seen in the training set.

Specificity: The model can only represent valid instances of the object.

Compactness: The variation is explained with few parameters.

To achieve this, we follow the principle of Occams razor which can be paraphrased: “simple descriptions generalize best.” As a quantitative measure of “simplicity,” we choose to apply the The MDL principle [15], [16]. This is based on the idea of transmitting data as a coded message, where the coding is based on some prearranged set of parametric statistical models. The full transmission then has to include not only the encoded data values, but also the coded model parameter values. Thus, MDL balances the model complexity, expressed in terms of the cost of sending the model parameters, against the quality of fit between the model and the data, expressed in terms of the coding length of the data. Comparison of description lengths calculated using models from different classes can be used as a way of solving the model selection problem [17]. However, our emphasis here is not on selecting the class of model, but on using the description length for a single class of model as an objective function for optimization of correspondence between the shapes.

We will use the simple two-part coding formulation of MDL. Although this does not give us a coding which is of the absolute minimum length [18], it does however give us a functional form which is computationally simple to evaluate, hence, suitable to be used as an objective function for numerical optimization.

A. The Model

Our training set of n_s shapes is sampled according to the parameterizations $\{\phi_i\}$ to give a set of n_p -D shape vectors $\{\mathbf{x}_i\}$. We choose to model this set of shape vectors using a multivariate Gaussian model. The initial step in constructing such a model is to change to a coordinate system whose axes are aligned with the principal axes of the data set. This corresponds to the orientation of the linear model defined earlier (1)

$$\mathbf{x}_i = \bar{\mathbf{x}} + \sum_{m=1}^{n_s-1} \mathbf{p}^m b_i^m. \quad (7)$$

The $n_s - 1$ eigenvectors \mathbf{p}^m lie in the space of shapes, they span the subspace which contains the training set and they are aligned with the principal axes. We now order these vectors in terms of nondecreasing eigenvalue and construct the orthonormal set

$$\tilde{\mathbf{p}}^m = \frac{\mathbf{p}^m}{\|\mathbf{p}^m\|}, m = 1 \text{ to } n_s - 1. \quad (8)$$

Our new coordinates with respect to this set of axes are defined, thus

$$y^m \equiv \mathbf{x} \cdot \tilde{\mathbf{p}}^m. \quad (9)$$

This corresponds to projecting \mathbf{x} into the subspace and then rotating the axes about the original origin. To describe this transformation, we have to transmit the set of $n_s - 1$, n_p -D vectors

$\{\tilde{\mathbf{p}}^m\}$. The code length for this transmission is a function of n_p and n_s only, hence, is constant for a given training set and number of sample points and will not be considered further.

B. The Description Length

For each direction $\tilde{\mathbf{p}}^m$, we now have to transmit the set of values $Y^m \equiv \{y_i^m : i = 1 \text{ to } n_s\}$. Since we have aligned our coordinate axes with the principal axes of the data, each direction is now modeled using a one-dimensional (1-D) Gaussian. In Appendix I, we derive an expression for the description length of one-dimensional, bounded, and quantized data, coded using a Gaussian model. To utilize this result, we first have to calculate a strict upper-bound R on the range of our data and also estimate a suitable value for a quantization parameter Δ .

Suppose that, for our original shape data, we know that the coordinates of our sample points are strictly bounded, thus

$$-\frac{r}{2} \leq x_{i\alpha} \leq \frac{r}{2} \text{ for all } \alpha = 1 \text{ to } n_p, i = 1 \text{ to } n_s. \quad (10)$$

Then, the upper-bound for the coordinates $\{y_i^m\}$ is given by

$$R = r\sqrt{n_p}, \text{ so that } -\frac{R}{2} \leq y_i^m \leq \frac{R}{2} \text{ for all } i, m. \quad (11)$$

The data quantization parameter Δ can be determined by quantizing the coordinates of our original sample points. Comparison of the original shape and the quantized shape then allows a maximum permissible value of Δ to be determined. For example, for boundaries obtained from pixellated images, Δ will typically be of the order of the pixel size. This also determines our lower bound on the modeled variance $\sigma_{\min} \equiv 2\Delta$. The parameters R and Δ are constant for a given training set, hence, we need not consider the description length for the transmission of these values.

Our original data values Y^m are now replaced by their quantized values \hat{Y}^m .¹ The quantized mean $\hat{\mu}^m$ of the quantized data in each direction is determined and the variance $(\sigma^m)^2$ of the data about this mean calculated, thus

$$(\sigma^m)^2 = \frac{1}{n_s} \sum_{i=1}^{n_s} (\hat{y}_i^m - \hat{\mu}^m)^2. \quad (12)$$

The description length D_m for each direction is then given by (see Appendix I)

- If $\sigma^m \geq \sigma_{\min}$

$$D_m = \ln\left(\frac{R}{\Delta}\right) + D^{(1)}(\hat{Y}^m, R, \Delta)$$

- If $\sigma^m < \sigma_{\min}$ but the range of $\hat{Y}^m > \Delta$

$$D_m = \ln\left(\frac{R}{\Delta}\right) + D^{(2)}(\hat{Y}^m, R, \Delta)$$

- Else

$$D_m = \ln\left(\frac{R}{\Delta}\right).$$

¹We will use \hat{a} to denote the quantized value of the corresponding continuum value a .

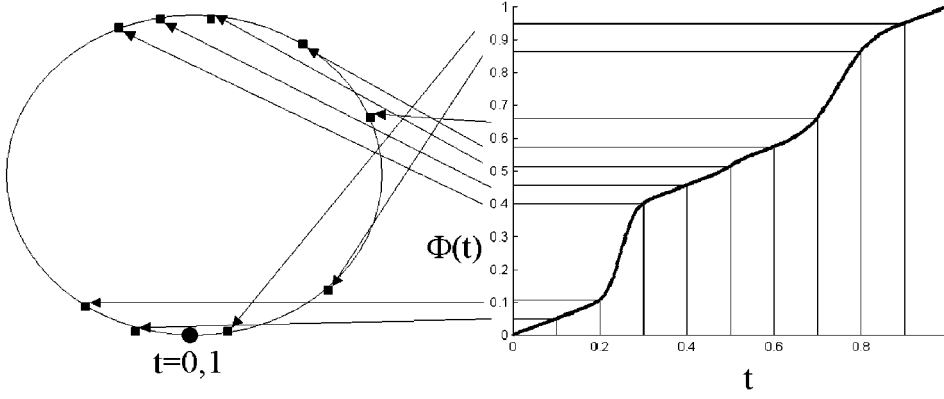


Fig. 3. A demonstration of how a circle is sampled according to its parameterization. To sample points on the shape, we uniformly sample the bottom axis of the parameterization function (t) and read the values off the vertical axis ($\phi(t)$). As there is a one-to-one correspondence between the vertical axis ($\phi(t)$) and the shape boundary; points are sampled according to the shape of the parameterization function.

C. The Objective Function

Let us define n_g to be the number of directions for which the first of the above criteria holds and n_{\min} the number which satisfy the second. Then, since the directions are ordered in terms of nonincreasing eigenvalue/variance, the total description length for our training set can be written, thus

$$D = (n_s - 1) \ln \left(\frac{R}{\Delta} \right) + \sum_{p=1}^{n_g} D^{(1)}(\hat{Y}^p, R, \Delta) + \sum_{q=n_g+1}^{n_g+n_{\min}} D^{(2)}(\hat{Y}^q, R, \Delta). \quad (13)$$

The leading term is the cost of transmitting the mean shape. For a given training set, the quantities R and Δ will be held constant, so that we can drop the leading term, to give the objective function

$$F = \sum_{p=1}^{n_g} D^{(1)}(\hat{Y}^p, R, \Delta) + \sum_{q=n_g+1}^{n_g+n_{\min}} D^{(2)}(\hat{Y}^q, R, \Delta). \quad (14)$$

We now consider the form of this objective function. For the linear model defined earlier (1)

$$n_p \lambda^m = \frac{1}{n_s} \sum_{i=1}^{n_s} (y_i^m - \bar{y}^m)^2 \text{ where } \bar{y}^m \equiv \frac{1}{n_s} \sum_{i=1}^{n_s} y_i^m. \quad (15)$$

In the limit $\Delta \rightarrow 0$, the quantized values of \hat{Y} and $\hat{\mu}$ approach their continuum values, so that

$$\hat{\mu}^m \rightarrow \bar{y}^m \text{ and } \sigma^m \rightarrow \sqrt{n_p \lambda^m}. \quad (16)$$

If we also consider the limit where n_s is sufficiently large, it can be seen that the functions $D^{(1)}$ and $D^{(2)}$ can be written in the form

$$D^{(1)}(Y^m, R, \Delta) \approx f(R, \Delta, n_s) + (n_s - 2) \ln \sigma^m \\ D^{(2)}(Y^m, R, \Delta) \approx f(R, \Delta, n_s) + (n_s - 2) \ln \sigma_{\min} \\ + \frac{(n_s + 3)}{2} \left[\left(\frac{\sigma^m}{\sigma_{\min}} \right)^2 - 1 \right] \quad (17)$$

where f is some function which depends only on R , Δ and n_s . So, in this dual limit, the part of the objective function which depends on the $\{\sigma^m\}$ contains terms similar to the determinant of the covariance matrix (that is $\sum \ln \lambda^m$) used by Kotcheff and Taylor [7]. However, our objective function is well defined, even in the limit $\lambda^m \rightarrow 0$, where in fact such a direction makes **no** contribution to the objective function. Whereas in the form used previously, without the addition of artificial correction terms, it would have an infinitely large contribution.

V. SELECTING CORRESPONDENCES

In order to build a statistical shape model, we need to sample a number of *corresponding* points on each shape. As demonstrated in Section II, the choice of correspondences determines the quality of the model.

We choose to cast this correspondence problem as one of defining the parameterization function ϕ_i , of each shape. These parameterization functions explicitly define how points are sampled on each shape (see Fig. 3). We can consequently minimize the value of F in (14) by manipulating the set of parameterization functions $\{\phi_i\}$.

Our training shapes are represented parametrically as curves in two dimensions

$$S_i(t) = (S_{xi}(t), S_{yi}(t)) \in \mathbb{R}^2 \text{ where } 0 \leq t \leq 1 \\ \text{and } i = 1 \text{ to } n_s. \quad (18)$$

In order to constrain the point ordering—and, hence, the correspondences—the parameterization $\phi_i(t)$ of each shape i must be a monotonically increasing function of t . That is

$$S_i \rightarrow S'_i \text{ so that} \\ S'_i(t) = S'_i(\phi_i(t)), \text{ where } \phi_i : [0, 1] \rightarrow [0, 1]. \quad (19)$$

Each $\phi_i(t)$ must be one-to-one, onto and invertible, i.e., a diffeomorphism of a circle (for closed curves) or a line (for open curves).

We describe below a novel, piecewise-linear representation of parameterization and describe how stochastic optimization can be used to find the set of parameterizations that minimize F .

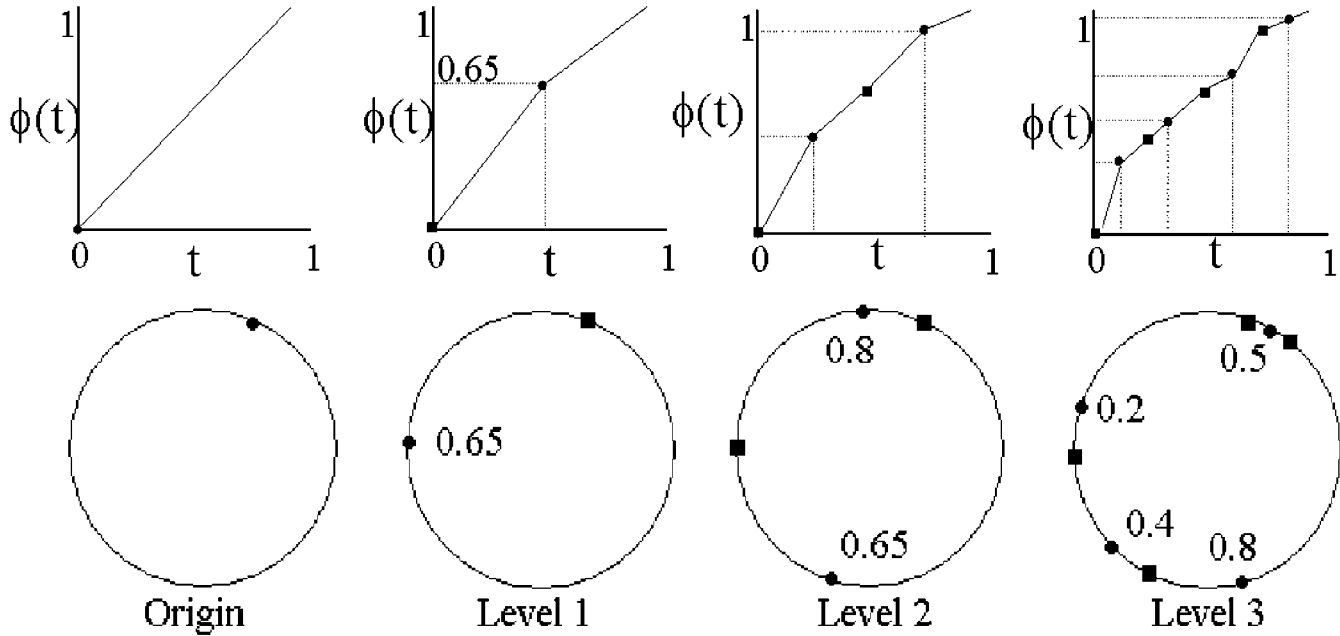


Fig. 4. This figure demonstrates the proposed method for representing the parameterization. The top row shows the parameterization and the bottom row shows how points would be sampled on a circle. The circles represent daughter nodes and squares parent nodes. The brackets on the parameterization show the range that each node is allowed to move. The node has a value of zero at the bottom of the bracket, one at the top, and 0.5 in the middle. The parameter values for this example are: [Origin, 0.65 (0.65 (0.4, 0.8), 0.8 (0.5, 0.2))].

A. Representing the Parameterizations

We choose to use a piecewise-linear representation of the parameterization. In two dimensions, we could ensure that $\phi_i(t)$ is monotonic by enforcing the ordering of points on the boundaries according to arc-length. On surfaces, however, no such ordering exists. To overcome this, we have developed a novel method of representation that guarantees a diffeomorphic mapping without using arc-length ordering—allowing straightforward extension to surfaces in three dimensions.

We define the piecewise linear parameterization for each training shape by recursively defining the parameterization, inserting nodes between those already present. The position of each daughter node is coded in terms of its fractional distance between its two parent nodes. Thus, by constraining the positions of daughter nodes to lie in the range $[0, 1]$ (where the node has a value of zero if it is positioned on its left neighbor, one on its right neighbor, and 0.5 in the center) we can enforce an *implicit ordering*. This is illustrated by the example in Fig. 4, which demonstrates the parameterization of a circle.

Additional nodes are added until the parameterization is suitably defined, where in general the degree of refinement required depends on the complexity of the training shapes.

The parameterization is fully described by the set of fractional distances defining the positions of the daughter nodes relative to their parents. Any such set of positive fractions describes a valid parameterization.

Once a parameterization $\phi_i(t)$ has been constructed in this way for each shape in the training set, an arbitrary number of corresponding points can be obtained by sampling the training shapes at equally spaced intervals of $\phi_i(t)$. These sampled shapes (and, hence, the generated correspondences) can then be evaluated using the objective function F in (14).

In summary, the algorithm for closed curves can be described as follows.

Recursive Parameterization of a Single Shape:

- An initial node defines an origin and endpoint.
- Given a set of nodes, a daughter node is created between each adjacent pair of parent nodes. The parameter describing each daughter node is then its fractional distance along the curve between the parent nodes.
- The set of parent and daughter nodes together form the initial set of nodes for the next level of recursion.

Optimization:

- Generate a parameterization for each shape recursively, to the same level.
- Sample the shapes according to the correspondence defined by the parameterization.
- Build a model from the sampled shapes.
- Calculate the objective function.
- Vary the parameterization of each shape until the optimum value of the objective function is found.

The representation of parameterization is similar to that used by Tagare [12] and Kotcheff and Taylor [7] but as it only uses an implicit ordering, it can be extended to build statistical shape models of surfaces in three dimensions (see Appendix II for details).

B. Optimizing the Parameterizations

We wish to manipulate the set of parameterizations $\{\phi_i\}$ in order to minimize our objective function F . One way that F can be minimized is for all the points on all shapes to collapse to a single part of the boundary. To avoid this, we must select a “reference shape” whose parameterization is fixed.

In practice, 10–30 nodes are required to sufficiently define the parameterization for each shape—creating a high-dimensional configuration space. The behavior of F over this space is highly nonlinear and contains many local minima leading us to prefer a stochastic optimization method such as simulated annealing [19] or genetic algorithm search [20]. We chose to use a genetic algorithm to perform the experiments reported in Section VI.

VI. RESULTS

We present qualitative and quantitative results of applying our method to several sets of outlines of 2-D biomedical objects. The parameters used for the genetic algorithm were (crossover rate: 100%; mutation rate: 0.05%; population size: 4000; nonelitist, sigma-scaled rank selection [21]). A single corresponding point was chosen on each shape and used as the origin.² Procrustes alignment [22] was performed on the shapes (parameterized by arc-length) to transfer them into a common coordinate frame. Four levels of refinement were used to give $1+2+4+8 = 15$ nodes.

We also investigated how our objective function behaves around the minimum and how it selects the correct dimensionality of the final model.

A. Shape Data

The method was evaluated using several training sets of biomedical objects:

Infarcts: Permanent focal cerebral ischaemia was induced in rats and multislice T2-weighted MRI was performed *in vivo*, as described previously [23]. For this study, only data from saline-treated animals were used, giving a total of 23 outlines. The shape of the infarct in this case represents the territory of the middle cerebral artery. Following segmentation, an atlas of anatomy [24] was used to select a single slice corresponding to an anatomic location 6.3 mm posterior to the bregma.

Kidneys: Wistar, Sprague-Dawley and Fisher rats were imaged using an MRI system (“Inova,” Varian, Palo Alto, CA) incorporating a 400 mm bore 4.7 T magnet (Oxford Instruments, Oxford, U.K.), a 150 mm bore 200 mT/m pulsed field gradient set (Oxford Instruments) and a 63 mm bore quadrature birdcage transceiver (Varian). Multislice T2-weighted MRI was performed in the transverse plane with repetition time (TR) 2 s; echo time (TE) 20 ms, and slice thickness 1 mm, with 41 contiguous slices. Images were acquired with a 64×64 mm field of view and a $256 \times 256 \times 41$ image matrix. Following segmentation, a single transverse slice was selected from the right kidney that included most evidence of the collecting apparatus.

Knee Cartilage: Sagittal images of the articular cartilage of the lateral femoral condyle were acquired from asymptomatic human subjects using T1-weighted MRI and segmented as described previously [25]. A single sagittal slice was chosen from the center of the lateral femoral condyles. As the width of the femur varies from subject to subject, comparable slices were identified by selecting the slices halfway between 1, the first evidence of the lateral aspect of the meniscal horn and 2, the full extent of the posterior cruciate ligament.

²The position of the origin can be optimized but the algorithm takes longer to converge.

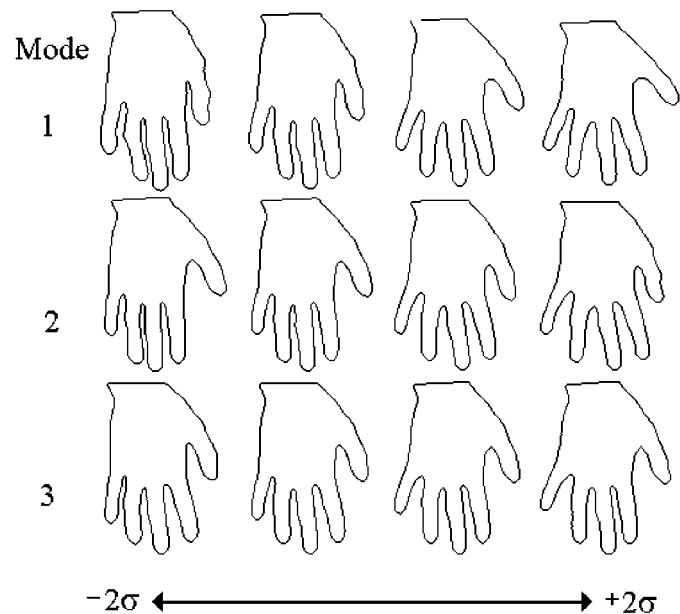


Fig. 5. The first three modes ($m = 1$, $m = 2$, and $m = 3$) of variation of the automatically generated model of the hand outlines. The shape instances were obtained by varying the values of $\{b^m\}$ by $(\pm 2\sigma)$.

Hand Outlines: 17 hand outlines were segmented from images of different poses from the same subject.

Hip Prostheses: 24 outlines of hip prostheses were segmented from clinical radiographs of subjects who had undergone total joint replacement [26]. Plane-film X-ray was used to acquire the image where the beam was centered on the symphysis pubis so that the radiograph captured the full pelvis and contralateral hip.

Left Ventricle: 38 outlines of the left ventricle of the heart were segmented from transcostal, long-axis echocardiograms [27]. The outlines were randomly selected from 33 image pairs of different subjects.

B. Results on 2-D Outlines

We tested our method on the training sets described in Section IV-A. In Figs. 5–10, we show qualitative results by displaying the variation captured by the first three modes of each model (b^m ($m = 1 \dots 3$)) varied by $(\pm 2[\text{standard deviations over training set}])$. We also give quantitative results in Tables I–VI, tabulating the value of F , the total variance and variance explained by each mode for each of the models, comparing the automatic result with those for models built using manual landmarking and arc-length parameterized (equally spaced) points.

The quantitative results show that the automatically generated models are significantly more compact (have less variance per mode) than either the models built by manual landmarking or by arc-length parameterization. It is interesting to note that the models produced by arc-length parameterization of the hip prostheses and heart ventricles have a lower value of the objective function than the manual model. This is because there are few salient anatomical landmarks, thus, errors in the manual annotation adds extra noise that is captured as statistical variation. It is also interesting to note that, although the variance of the

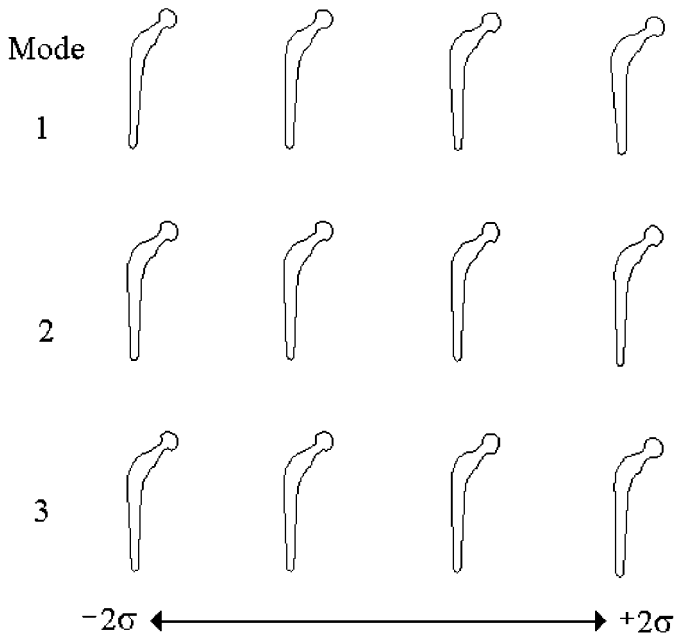


Fig. 6. The first three modes ($m = 1$, $m = 2$, and $m = 3$) of variation of the automatically generated model of the hip prostheses.

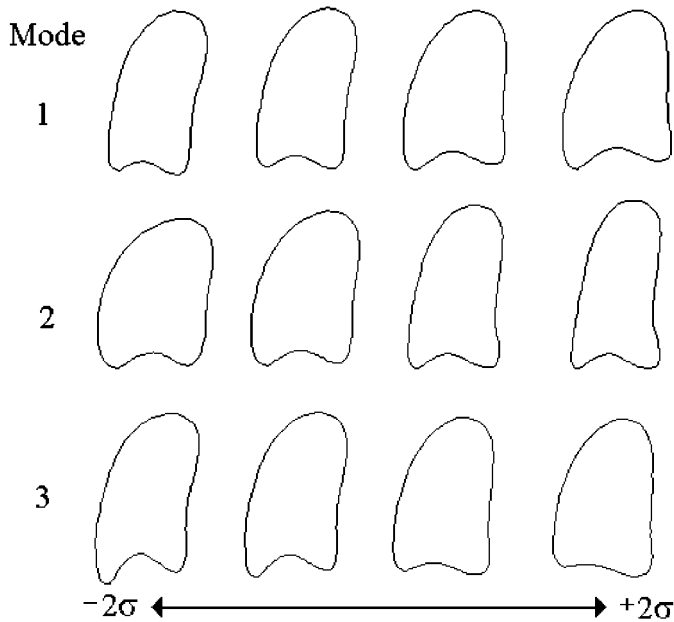


Fig. 7. The first three modes ($m = 1$, $m = 2$, and $m = 3$) of variation of the automatically generated model of the heart ventricles.

manual heart ventricle model is less than the arc-length model, the value of the objective function is higher.

To test the generalization ability of the models, we performed leave-one-out tests on each to determine the accuracy of with which the model was able to approximate unseen examples of the same class. In Fig. 11, we report the results for the hand outlines, although the same trends appear in all datasets. As can be seen from the figure, the optimized model performs significantly better than both the manual and arc-length parameterized models for any number of retained modes, indicating better generalization ability.

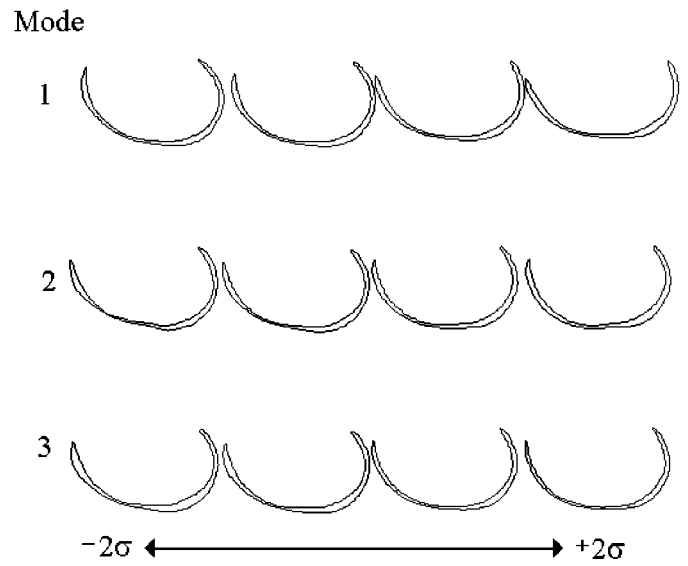


Fig. 8. The first three modes ($m = 1$, $m = 2$, and $m = 3$) of variation of the automatically generated model of the knee cartilage.

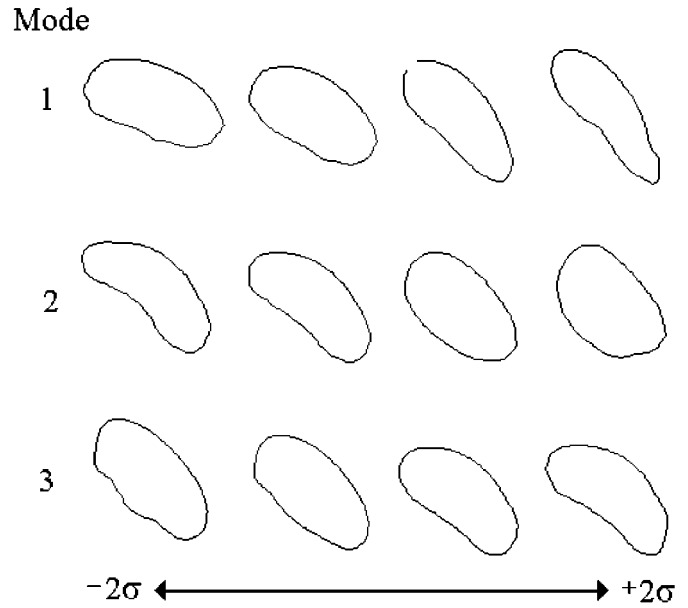


Fig. 9. The first three modes ($m = 1$, $m = 2$, and $m = 3$) of variation of the automatically generated model of the infarcts.

C. The Behavior of F

In Fig. 12, we show the effect on F of adding Gaussian random noise to the position of each point on each shape for the automatically generated hand model. Points are constrained to move along the curve, so that the object outline is effectively unchanged, but the parameterization is perturbed. As can be seen from the figure, the objective function gives a clear minimum for zero noise, corresponding to our original optimal solution.

D. The Effective Dimensionality of the Model

We generated a set of 49 artificial shapes, generated to have exactly one mode of variation (varying the position of a semi-circular bump along an edge of a rectangle). Taking the orig-

Mode

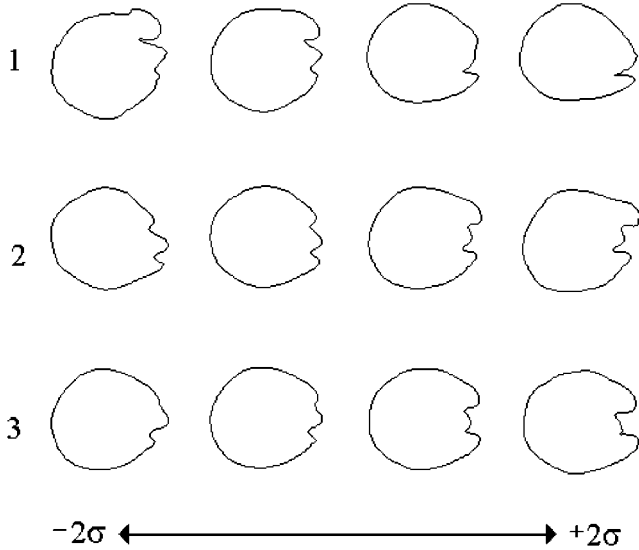
Fig. 10. The first three modes ($m = 1$, $m = 2$, and $m = 3$) of variation of the automatically generated model of the kidneys.

TABLE I

A QUANTITATIVE COMPARISON OF THE HAND OUTLINE MODELS SHOWING THE VARIANCE EXPLAINED BY EACH MODE. V_T IS THE TOTAL VARIANCE AND F IS THE VALUE OF THE OBJECTIVE FUNCTION

Mode	Manual	Arc-length	Automatic
1	3.35	6.91	3.33
2	1.84	2.47	1.57
3	0.96	1.71	0.89
4	0.80	1.05	0.48
5	0.24	0.57	0.16
6	0.23	0.41	0.09
V_T	7.94	13.7	6.69
F	3 081	3 114	2 926

TABLE II

A QUANTITATIVE COMPARISON OF THE HIP PROSTHESES MODELS SHOWING THE VARIANCE EXPLAINED BY EACH MODE. V_T IS THE TOTAL VARIANCE AND F IS THE VALUE OF THE OBJECTIVE FUNCTION

Mode	Manual	Arc-length	Automatic
1	1.25	1.39	1.38
2	0.62	0.27	0.22
3	0.21	0.12	0.11
4	0.18	0.01	0.02
5	0.10	0.01	0.01
6	0.05	0.01	0.01
V_T	2.49	1.84	1.76
F	5 655	4 837	4 815

inal arc-length correspondence and the final optimized correspondence, we then calculated the contribution to the objective function from each dimension of the model (see Fig. 13). As can be seen from the figure, the initial parameterization gives a nonzero variance, hence, nonzero F for all $n_s - 1$ dimensions of

TABLE III

A QUANTITATIVE COMPARISON OF THE HEART VENTRICLE MODELS SHOWING THE VARIANCE EXPLAINED BY EACH MODE. V_T IS THE TOTAL VARIANCE AND F IS THE VALUE OF THE OBJECTIVE FUNCTION

Mode	Manual	Arc-length	Automatic
1	2.45	2.48	1.98
2	1.21	1.98	1.56
3	0.81	0.87	0.69
4	0.35	0.31	0.25
5	0.26	0.21	0.17
6	0.17	0.14	0.11
V_T	5.65	6.35	5.07
F	12 396	12 391	12 064

TABLE IV

A QUANTITATIVE COMPARISON OF THE KNEE CARTILAGE MODELS SHOWING THE VARIANCE EXPLAINED BY EACH MODE. V_T IS THE TOTAL VARIANCE AND F IS THE VALUE OF THE OBJECTIVE FUNCTION

Mode	Manual	Arc-length	Automatic
1	171.5	171.6	146.5
2	27.7	27.8	12.7
3	14.3	14.2	6.8
4	4.8	4.9	3.7
5	3.8	3.8	2.4
6	2.3	2.3	1.8
V_T	229.1	229.4	177.1
F	1 443	1 443	1 403

TABLE V

A QUANTITATIVE COMPARISON OF THE INFRACT MODELS SHOWING THE VARIANCE EXPLAINED BY EACH MODE. V_T IS THE TOTAL VARIANCE AND F IS THE VALUE OF THE OBJECTIVE FUNCTION

Mode	Arc-length	Automatic
1	931.0	306.6
2	283.0	59.5
3	251.9	54.7
4	57.4	36.6
5	40.7	27.1
6	35.1	22.3
V_T	1 710	1 239
F	3 767	3 604

TABLE VI

A QUANTITATIVE COMPARISON OF THE KIDNEY MODELS SHOWING THE VARIANCE EXPLAINED BY EACH MODE. V_T IS THE TOTAL VARIANCE AND F IS THE VALUE OF THE OBJECTIVE FUNCTION

Mode	Arc-length	Automatic
1	499.3	128.2
2	164.3	54.0
3	74.9	33.7
4	42.2	28.6
5	25.8	14.5
6	17.8	8.6
V_T	862.3	284.5
F	2 046	1 912

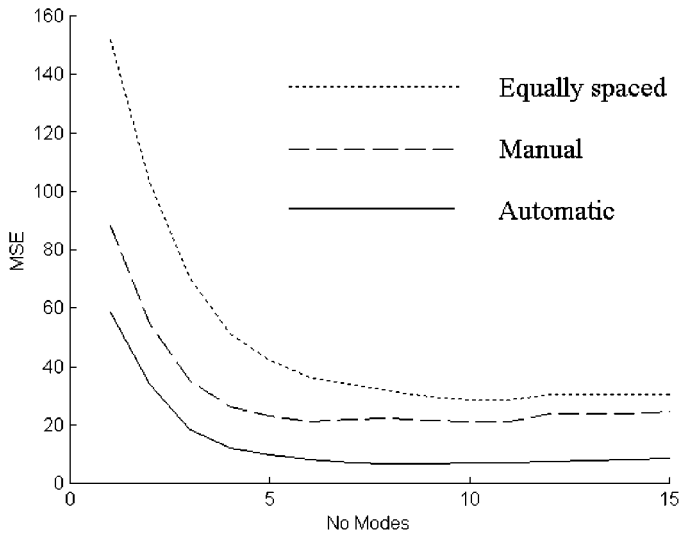


Fig. 11. The result of running leave-one-out reconstructions on each model of the hand outlines. The model is built with all but one example and then fitted to the unseen example. The plot shows the mean squared approximation error against the number of modes used. This measures the ability of the model to represent unseen shape instances of the object.

the model. For the final optimized parameterization, the objective function for all modes other than the first is identically zero, showing that we have correctly matched the dimensionality of the modeled variation to the actual known dimensionality of the shape variation.

VII. DISCUSSION AND CONCLUSION

We have derived an objective function that provides a principled measure of the quality of a statistical shape model. The expression we use is grounded in information theory and considers both model complexity and quality of model fit to data in a unified way. The objective function bears some similarities to that used by Kotcheff and Taylor [7], but does not suffer from the problem of requiring ad hoc correction terms.

We have also described a new method of representing parameterization of curves in two dimensions, that has a natural extension to the parameterization of surfaces in three dimensions.

When coupled with stochastic optimization, the objective function allows us to automatically build models that are substantially better than those built using manual landmarking. We have shown that these automatic models, as well as being quantitatively more compact than manually landmarked models (as measured in terms of the total variance), also have improved generalization ability.

As with any stochastic optimization technique, our search requires a large number of function evaluations. Each of the results in this paper typically took several hours to produce. Although this is a one-off, off-line process, it is likely to become impractical when a larger set of training shapes is used because its complexity is at least $O(n_s^2)$. We are currently working on finding faster ways of locating the optimum.

We note that the MDL based objective function, being a general measure of model quality, can also be used to determine the values of other quantities than the parameterization. We intend

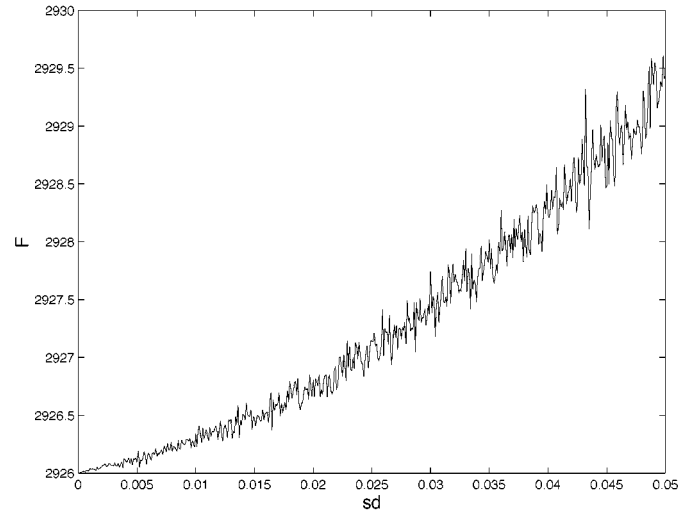


Fig. 12. A plot of F against the standard deviation of the random Gaussian noise on each control point to show how noise on the control points affects the value of the objective function.

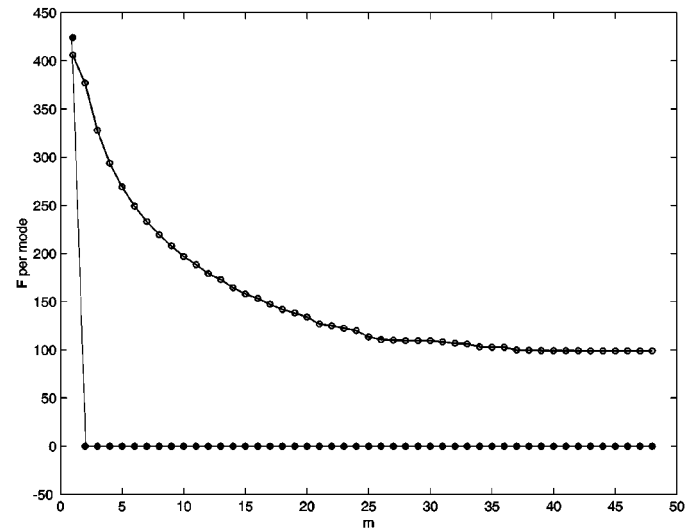


Fig. 13. A plot of the value of F for each mode for the initial parameterization (open circles) and the final optimized parameterization (closed circles) for the artificial data example.

to investigate this and use the objective function to improve on such things as the Procrustes alignment. Hence, different aspects of model building will then be combined into a single, unified optimization framework.

APPENDIX I DESCRIPTION LENGTH FOR 1-D GAUSSIAN MODELS

In this appendix, we show how to construct an expression for the description length required to send a 1-D data set using a Gaussian model. The total description length is computed as the sum of the description length for coding the parameters of the Gaussian model and the description length for coding the data using the model.

We take our data set $Y = \{y_i : i = 1 \text{ to } n_s\}$ to lie within a strictly bounded region. We then quantize the data values using

a parameter Δ , so that $Y \rightarrow \hat{Y} = \{\hat{y}_i : i = 1 \text{ to } n_s\}$,³ where for any quantized value \hat{y} from any possible data set

$$-\frac{R}{2} \leq \hat{y} \leq \frac{R}{2} \text{ and } \hat{y} = m\Delta, m \in \mathbb{Z}. \quad (20)$$

Our data model is the family of Gaussian distributions, defined by parameters σ and μ

$$\rho(y; \sigma, \mu) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(y - \mu)^2\right). \quad (21)$$

We will use the fundamental result that the ideal-coding codeword length for a value \hat{y} , encoded using a statistical model $P(\hat{y})$ is given by the Shannon Coding codeword length [28]⁴

$$\mathcal{L}(\hat{y}; P) = -\log_2 P(\hat{y}) \text{ bits, or } -\ln P(\hat{y}) \text{ nats.} \quad (22)$$

A. Coding the Parameters

Consider first the parameter μ . This should obviously be related to the mean of the actual data. Given that the data is bounded, we can see that the allowed values of μ should be similarly bounded. We here make the simplest modeling choice, which is to quantize the allowed values of μ to the same accuracy as our data. So, we define the allowed values of the quantized parameter $\hat{\mu}$ to be

$$\hat{\mu} = \left(m \pm \frac{1}{2}\right) \Delta \text{ where } m \in \mathbb{Z} \text{ and } -\frac{R}{2} \leq \hat{\mu} \leq \frac{R}{2} \quad (23)$$

where we choose the value $\hat{\mu}$ which is closest to the actual mean of the quantized data. Given the absence of any prior knowledge, we will assume a flat distribution for $\hat{\mu}$ over this range. This then gives us the ideal codeword length for $\hat{\mu}$ of

$$\mathcal{L}_{\hat{\mu}} = \ln\left(\frac{R}{\Delta}\right). \quad (24)$$

For the parameter $\hat{\sigma}$, we will assume that its allowed values are bounded and quantized, thus

$$\hat{\sigma} = n\delta, n \in \mathbb{N} \text{ and } \sigma_{\min} \leq \hat{\sigma} \leq \sigma_{\max} \quad (25)$$

which then gives us a codeword length

$$\mathcal{L}_{\hat{\sigma}} = \ln\left(\frac{\sigma_{\max} - \sigma_{\min}}{\delta}\right). \quad (26)$$

Note that our receiver cannot decrypt the value of $\hat{\sigma}$ without knowing the value of δ . So we now have to consider the codeword length for transmitting δ . Assuming the quantization parameter δ is of the form

$$\delta = 2^{\pm k}, k \in \mathbb{N} \quad (27)$$

then it can easily be seen that it can be coded directly with a codeword length

$$\mathcal{L}_{\delta} = 1 + |\log_2 \delta| \text{ bits} \approx 1 + |\ln \delta| \text{ nats} \quad (28)$$

where the additional bit/nat codes for the sign in the exponent of δ .

³We will use α to denote continuum values and $\hat{\alpha}$ to denote the corresponding quantized value.

⁴In what follows, we will restrict ourselves to natural logarithms and work in units of nats. However, expressions can easily be converted back into bit-lengths by noting that 1 bit $\equiv \ln 2$ nats.

So, our total code length for transmitting the parameters is given by

$$\mathcal{L}_{\text{parameters}} = \mathcal{L}_{\hat{\mu}} + \mathcal{L}_{\hat{\sigma}} + \mathcal{L}_{\delta}. \quad (29)$$

B. Coding the Data

For our Gaussian data model, the probability $P(\hat{y})$ associated with a bin centered at \hat{y} is given by

$$\begin{aligned} P(\hat{y}) &= \int_{\hat{y}-\Delta/2}^{\hat{y}+\Delta/2} dk \rho(k; \sigma, \mu) \\ &\approx \frac{\Delta}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(\hat{y} - \mu)^2\right). \end{aligned} \quad (30)$$

It can be shown numerically that this is a very good approximation (to better than 99% of the exact value) for all values $\sigma \geq 2\Delta$, hence, we will take

$$\sigma_{\min} = 2\Delta. \quad (31)$$

The code length for the data set is then

$$\mathcal{L}_{\text{data}}(\hat{\sigma}) = -n_s \ln \Delta + \frac{n_s}{2} \ln(2\pi\hat{\sigma}^2) + \frac{1}{2\hat{\sigma}^2} \sum_{i=1}^{n_s} (\hat{y}_i - \hat{\mu})^2. \quad (32)$$

The variance of the quantized data about the quantized mean is

$$\sigma^2 = \frac{1}{n_s} \sum_{i=1}^{n_s} (\hat{y}_i - \hat{\mu})^2 \text{ and } \sigma_{\max} = \frac{R}{2}. \quad (33)$$

In general, σ will differ from its quantized value $\hat{\sigma}$, thus

$$\hat{\sigma} = \sigma + d_{\sigma}, |d_{\sigma}| \leq \frac{\delta}{2}. \quad (34)$$

So, averaging over an ensemble of data sets and assuming a flat distribution for d_{σ} over this range, we find

$$\begin{aligned} \langle d_{\sigma}^2 \rangle &= \frac{\delta^2}{12}, \quad \frac{1}{\hat{\sigma}^2} = \frac{1}{\sigma^2} \left(1 + \frac{\delta^2}{4\sigma^2}\right) + O\left(\frac{\delta^4}{\sigma^4}\right) \\ \langle \ln \hat{\sigma}^2 \rangle &= \ln \sigma^2 - \frac{\delta^2}{12\sigma^2} + O\left(\frac{\delta^4}{\sigma^4}\right). \end{aligned} \quad (35)$$

Substituting these expressions into (32) then gives us the following expression for the description length of the data:

$$\begin{aligned} \mathcal{L}_{\text{data}} &= -n_s \ln \Delta + \frac{n_s}{2} \ln(2\pi\sigma^2) - \frac{n_s\delta^2}{24\sigma^2} \\ &\quad + \frac{1}{2\sigma^2} \left(1 + \frac{\delta^2}{4\sigma^2}\right) \sum_{i=1}^{n_s} (\hat{y}_i - \hat{\mu})^2 + O\left(\frac{\delta^4}{\sigma^4}\right). \end{aligned} \quad (36)$$

Substituting from (33) and keeping only terms up to $O(\delta^2/\sigma^2)$

$$\mathcal{L}_{\text{data}} = -n_s \ln \Delta + \frac{n_s}{2} \ln(2\pi\sigma^2) + \frac{n_s}{2} + \frac{n_s\delta^2}{12\sigma^2}. \quad (37)$$

The total description length is then

$$\begin{aligned} \mathcal{L}^{(1)} &= \mathcal{L}_{\text{parameters}} + \mathcal{L}_{\text{data}} \\ &= 1 + \ln\left(\frac{R}{\Delta}\right) + \ln\left(\frac{\sigma_{\max} - \sigma_{\min}}{\delta}\right) + |\ln \delta| \\ &\quad - n_s \ln \Delta + \frac{n_s}{2} \ln(2\pi\sigma^2) + \frac{n_s}{2} + \frac{n_s\delta^2}{12\sigma^2}. \end{aligned} \quad (38)$$

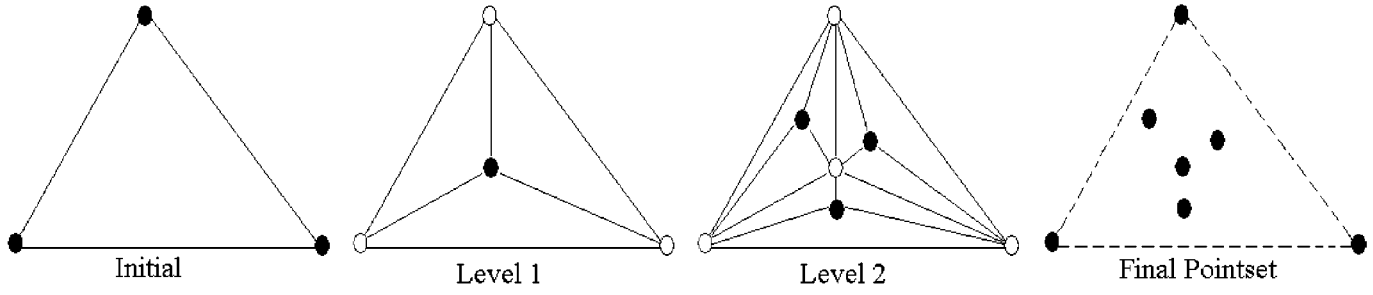


Fig. 14. A diagram to demonstrate the representation of parameterization on a surface. The solid circles represent new daughter nodes and hollow points represent nodes that are already in place. Each daughter node is allowed to move anywhere inside the spherical triangle formed by its three parent nodes.

By differentiating with respect to δ and setting the derivative to zero, we find that the optimum value of δ is

$$\delta^*(\sigma, n_s) = \min \left(1, \sigma \sqrt{\frac{12}{n_s}} \right) \quad (39)$$

which then allows us to write the above expression as

$$\mathcal{L}^{(1)} = \ln \left(\frac{R}{\Delta} \right) + D^{(1)}(\hat{Y}, R, \Delta). \quad (40)$$

In the case where $\sigma < \sigma_{\min}$, but the data occupies more than one bin, we will model the data using a Gaussian of width σ_{\min} and a quantization parameter $\delta = \delta^*(\sigma_{\min}, n_s)$. An analogous derivation to that given above then gives us the description length

$$\begin{aligned} \mathcal{L}^{(2)} = & 1 + \ln \left(\frac{R}{\Delta} \right) + \ln \left(\frac{\sigma_{\max} - \sigma_{\min}}{\delta} \right) + |\ln \delta| \\ & - n_s \ln \Delta + \frac{n_s}{2} \ln (2\pi\sigma_{\min}^2) - \frac{n_s \delta^2}{24\sigma_{\min}^2} \\ & + \frac{1}{2\sigma_{\min}^2} \left(1 + \frac{\delta^2}{4\sigma_{\min}^2} \right) \sum_{i=1}^{n_s} (\hat{y}_i - \hat{\mu})^2 \\ \equiv & \ln \left(\frac{R}{\Delta} \right) + D^{(2)}(\hat{Y}, R, \Delta). \end{aligned} \quad (41)$$

Note that this contains terms of a similar form to those in the expression for $\mathcal{L}_{\text{data}}$ in (36), but with σ_{\min} rather than σ . The explicit data term is, hence, left in and allows us to calculate the increase in description length caused by the nonoptimal choice of $\sigma = \sigma_{\min}$. The remaining case is where all the data lies in one bin; we then only have to transmit the position of the mean, with a description length

$$\mathcal{L}^{(3)} = \ln \left(\frac{R}{\Delta} \right). \quad (42)$$

APPENDIX II

EXTENSION OF SHAPE PARAMETERIZATION TO THREE DIMENSIONS

In this appendix, we describe how the representation of the parameterization can be extended to build statistical shape models from surfaces in three dimensions.

A similar construction can be performed for the case of open surfaces.

Each surface in our training set is represented as a triangulated mesh. For surfaces which are topologically equivalent to

spheres,⁵ we obtain an initial parameterization by mapping each mesh to a unit sphere, where the mapping must be such that there is no folding or tearing. Each mapped mesh can then be represented, thus

$$S_i = S_i(\theta, \psi), \quad S_i \in \mathbb{R}^3 \quad (43)$$

where S_i is the set of original positions of the mesh vertices for the i th surface in Euclidean space and (θ, ψ) are the spherical polar coordinates of each mapped vertex. Various approaches have been described to achieve such mappings [29]–[32]. Since the final parameterization is obtained by optimizing the objective function, the final result will, in general, not depend on the particular initial mapping chosen.

Changes in parameterization of a given surface, hence, correspond to altering the positions of the mapped vertices on the sphere. That is

$$\begin{aligned} S_i &\rightarrow S'_i, \quad \theta \rightarrow \theta', \quad \psi \rightarrow \psi' \\ \text{where } S_i(\theta, \psi) &= S'_i(\theta', \psi') \\ \text{and } \theta' &= \phi_i^\theta(\theta, \psi), \quad \psi' = \phi_i^\psi(\theta, \psi). \end{aligned} \quad (44)$$

Note that we have separate parameterization functions $\phi_i = (\phi_i^\theta, \phi_i^\psi)$ for each surface. Valid parameterization functions ϕ_i correspond to exact homeomorphic mappings of the sphere. That is, mappings that are continuous, one-to-one and onto. The piecewise-linear representation of the parameterization proposed in Section V-A must now be extended. As in two dimensions, we construct an explicit representation of the parameterization by the use of a recursive process.

The construction is initialized by first selecting four nodes on the sphere, which form the initial mesh of four spherical triangles. Given a spherical triangulated mesh of nodes on the sphere, the next recursive level is defined by adding new nodes. Each spherical triangle in the mesh of nodes is subdivided into three smaller triangles by adding a new daughter node, which is constrained to lie inside the parent spherical triangle. If \mathbf{q}_1 , \mathbf{q}_2 and \mathbf{q}_3 are the position vectors (with respect to the center of the sphere) of the parent nodes at the vertices of the spherical triangle, then the new daughter node can be represented by

$$\mathbf{q}_0 = \alpha \mathbf{q}_1 + \beta \mathbf{q}_2 + \gamma \mathbf{q}_3. \quad (45)$$

The constraint that \mathbf{q}_0 lies within the *planar* triangle formed by \mathbf{q}_1 , \mathbf{q}_2 and \mathbf{q}_3 can be satisfied, thus

$$\alpha = 1 - \beta - \gamma, \quad \text{and } \alpha \geq 0, \beta \geq 0, \gamma \geq 0 \quad (46)$$

⁵The same method can be used on open surfaces but for clarity, we limit our discussion to the closed case.

which then gives two degrees of freedom to optimize for each new node. The position of the daughter node on the sphere is then given by projecting, thus

$$\mathbf{q}_0 \rightarrow \frac{\mathbf{q}_0}{\|\mathbf{q}_0\|}. \quad (47)$$

The new mesh of nodes is then given by the set of parent nodes and their daughters and the process is repeated. The construction is illustrated in Fig. 14.

ACKNOWLEDGMENT

The authors would like to thank A. Brett for his contribution to this work.

REFERENCES

- [1] T. Cootes, A. Hill, C. Taylor, and J. Haslam, "The use of active shape models for locating structures in medical images," *Image Vis. Comput.*, vol. 12, pp. 355–366, 1994.
- [2] T. Cootes, C. Taylor, D. Cooper, and J. Graham, "Active shape models—Their training and application," *Comput. Vis. Image Understanding*, vol. 61, pp. 38–59, 1995.
- [3] A. Baumberg and D. Hogg, "Learning flexible models from image sequences," in *Proc. Eur. Conf. Computer Vision*, Stockholm, Sweden, 1994, pp. 299–308.
- [4] A. Hill and C. Taylor, "Automatic landmark generation for point distribution models," in *Proc. Br. Machine Vis. Conf.*, 1994, pp. 429–438.
- [5] A. Hill and C. J. Taylor, "A framework for automatic landmark identification using a new method of nonrigid correspondence," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, pp. 241–251, Mar. 2000.
- [6] A. Kelemen, G. Szekely, and G. Gerig, "Elastic model-based segmentation of 3-D neuroradiological data sets," *IEEE Trans. Med. Imag.*, vol. 18, pp. 828–839, Oct. 1999.
- [7] A. C. W. Kotcheff and C. J. Taylor, "Automatic construction of eigen-shape models by direct optimization," *Med. Image Anal.*, vol. 2, pp. 303–314, 1998.
- [8] D. Rueckert, F. Frangi, and J. A. Schnabel, "Automatic construction of 3D statistical deformation models using nonrigid registration," in *Proc. MICCAI'01*, 2001, pp. 77–84.
- [9] Y. Wang, B. S. Peterson, and L. H. Staib, "Shape-based 3D surface correspondence using geodesics and local geometry," in *Proc. CVPR 2000*, vol. 2, 2000, pp. 644–651.
- [10] A. Benayoun, N. Ayache, and I. Cohen, "Adaptive meshes and non-rigid motion computation," in *Proc. Int. Conf. Pattern Recognition*, Jerusalem, Israel, 1994, pp. 730–732.
- [11] C. Kambhamettu and D. B. Goldgof, "Point correspondence recovery in nonrigid motion," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 1992, pp. 222–227.
- [12] H. D. Tagare, "Shape-based nonrigid correspondence with application to heart motion analysis," *IEEE Trans. Med. Imag.*, vol. 18, pp. 434–449, July 1999.
- [13] F. L. Bookstein, "Landmark methods for forms without landmarks: Morphometrics of group differences in outline shape," *Med. Image Anal.*, vol. 1, no. 3, pp. 225–243, 1997.
- [14] A. Rangarajan, H. Chui, and F. L. Bookstein, "The softassign procrustes matching algorithm," in *Proc. IPMI'97*, 1997, pp. 29–42.
- [15] J. R. Rissanen, "A universal prior for integers and estimation by minimum description length," *Ann. Statist.*, vol. 11, no. 2, pp. 416–431, 1983.
- [16] —, *Stochastic Complexity in Statistical Inquiry*, Singapore: World Scientific, 1989, vol. 15, Computer Science.
- [17] M. H. Hansen and B. Yu, "Model Selection and the Principle of Minimum Description Length," Bell Labs, Murray Hill, N.J, Tech. Memo., 1998.
- [18] J. R. Rissanen, "Fisher information and stochastic complexity," *IEEE Trans. Inform. Theory*, vol. 42, pp. 40–47, Jan. 1996.
- [19] S. Kirkpatrick, C. Gelatt, and M. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, pp. 671–680, 1983.
- [20] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*. Reading, MA: Addison Wesley, 1989.
- [21] M. Mitchell, *An Introduction to Genetic Algorithms*. Cambridge, MA: MIT Press, 1998.
- [22] C. Goodall, "Procrustes methods in the statistical analysis of shape," *J. Roy. Statist. Soc. B*, vol. 53, no. 2, pp. 285–339, 1991.
- [23] J. C. Waterton, B. J. Middleton, R. Pickford, C. P. Allott, D. Checkley, and R. A. Keith, "Reduced animal use in efficacy testing in disease models with use of sequential experimental designs," *Dev. Animal Veterinary Sci.*, vol. 31, pp. 737–745, 2000.
- [24] G. Paxinos and C. Watson, *The Rat Brain in Stereotaxic Coordinates*. San Diego, CA: Academic, 1986.
- [25] S. Solloway, C. E. Hutchinson, J. C. Waterton, and C. J. Taylor, "The use of active shape models for making thickness measurements of articular cartilage from MRI," *Magn. Reson. Med.*, vol. 47, pp. 943–952, 1997.
- [26] A. L. Redhead, A. C. W. Kotcheff, C. J. Taylor, M. L. Porter, and D. W. L. Hukins, "An automated method for assessing routine radiographs of patients with total hip replacement," in *Proc. Inst. Mech.Eng.*, vol. 211(H), 1997.
- [27] A. D. Parker, A. Hill, C. J. Taylor, T. F. Cootes, X. Y. Jin, and D. G. Gibson, "Application of point distribution models to the automated analysis of echocardiograms," in *Proc. Comput. Cardiol.*, 1994, pp. 25–28.
- [28] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379 and 623–423 and 656, 1948.
- [29] S. Angenent, S. Haker, A. Tannenbaum, and R. Kikinis, "On the Laplace-Beltrami operator and brain surface flattening," *IEEE Trans. Med. Imag.*, vol. 18, pp. 700–711, Aug. 1999.
- [30] M. Eck, T. DeRose, T. DuChamp, H. Hoppe, M. Lounsberry, and W. Stuetzle, "Multiresolution analysis of arbitrary meshes," in *Proc. SIGGRAPH '95 (Computer Graphics)*, 1995, pp. 173–182.
- [31] C. Brechbühler, G. Gerig, and O. Kubler, "Parameterization of closed surfaces for 3-D shape description," *Comput. Vis. Image Understanding*, vol. 61, no. 2, pp. 154–170, 1995.
- [32] D. Tosun and J. L. Prince, "Hemispherical map for the human brain cortex," in *Proc. SPIE Medical Imaging 2001*, 2001, pp. 290–300.