

3D Statistical Shape Models Using Direct Optimisation of Description Length

Rhodri H. Davies¹, Carole J. Twining¹, Tim F. Cootes¹, John C. Waterton²,
and Chris J. Taylor¹

¹ Division of Imaging Science, Stopford Building, Oxford Road,
University of Manchester, Manchester, M13 9PT, UK

`rhodri.h.davies@stud.man.ac.uk`

² AstraZeneca, Alderley Park, Macclesfield,
Cheshire, SK10 4TG, UK

Abstract. We describe an automatic method for building optimal 3D statistical shape models from sets of training shapes. Although shape models show considerable promise as a basis for segmenting and interpreting images, a major drawback of the approach is the need to establish a dense correspondence across a training set of example shapes. It is important to establish the correct correspondence, otherwise poor models can result. In 2D, this can be achieved using manual ‘landmarks’, but in 3D this becomes impractical. We show it is possible to establish correspondences *automatically*, by casting the correspondence problem as one of finding the ‘optimal’ parameterisation of each shape in the training set. We describe an explicit representation of surface parameterisation, that ensures the resulting correspondences are legal, and show how this representation can be manipulated to minimise the *description length* of the training set using the model. This results in compact models with good generalisation properties. Results are reported for two sets of biomedical shapes, showing significant improvement in model properties compared to those obtained using a uniform surface parameterisation.

1 Introduction

Statistical models of shape show considerable promise as a basis for segmenting and interpreting images in 2D [5]. 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 parameterisation of this variability, providing a compact representation of shape and allowing shape constraints to be applied effectively during image interpretation [6]. A key step in building a model involves establishing a dense correspondence between shape boundaries/surfaces over a reasonably large set of training images. It is important to establish the ‘correct’ correspondences, otherwise an inefficient parameterisation of shape can result, leading to difficulty in defining shape constraints effectively. In 2D, correspondence is often established using manually defined ‘landmarks’ but this is a time-consuming, error-prone

and subjective process. In principle, the method extends to 3D, but in practice, manual landmarking becomes impractical. In this paper we show how an ‘optimal’ model can be built by *automatically* defining correspondences across a training set of 3D shapes.

Several previous attempts have been made to build 3D statistical shape models [3,4,10,12,13,14,21,25]. The problem of establishing dense correspondence over a set of training shapes can be posed as that of defining a parameterisation for each of the training set, assuming correspondence between equivalently parameterised points. Kelemen *et. al* [14] and Hill *et. al* [12] use different arbitrary parameterisations of the training shapes. Christensen *et al.* [7], Szekely and Lavalley [23] and Rueckert *et al.* [21] describe methods for warping the space in which the shapes are embedded. Models can then be built from the resulting deformation field [10,13,21]. Brett and Taylor [3,4] and Wang *et. al* [25] use shape ‘features’ (e.g. regions of high curvature) to establish point correspondences.

The correspondences found using the methods described above are not, in any obvious sense, the correct ones. We show in section 2 (fig. 1) that unsatisfactory models can result if correspondences are chosen inappropriately. We start from the position that the correct correspondences are, by definition, those that build the ‘best’ model. We define the ‘best’ model as that with optimal compactness, specificity and generalisation ability. We have shown elsewhere [8] that a model with these properties can be obtained using an objective function based on the minimum description length principle [18]. We have also described a method that uses the objective function to build models in 2D that are better than the best models we could build using manual landmarks [9]. The representation of the parameterisation did not, however, extend to surfaces in 3D. In this paper we describe the derivation of the objective function. We also describe a novel method of representing and manipulating the parameterisation of a surface allowing the construction of shape models in 3D. The method ensures that only valid correspondences can be represented.

In the remainder of the paper, we establish notation, and outline the model-building problem. We then provide a summary of the derivation of the minimum description length objective function. We show how a set of surface parameterisations can be represented explicitly and manipulated to build an optimal model. Finally, we present qualitative and quantitative results of applying the method to surfaces obtained from 3D MR images of brain ventricles and rat kidneys.

2 Statistical Shape Models

A 2d (3d) statistical shape model is built from a training set of example outlines (surfaces), 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 regularly sampled on the shape, as defined by some parameterisation ϕ_i . This allows each shape S_i to be represented by an n_p -dimensional 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 1 to $n_s - 1$.

To calculate the non-zero 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 $\mathbf{D} = \frac{1}{n_p n_s} \mathbf{W}\mathbf{W}^T$ 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, $\mathbf{D}' = \frac{1}{n_p n_s} \mathbf{W}^T \mathbf{W}$ then:

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

pre-multiplying 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 [15], 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 $\bar{S} = \frac{1}{n_s} \sum_{i=1}^{n_s} S_i$ is the mean shape and $S_i(\phi_i)$ is a continuous representation of S_i parameterised by ϕ_i . The integral can be evaluated numerically.

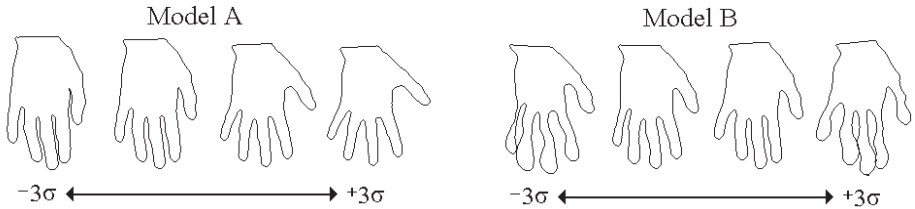


Fig. 1. The first mode of variation ($\pm 3\sigma$) of two shape models built from the training set of hand outlines but parameterised differently. Model *A* was parameterised using manual ‘landmarks’ and model *B* was parameterised using arc-length parameterisation. The figure demonstrates that model *B* can represent invalid shape instances.

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 parameterisations $\{\phi_i\}$ 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, figure 1 shows two 2D models generated from a set of 17 hand outlines. Model *A* uses a set of parameterisations 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 path length parameterisation to position the other sample points. The variances 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 and is thus of limited utility for imposing shape constraints when the model is used for image search, see fig. 1.

The set of parameterisations used for model *A* were obtained by marking ‘natural’ landmarks manually on each training example, then using simple path length parameterisation 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 modelling approach extends to 3D, but in practice, manual landmarking becomes impractical. We propose to overcome this problem by *automatically* defining correspondences between a training set of example shapes.

3 An Information Theoretic Objective Function

We wish to define a criterion for selecting the set of parameterisations $\{\phi_i\}$ 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. Ideally, we would like a model with optimal:

Generalisation 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 Occam’s razor which can be paraphrased: ‘simple descriptions generalise best’. As a quantitative measure of ‘simplicity’, we choose to apply The Minimum Description Length (MDL) Principle [18,19]. The MDL principle is based on the idea of transmitting a data set as an encoded message, where the code is based on some pre-arranged 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 [11]. 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 optimisation 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 [20], 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 optimisation.

3.1 The Model

Our training set of n_s shapes is sampled according to the parameterisations $\{\phi_i\}$ to give a set of n_p -dimensional 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 $n_s - 1$ mutually-orthogonal eigenvectors $\{\mathbf{p}^m\}$ span the subspace which contains the training set, and, by appropriate scaling, can be transformed into an orthonormal basis set for this subspace. We will also order these vectors in terms of non-decreasing eigenvalue to give us our final orthonormal basis set $\{\tilde{\mathbf{p}}^m\}$.

To transmit a shape \mathbf{x}_i using this model, we first have to transmit the mean shape $\bar{\mathbf{x}}$, then the deviations from the mean shape, which can be written thus:

$$y_i^m \equiv \tilde{\mathbf{p}}^m \cdot (\mathbf{x}_i - \bar{\mathbf{x}}). \quad (8)$$

We will assume that the code length for the transmission of the mean shape $\bar{\mathbf{x}}$ is a constant for a given training set and number of sample points. Furthermore, the code length for the transmission of the set of $n_s - 1$, n_p -dimensional basis vectors $\{\tilde{\mathbf{p}}^m\}$ is also constant for a given training set and number of sample points and need not be considered further.

3.2 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, and aligned our origin with the mean shape, each direction can now be modelled using a one-dimensional, centred Gaussian. In the Appendix, we derive an expression for the Description Length of one-dimensional, bounded and quantised data, coded using a centred Gaussian model. To utilise 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 the data quantisation parameter Δ .

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

$$-\frac{r}{2} \leq x_{i\alpha} \leq \frac{r}{2} \text{ for all } \alpha = 1 \text{ to } n_p, i = 1 \text{ to } n_s. \quad (9)$$

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

$$R = r\sqrt{n_p}, \text{ so that } |y_i^m| \leq R \text{ for all } i, m. \quad (10)$$

The data quantisation parameter Δ can be determined by quantising the coordinates of our original sample points. Comparison of the original shape and the quantised shape then allows a maximum permissible value of Δ to be determined. For example, for boundaries obtained from voxelated images, Δ will typically be of the order of the voxel size. This also determines our lower bound on the modelled variance $\sigma_{min} \equiv 2\Delta$. The parameters R and Δ are held 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 quantised¹ values \hat{Y}^m . The variance of the quantized data is then calculated thus:

$$(\sigma^m)^2 = \frac{1}{n_s} \sum_{i=1}^{n_s} (\hat{y}_i^m)^2. \quad (11)$$

The Description Length D_m for each direction is then given by (see Appendix):

- If $\sigma^m \geq \sigma_{min}$: $D_m = D^{(1)}(\sigma^m, n_s, R, \Delta)$
- If $\sigma^m < \sigma_{min}$ but the range of $\hat{Y}^m \geq \Delta$: $D_m = D^{(2)}(\sigma^m, n_s, R, \Delta)$
- Else: $D_m = 0$.

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

3.3 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 non-increasing eigenvalue/variance, the total Description Length for our training set, and our Objective function, can be written thus:

$$F = \sum_{p=1}^{n_g} D^{(1)}(\sigma^p, n_s, R, \Delta) + \sum_{q=n_g+1}^{n_g+n_{min}} D^{(2)}(\sigma^q, n_s, R, \Delta). \quad (12)$$

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

$$n_p \lambda^m = \frac{1}{n_s} \sum_{i=1}^{n_s} (y_i^m)^2 \quad (13)$$

In the limit $\Delta \rightarrow 0$, the quantised values in \hat{Y} approach their continuum values, so that:

$$\sigma^m \rightarrow \sqrt{n_p \lambda^m}. \quad (14)$$

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:

$$D^{(1)}(\sigma^m, n_s, R, \Delta) \approx f(R, \Delta, n_s) + (n_s - 2) \ln \sigma^m \quad (15)$$

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

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 [15]. 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.

4 Manipulating the Parameterisation of a Surface

In order to build a statistical shape model, we need to sample a number of *corresponding* points on each shape. As demonstrated in section 2, the choice of correspondences determines the quality of the model. We have chosen to cast this correspondence problem as that of defining the parameterisation ϕ_i , of each shape so as to minimise the value of F in (12). We propose to solve for $\{\phi_i\}$ using numerical optimisation, which requires that we formulate an explicit (and ideally, compact) representation of the ϕ_i that can be manipulated. We would also like the computational complexity to be minimal (i.e. it should only involve the evaluation of elementary functions).

Each surface in our training set is originally represented as a triangular mesh that is topologically equivalent to a sphere. We obtain an initial parameterisation by mapping each surface 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:

$$S_i = S_i(\theta, \psi), \quad S_i \in \mathbb{R}^3 \quad (16)$$

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 [1, 2, 24]. Since we intend to optimise the parameterisations, the final result should significantly not depend on this initial mapping. We have used the method described by Brechbulher [2] to produce the results reported below.

Changes in parameterisation of a given surface correspond to altering the positions of the mapped vertices on the sphere. That is:

$$S_i \rightarrow S'_i, \quad \theta \rightarrow \theta', \quad \psi \rightarrow \psi' \quad (17)$$

$$\text{where } S_i(\theta, \psi) = S'_i(\theta', \psi') \quad \text{and} \quad \theta' = \phi_i^\theta(\theta, \psi), \quad \psi' = \phi_i^\psi(\theta, \psi).$$

Note that we have separate parameterisation functions $\phi_i = (\phi_i^\theta, \phi_i^\psi)$ for each surface. Valid parameterisation functions ϕ_i correspond to exact homeomorphic mappings of the sphere. That is, mappings that are continuous, one-to-one and onto. In the following sections, we present a number of such mappings.

4.1 Symmetric Theta Transformations

Consider an arbitrary point P on the unit sphere. For simplicity, assume that the spherical polar co-ordinates (θ, ψ) on the sphere have been redefined so that P corresponds to the point $\theta = 0$. Let us first consider a rotationally symmetric mapping that reparameterises the θ coordinate:

$$\theta \rightarrow f(\theta). \quad (18)$$

For the mapping to be homeomorphic and continuous with the identity, f must be a differentiable non-decreasing monotonic function over the range $0 < \theta < \pi$, with $f(0) = 0$, $f(\pi) = \pi$. Any such monotonic function f can be rewritten in terms of the cumulative distribution function of some density function $\rho(\theta)$, defined over the range $0 \leq \theta \leq \pi$.

As our normalised density function, we take a constant term plus a wrapped Cauchy distribution. The wrapped Cauchy [16] is a normalisable, uni-modal distribution for circular data, of variable width, which has an analytic indefinite integral:

$$\rho(\theta) = \frac{1}{N} \left[1 + A \left(\frac{1 - \alpha^2}{1 + \alpha^2 - 2\alpha \cos \theta} \right) \right] \quad (19)$$

where $N = \pi [1 + A]$. Hence:

$$f(\theta) = \pi \int_0^\theta ds \rho(s) = \frac{1}{1 + A} \left[\theta + A \arccos \left(\frac{(1 + \alpha^2) \cos \theta - 2\alpha}{1 + \alpha^2 - 2\alpha \cos \theta} \right) \right] \quad (20)$$

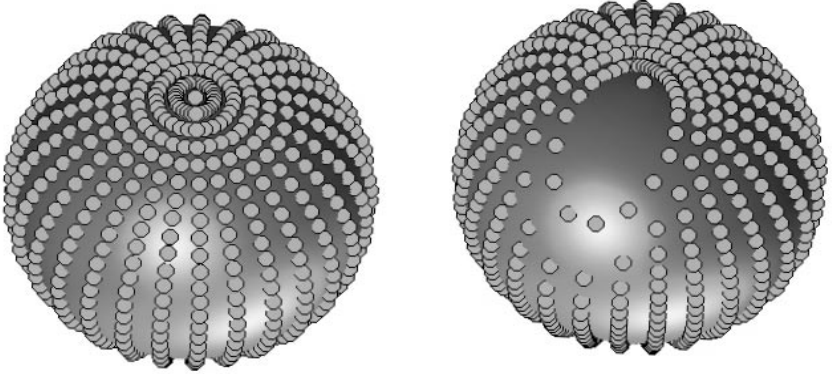


Fig. 2. *Left:* Original sphere. *Right:* Sphere after asymmetric θ transformation.

where α ($\alpha \equiv e^{-a}$, $a \in \mathbb{R}$) is the width, and A ($A \geq 0$) is the amplitude of the Cauchy.

The constant term is included so that $f(\theta) = \theta$ when $A = 0$. i.e. the parameterisation is unchanged when the Cauchy has zero magnitude.

4.2 Asymmetric Theta Transformations

We can also perform non-symmetric transformations:

$$\theta \rightarrow f(\theta, \psi). \quad (21)$$

Define ψ' to be the ψ coordinates redefined so that a point ($Q \neq P$) corresponds to $\psi' = 0$. An asymmetric transformation around the point P , towards a point Q can be achieved using (20) and making the amplitude A a smooth function of the ψ coordinate:

$$A \rightarrow A(\psi'). \quad (22)$$

One such way to do this is to use the wrapped Cauchy distribution to obtain:

$$A(\psi') = A_0 \left[\frac{1 - \beta^2}{1 + \beta^2 - 2\beta \cos \psi'} - \frac{1 - \beta^2}{(1 + \beta)^2} \right] \quad (23)$$

where β ($\beta \equiv e^{-b}$) is the width of the subsidiary Cauchy. We have chosen the formulation such that $A(\psi')$ has a minimum value of zero. An example of an asymmetric transformation is shown in figure 2.

4.3 Shear and Twist

We also consider transformations of the ψ coordinate. This is equivalent to shearing and twisting the sphere about the axis defined by the point P . So, for example, we could consider a reparameterisation of the form:

$$\psi \rightarrow \psi + g(\theta) \quad (24)$$

where

$$g(\theta) = \left(\frac{B}{2\pi} \right) \frac{1 - \gamma^2}{1 + \gamma^2 - 2\gamma \cos(\theta - \theta_0)} \quad (25)$$

where B is the amplitude, γ ($\gamma \equiv e^{-c}$, $c \in \mathbb{R}$) is the width and θ_0 is the position of the centre. This transformation is continuous with the identity at $B = 0$ (i.e. the transformation has no affect when $B = 0$). It can also be localised about $\theta = \theta_0$ in the limit of zero width. An example of such a transformation is shown in Figure 3. Figure 4 shows an example of applying a combination of all the transformations described above.

5 Optimising the Parameterisations

We now wish to manipulate $\{\phi_i\}$ so as to minimise F in (12). We have found that, for the objects modelled in this paper, the symmetric theta transformations, alone, provide a sufficient group of reparameterisations. To manipulate the parameterisations, we fix the position P and width a of each of the Cauchy kernels and vary its magnitude A .

To select the positions P of the kernels, we (approximately) uniformly sample the sphere and centre the kernels at these sample points. It is not strictly possible to position an arbitrary number of equidistant points on the sphere, but a good

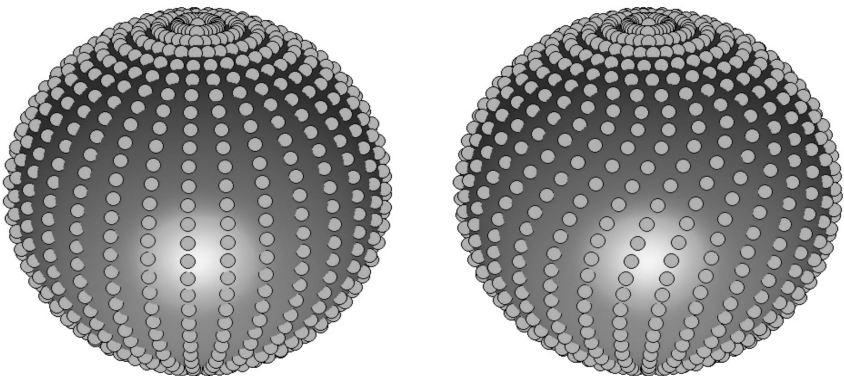


Fig. 3. *Left:* Original sphere. *Right:* Sphere after shear transformation.

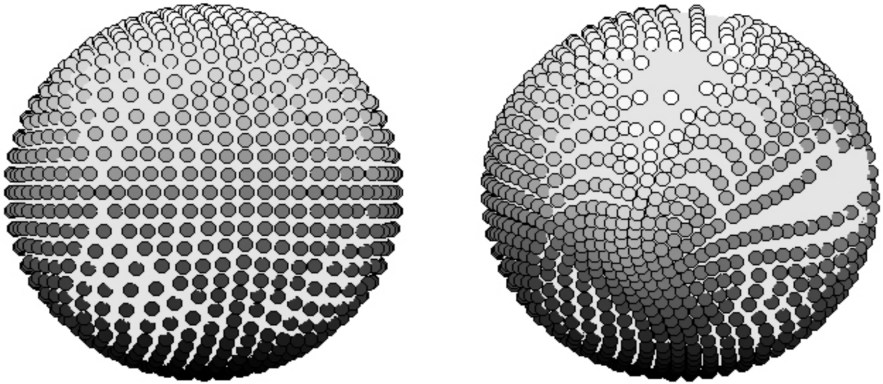


Fig. 4. *Left:* Original sphere. *Right:* Sphere after combined transformation.

approximation can be obtained by recursive subdivision of a polyhedron (initially an octohedron) and projecting the points onto the sphere surface. At each level of recursion, each triangle is divided into 4 smaller triangles by placing 3 new vertices halfway along each edge - this gives $12 \times 4^{k-2}$ new kernels on the k^{th} level of recursion.

We choose to use a multiresolution approach to the optimisation. The basic idea is to begin with broad Cauchies and to iteratively refine the parameterisation by introducing additional, narrower Cauchies between the existing ones. We have found, by cross validation, that the best value for the width parameter, is $a_k = 1/2^{k-2}$, i.e. the widths are halved at each level of recursion.

A local optimisation algorithm is employed to find the optimum magnitude of each kernel. At each level of recursion, we have $12 \times 4^{k-2}$ kernels for each shape, creating a $12n_s \times 4^{k-2}$ -dimensional configuration space. This is generally too difficult to optimise robustly and reliably. We have found, however, that optimising the magnitude of a single kernel in turn on each shape gives better results and is substantially quicker. We used the Nelder-Mead simplex algorithm [17] to obtain the results reported in section 6.

6 Results

We present qualitative and quantitative results of applying our method to a training set of 16 rat kidneys and 8 anterior horns of brain ventricles. In each case the shapes were segmented by hand from a set of 3D magnetic resonance images. The algorithm was run for three levels of recursion, giving a total of 66 kernels per shape. We compare the results to models built by uniformly sampling the surface.

In figures 5 and 6, we show qualitative results by displaying the variation captured by the first three modes of each model (b^m varied by ± 2 [standard deviations seen across the training set]). We show quantitative results in

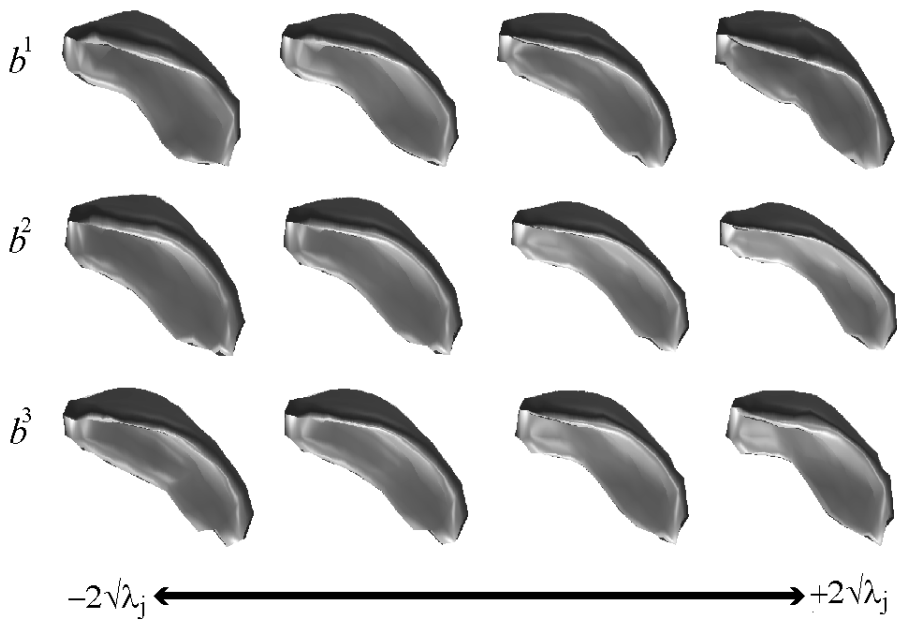


Fig. 5. The first three modes of variation $\pm 2\lambda^m$ of the automatically produced model of the brain ventricle

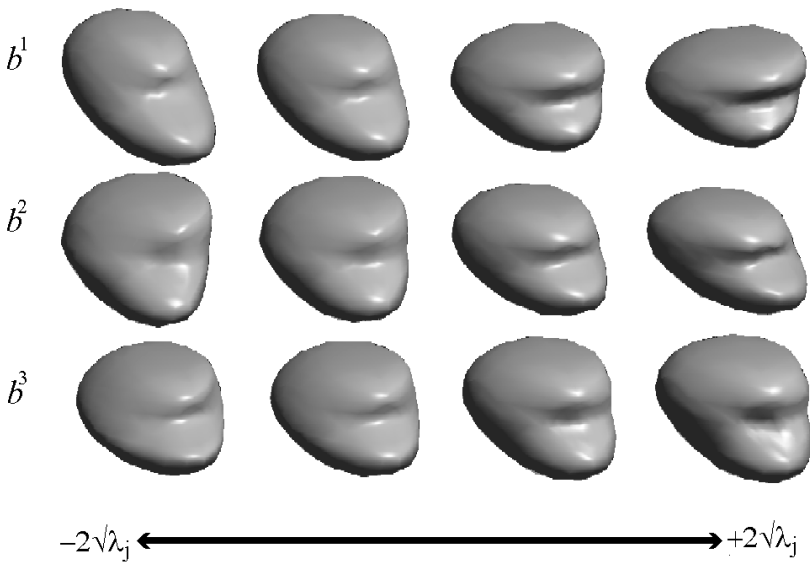


Fig. 6. The first three modes of variation $\pm 2\lambda^m$ of the automatically produced model of the rat kidneys

Table 1. A quantitative comparison of each model showing the variance explained by each mode. F is the value of the objective function and V_T is the total variance.

Kidneys			Ventricles		
Mode	Automatic	Uniform	Mode	Automatic	Uniform
1	3.06	3.23	1	11.98	12.58
2	2.75	3.01	2	1.41	1.77
3	1.35	1.70	3	1.16	1.49
4	0.94	1.08	4	0.69	0.79
5	0.76	0.87	5	0.36	0.45
6	0.46	0.57	6	0.22	0.24
V_T	10.51	11.98	V_T	15.90	17.46
F	1396	1417	F	366	371

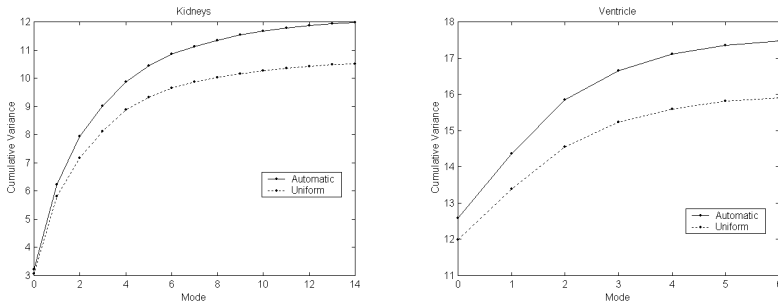


Fig. 7. A plot showing the cumulative variance described by each mode of the model. This is a measure of the compactness of each model.

table 1, tabulating the variance explained by each mode, the total variance and the value of F . In both cases F is substantially better for the optimised model. Figure 7 shows the cumulative variance plotted against the number of modes used for each model; this measures the compactness of the model.

The plots show that, for the entire range of modes, the optimised models are substantially more compact than those obtained by uniformly-sampling the surface. To test the generalisation ability of the models, we performed leave-one-out tests on each model. In figure 8 we show the approximation error for representing an unseen shape as a function of the number of modes used in the reconstruction. The optimised models perform substantially better than the uniformly sampled models whatever the number of modes used, demonstrating superior generalisation ability.

7 Discussion and Conclusions

We have described a method for building 3D statistical shape models by automatically establishing optimal correspondences between sets of shapes. We have shown that the method produces models that are more compact than those

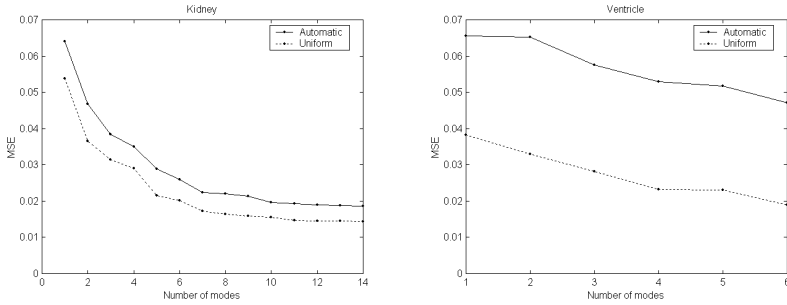


Fig. 8. Leave one out tests on the models. The plot shows the number of modes used against the mean squared approximation error. This measures the ability of each model to generalise to *unseen* examples.

based on uniformly-sampled shapes and have substantially better generalisation ability.

We have described a novel method of reparameterising closed surfaces. The method guarantees that the resulting reparameterisation is homeomorphic - an essential constraint. With a suitable choice of boundary conditions, the representation can also be applied to reparameterise open surfaces.

Although we have only reported results on a relatively small set of example shapes, we believe the method scales easily to deal with sets of 40-50 shapes. The local optimisation algorithm employed will, however, be substantially slower. We are currently looking for more robust and faster methods of finding the minimum.

The MDL objective function and optimisation approach provides a unified framework for dealing with all aspects of model building. For example, we plan to investigate including the alignment of the shapes in the optimisation.

Appendix: Description Length for One-Dimensional Centred Gaussian Models

In this Appendix, we show how to construct an expression for the description length required to send a one-dimensional, centred data set using a Gaussian model. Our data model is the family of centred Gaussian distributions:

$$\rho(y; \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{y^2}{2\sigma^2}\right). \quad (26)$$

Following the two-part coding scheme [19], the total description length is computed as the sum of two parts; the description length for sending the value and accuracy of the parameter σ , and the description length for coding the data according to this model. To calculate these description lengths, we use the fundamental result that the ideal-coding codeword length for a discrete event A , encoded using a statistical model with associated event probabilities $P(A)$ is given by the Shannon Coding codeword length [22]:

$$\mathcal{L}(A; P) = -\log_2 P(A) \text{ bits, or: } -\ln P(A) \text{ nats.}^2 \quad (27)$$

We take our centred data set to be $Y = \{y_i : i = 1 \text{ to } n_s\}$, where the data is known to lie within a strictly bounded region. To reduce the continuum values $\{y_i\}$ to a set of discrete events, we 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:

$$-R < \hat{y} < R \text{ and } \hat{y} = m\Delta, \quad m \in \mathbb{Z}. \quad (28)$$

Coding the Parameters

We will assume that the parameter σ is described to an accuracy δ . We will also assume that our now quantized parameter $\hat{\sigma}$ is bounded, hence has the allowed values:

$$\hat{\sigma} = n\delta, \quad n \in \mathbb{N} \text{ and } \sigma_{min} \leq \hat{\sigma} \leq \sigma_{max}. \quad (29)$$

Given the absence of any prior knowledge, we assume a flat distribution for $\hat{\sigma}$ over this range, which gives us the codeword length:

$$\mathcal{L}_{\hat{\sigma}} = \ln \left(\frac{\sigma_{max} - \sigma_{min}}{\delta} \right). \quad (30)$$

Note that our receiver cannot decrypt the value of $\hat{\sigma}$ without knowing the value of δ . If we assume that the quantization parameter δ is of the form:

$$\delta = 2^{\pm k}, \quad k \in \mathbb{N} \quad (31)$$

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

$$\mathcal{L}_{\delta} = 1 + |\log_2 \delta| \text{ bits} \approx 1 + |\ln \delta| \text{ nats} \quad (32)$$

where the additional bit/nat codes for the sign in the exponent of δ . The total code length for transmitting the parameters is then given by $\mathcal{L}_{\hat{\sigma}} + \mathcal{L}_{\delta}$.

Coding the Data

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

$$P(\hat{y}) = \int_{\hat{y}-\Delta/2}^{\hat{y}+\Delta/2} dk \rho(k; \hat{\sigma}) \approx \frac{\Delta}{\hat{\sigma}\sqrt{2\pi}} \exp(-\hat{y}^2/2\hat{\sigma}^2). \quad (33)$$

² In what follows, we will restrict ourselves to natural logarithms, and codeword lengths in units of nats. However, expressions can be easily converted to the more familiar binary lengths by noting that 1 bit $\equiv \ln 2$ nats.

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

It can be shown numerically that this is a very good approximation (mean fractional error $\leq 1.0\% \pm 0.8\%$) for all values $\hat{\sigma} \geq 2\Delta$, hence we will take $\sigma_{min} = 2\Delta$. The code length for the data is then:

$$\mathcal{L}_{data} = -\sum_{i=1}^{n_s} \ln P(\hat{y}_i) = -n_s \ln \Delta + \frac{n_s}{2} \ln(2\pi\hat{\sigma}^2) + \frac{1}{2\hat{\sigma}^2} \sum_{i=1}^{n_s} \hat{y}_i^2 \quad (34)$$

The variance of the quantized data is:

$$\sigma^2 = \frac{1}{n_s} \sum_{i=1}^{n_s} \hat{y}_i^2 \text{ and } \sigma_{max} = R. \quad (35)$$

In general σ differs from the nearest quantized value $\hat{\sigma}$ thus:

$$\hat{\sigma} = \sigma + d_\sigma, \quad |d_\sigma| \leq \frac{\delta}{2}. \quad (36)$$

So, averaging over an ensemble of data sets, and assuming a flat distribution for d_σ over this range, we find:

$$\langle d_\sigma^2 \rangle = \frac{\delta^2}{12}, \quad \frac{1}{\hat{\sigma}^2} \approx \frac{1}{\sigma^2} \left(1 + \frac{\delta^2}{4\sigma^2} \right), \quad \langle \ln \hat{\sigma}^2 \rangle \approx \ln \sigma^2 - \frac{\delta^2}{12\sigma^2}. \quad (37)$$

By substituting these expressions into (34) and using (35) gives the following expression for the Description Length of the data:

$$\mathcal{L}_{data} = -n_s \ln \Delta + \frac{n_s}{2} \ln(2\pi\sigma^2) + \frac{n_s}{2} + \frac{n_s\delta^2}{12\sigma^2}. \quad (38)$$

The total Description Length for the parameters and the data is then:

$$\mathcal{L}^{(1)} = 1 + \ln \left(\frac{\sigma_{max} - \sigma_{min}}{\delta} \right) + |\ln \delta| - n_s \ln \Delta + \frac{n_s}{2} \ln(2\pi\sigma^2) + \frac{n_s}{2} + \frac{n_s\delta^2}{12\sigma^2}. \quad (39)$$

By differentiating w.r.t. δ and setting the derivative to zero, we find that the optimum parameter accuracy δ is:

$$\delta^*(\sigma, n_s) = \min \left(1, \sigma \sqrt{\frac{12}{n_s}} \right) \quad (40)$$

which then allows us to write the above expression as:

$$\mathcal{L}^{(1)} = D^{(1)}(\sigma, n_s, R, \Delta). \quad (41)$$

In the case where $\sigma < \sigma_{min}$, but the quantized 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:

$$\begin{aligned} \mathcal{L}^{(2)} = & 1 + \ln \left(\frac{\sigma_{max} - \sigma_{min}}{\delta} \right) + |\ln \delta| - n_s \ln \Delta + \frac{n_s}{2} \ln(2\pi\sigma_{min}^2) \\ & - \frac{n_s\delta^2}{24\sigma_{min}^2} + \frac{n_s\sigma^2}{2\sigma_{min}^2} \left(1 + \frac{\delta^2}{4\sigma_{min}^2} \right) \equiv D^{(2)}(\sigma, n_s, R, \Delta). \end{aligned} \quad (42)$$

The only remaining case is where all the quantized data lies in one bin at the origin. This requires no further information to describe it fully, hence has a description length of zero.

Acknowledgements. The authors would like to thank Alan Brett and Johann Kim for their contribution to this work. Rhodri Davies would like to acknowledge the BBSRC and AstraZeneca⁴ for their financial support.

References

1. Angenent, S., S. Haker, A. Tannenbaum and R. Kikinis: "On the laplace-beltrami operator and brain surface flattening". *IEEE Trans. Medical Imaging*, 18: p. 700-711. 1999
2. Brechbuhler, C., G. Gerig and O. Kubler, "Parameterisation of closed surfaces for 3-D shape description", in *Computer Vision and Image Understanding*, 61(2), p154-170, 1995
3. Brett A.D. and C.J. Taylor, "A Method of automated landmark generation for automated 3D PDM construction", in *Image and Vision Computing*, 18(9), p. 739-748, 2000
4. Brett A.D. and C.J. Taylor, "Construction of 3D Shape Models of Femoral Articular Cartilage Using Harmonic Maps", in *MICCAI'00*, p. 1205-1214, 2000
5. Cootes, T., A. Hill, C. Taylor, and J. Haslam, "The use of Active shape models for locating structures in medical images". *Image and Vision Computing*, 12: p. 355-366. 1994.
6. Cootes, T., C. Taylor, D. Cooper and J. Graham, "Active shape models - their training and application". *Computer Vision and Image Understanding*, 61: p. 38-59. 1995.
7. Christensen, G. E., S. C. Joshi, M.I. Miller, "Volumetric Transformation of Brain Anatomy", in *IEEE Trans. Medical Imaging*, 16, p864-877, 1997
8. Davies, Rh. H., C.J. Twining, T.F. Cootes, J.C. Waterton and C.J. Taylor, "A Minimum Description Length Approach to Statistical Shape Modelling", in *IEEE trans. Medical Imaging*, To appear
9. Davies, Rh. H., T.F. Cootes, C.J. Twining and C.J. Taylor, "An Information Theoretic Approach to Statistical Shape Modelling", in *British Machine Vision Conference - BMVC'01*, p 3-12, 2001
10. Flueté, M., S. Lavalee, "Building a Complete Surface Model from Sparse Data Using Statistical Shape Models: Application to Computer Assisted Knee Surgery", in *MICCAI'98*, p878-887, 1998
11. M. H. Hansen and B. Yu, "Model Selection and the Principle of Minimum Description Length", Technical Memorandum, Bell Labs, Murray Hill, N.J. 1998.
12. Hill, A., A. Thornham and C.J. Taylor, "Model based interpretation of 3D medical images", in *British Machine Vision Conference, BMVC'93*, p339-348, 1993
13. Joshi *et. al* , "Gaussian Random Fields on Sub-Manifolds for Characterizing Brain Surfaces", in *IPMI'97*, p381-386, 1997
14. Kelemen, A., G. Szekely, and G. Gerig, "Elastic model-based segmentation of 3-D neuroradiological data sets". *IEEE Transactions On Medical Imaging*, 18(10): p. 828-839. 1999.

⁴ AstraZeneca, Alderley Park, Macclesfield, Cheshire, UK

15. Kotcheff, A.C.W. and C.J. Taylor, "Automatic Construction of Eigenshape Models by Direct Optimisation". *Medical Image Analysis*, 2: p. 303-314. 1998.
16. Mardia, K. V., *Statistics of Directional Data*, Academic Press, London, 1972
17. Press, W.H., S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, *Numerical Recipes in C*, Cambridge University Press; 1993
18. Rissanen, J. R. "A universal prior for integers and estimation by minimum description length", *Annals of Statistics*, vol. 11, no. 2, pp. 416-431, 1983.
19. Rissanen, J. R. *Stochastic Complexity in Statistical Inquiry*, World Scientific Series in Computer Science, Vol. 15, World Scientific Publishing Co., Singapore, 1989.
20. Rissanen, J. R. "Fisher Information and Stochastic Complexity", *IEEE Transactions on Information Theory*, vol. 42, no. 1, pp. 40-47, 1996.
21. Rueckert, D., Frangi, F., Schnabel, J.A., "Automatic construction of 3D statistical deformation models using non-rigid registration", *MICCAI'01* p77-84. 2001.
22. Shannon, C. E., "A mathematical theory of communication," , *Bell System Technical Journal*, vol. 27, pp. 379-423 and 623-656, 1948.
23. Szeleky, G., S. Lavalee, "Matching 3-D Anatomical Surface with Non-Rigid Deformations using Octree-Splines", in *International Journal of Computer Vision*, 18(2), p171-186, 1996
24. D. Tosun and J.L. Prince, "Hemispherical map for the human brain cortex," *Proc. SPIE Medical Imaging 2001*, p 290-300. 2001.
25. Wang, Y., B. S. Peterson, and L. H. Staib. "Shape-based 3D surface correspondence using geodesics and local geometry". *CVPR 2000*, v. 2: p. 644-51. 2000.