

Diffeomorphic Statistical Shape Models

T.F. Cootes, C.J. Twining and C.J. Taylor
Imaging Science and Biomedical Engineering,
University of Manchester, M13 9PT, UK
t.cootes@man.ac.uk

Abstract

We describe a method of constructing parametric statistical models of shape variation which can generate continuous diffeomorphic (non-folding) deformation fields. Traditional statistical shape models are constructed by analysis of the positions of a set of landmark points. Here we analyse the parameters of continuous warp fields, constructed by composing simple parametric diffeomorphic warps. The warps are composed in such a way that the deformations are always defined in a reference frame. This allows the parameters controlling the deformations to be meaningfully compared from one example to another. A linear model is learnt to represent the variations in the warp parameters across the training set. This model can then be used to generalise the deformations. Models can be built either from sets of annotated points, or from unlabelled images. In the latter case, we use techniques from non-rigid registration to construct the warp fields deforming a reference image into each example. We describe the technique in detail and give examples of the resulting models.

1 Introduction

This paper proposes a method of building generative statistical models of diffeomorphic (ie smooth, invertible) deformation fields. It is part of a larger programme aimed at exploiting the synergy between work on constructing statistical models of shape and that on non-rigid registration.

Statistical models of shape variation [5] have been shown to be powerful tools for image interpretation. Most approaches to constructing such models assume (either explicitly or implicitly) that the object shape can be represented by a set of points which exist on all examples, and essentially define the correspondences across those examples.

The act of annotating a training set with such points is essentially the same as the goal of non-rigid image registration, widely used in the medical image analysis community. In such cases the aim is to find the deformation field which maps one image into another. Such a field defines a dense correspondence between two images. Given a set of such dense correspondences, one can build a statistical model of the deformation field [11].

When registering two images one usually assumes that similar structures are present in each. Since in general the ordering should not be important, it is desirable that the deformation field be smooth and invertible (so that every point in one image has a corre-

sponding point in the other). Such smooth, invertible transformations are called *diffeomorphisms*.¹

In the non-rigid registration literature there are a variety of ways of constructing diffeomorphic maps between pairs of images. The diffeomorphism is either achieved by placing constraints on the Jacobian of more arbitrary warps [3, 11, 2] or by composing many simple diffeomorphisms [6].

There has been little attention paid to generating parameterisable diffeomorphisms suitable for representing the statistics of shape change across a *set* of examples.

Point-based statistical models of shape [5] give us a parameterised representation of the movement of a set of landmark points. If we could interpolate the deformation field between the points we would have a statistical model of continuous deformation. Unfortunately, interpolating in such a way as to ensure the resulting mapping is diffeomorphic is difficult.

For instance, the simplest approach, that of piecewise linear interpolation using a triangulation of the points, is clearly unsatisfactory. Not only is the resulting deformation field not smooth, it is quite likely to fold up on itself (see for instance, Figure 1).

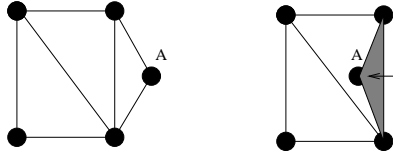


Figure 1: Problems when interpolating with triangles. If point A moves as shown, the triangle flips and there is no longer a 1-1 mapping between the two regions.

Smoother interpolation schemes such as Thin-Plate Splines [1] are also prone to tearing space. However, Twining *et al.* describe Geodesic Interpolating Splines, which are capable of constructing a diffeomorphism between two sets of landmark points [13]. Unfortunately the method requires a relatively complex optimisation, so is not very efficient for large numbers of points.

Thus a conceptually simple approach, that of using a standard point based statistical model, then interpolating to estimate the diffeomorphism between the mean shape and points generated by the model, can be rather slow.

In this work we describe how we can construct statistical models of continuous deformation fields in such a way that they can only generate smooth invertible mappings. The key to the approach is to create parameterised deformations by composing simple diffeomorphic functions, then to apply statistics in the space of the parameters, rather than to the point positions directly. It is possible to generate functions such that linear interpolation in the parameter space always leads to legal deformations.

In the following we will describe the construction of parameterised diffeomorphic warps, how their parameters can be estimated and how one can model their parameters given a training set of registered data. We will give examples of the models and discuss how such models can be used.

¹In cases where structures appear or disappear between one image and the next, we believe that these should be explicitly modelled as creation or destruction processes – such processes will not be addressed in this paper.

1.1 Related Work

Rueckert *et al.*[11] describe statistical shape models in which deformation is represented by the control points of a B-spline. Since the B-spline is linear in the control point positions, the resulting model is essentially identical to the original PDMs of Cootes *et al.*[5]. Such a model is prone to generating non-diffeomorphic deformations. However, one could use the constraints on the control point displacements described by Chio and Lee [2] to ensure that B-spline based deformations are diffeomorphic.

The work by Pizer’s group on medial representations (M-Reps) [12] explores modelling compound structures and their deformation. In work by Fletcher *et al.*[8] they investigate statistically modelling the variation in shape in such a way as to preserve correspondence.

2 Representing Diffeomorphisms

We can construct complex diffeomorphic functions by composing simple diffeomorphisms.

Let $f(\mathbf{x}|\phi_i)$ be a diffeomorphic mapping controlled by parameters ϕ_i . We define $f_1 f_2(\mathbf{x}) = f_1(f_2(\mathbf{x}))$.

We can define the composition of a set of diffeomorphisms as

$$F(\mathbf{x}|\Phi) = f_1 f_2 \dots f_{n-1} f_n(\mathbf{x}) \quad (1)$$

where the parameter vector Φ is simply a concatenation of the parameter vectors for each individual function, $\{\phi_i\}$. As long as the component functions are all diffeomorphic, so is F .

In the following we will consider diffeomorphisms of the form

$$f_A F(\mathbf{x}|\Phi) = f_A f_1 f_2 \dots f_{n-1} f_n(\mathbf{x}) \quad (2)$$

in which $f_A(\mathbf{x})$ is an affine transformation, and each f_i is a parameterised diffeomorphic of increasing complexity and decreasing scale. In particular, we use grid based deformations, in which the deformation of a region is controlled by the position of a set of grid nodes (see Appendix A). We use a coarse grid with few nodes for f_1 , and increase the number of nodes for each subsequent finer deformation f_i .

Note that these deformations are similar to the B-spline free form deformations used by Rueckert *et al.*[11]. However, that work effectively represents the deformation field with the final, single grid based warp, and doesn’t necessarily force the result to be diffeomorphic.

The ordering of the transformations is important. To apply a warp to a structure in the reference frame, we first apply the finest (small scale) deformations f_n . We then apply deformations of increasing scale. Finally we apply an affine transformation. The grid based warps described in Appendix A are controlled by the position of the nodes. If we apply the deformations in a fine-to-coarse manner, each grid is always defined in the reference frame, and the displacements of each node can be represented in this frame. This means that it is reasonable to compare the parameters for different deformations, as they are computed in the same domain. It is this property which will allow us to perform meaningful analysis on the parameter vectors when building the models of deformation (see the next section).

This fine-to-coarse approach can be considered as applying a series of object-centred deformations. We first make small changes to the object, then make increasingly large changes, carrying the smaller changes along - see Figure 2. If we applied the coarser warps first then the finer warps would affect different parts of the reference object each time, depending upon which coarse warps are used.

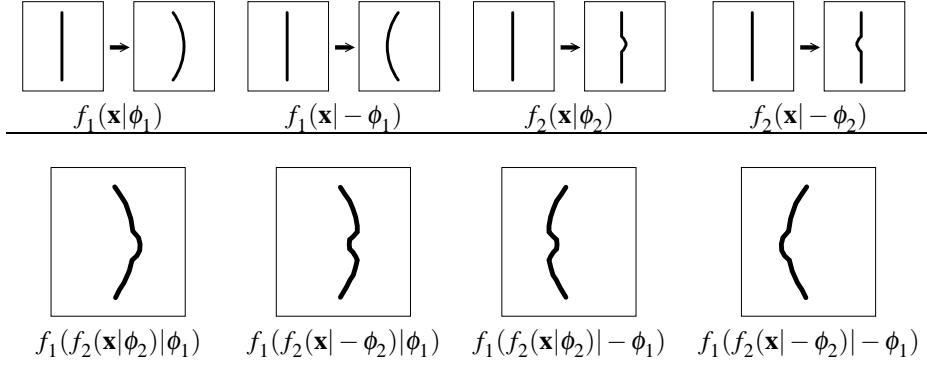


Figure 2: Example of composing warps. f_1 is a coarse warp, f_2 a fine scale warp. By applying f_2 first we make small changes to the shape, which are then carried along when we apply f_1 .

3 Estimating Diffeomorphisms

We will consider estimating the parameters for the compositional warps described above for two cases; one in which we have a set of landmark positions on a reference shape and a target shape, and one in which we have only un-annotated images (one reference, one target). In both cases we wish to find the parameters of the compositional warp, ϕ_A (the affine component) and Φ (the non-rigid components), which best match the reference to the target data.

3.1 Estimating from Sets of Points

Suppose we have a set of points on a reference shape. Assume (for notational convenience) that their co-ordinates are concatenated together into the vector \mathbf{X} . Suppose we also have a set of points in the target shape, \mathbf{X}' . We wish to find the parameters which minimise $|f_A F(\mathbf{X}|\Phi) - \mathbf{X}'|^2$.

For small numbers of points there are likely to be multiple possible solutions, as similar deformations can be represented by grids at different scales. We thus tackle the problem in a greedy fashion:

- Find the affine parameters ϕ_A which minimise $|f_A(\mathbf{X}) - \mathbf{X}'|^2$
- For fixed ϕ_A find ϕ_1 to minimise $|f_A f_1(\mathbf{X}) - \mathbf{X}'|^2$
- For fixed ϕ_A, ϕ_1 find ϕ_2 to minimise $|f_A f_1 f_2(\mathbf{X}) - \mathbf{X}'|^2$
- Repeat for subsequent functions, fixing previous parameters

3.2 Estimating from Images

Given a pair of unlabelled images, I_0 and I_1 , we wish to find the parameters of the deformation which best matches one onto the other. We will use a modified version of Rueckert's free-form non-rigid registration algorithm [11]. In the results section below we demonstrate that the compositional technique described here is able to accurately register a pair of images.

Let \mathbf{X} be a set of points densely placed on image I_0 (for instance, the pixel positions). Let $\mathbf{s}_0 = I_0(\mathbf{X})$ be the intensities sampled at these positions in image I_0 . Let $\mathbf{s}_1 = I_1(W(\mathbf{X}))$ be the intensities sampled at warped positions, $W(\mathbf{X})$ in image I_1 .

We will assume that the quality of match between two images is evaluated using a function $Q(\mathbf{s}_0, \mathbf{s}_1)$. In the experiments which follow we use a simple sum of absolute differences, which gives some robustness to outliers. However we could equally use mutual information (MI), normalised MI, or other statistical measures.

The registration process proceeds as follows:

- Find the parameters ϕ_A of the affine transformation which best maps I_0 onto I_1 by minimising $Q(\mathbf{s}_0, I_1(f_A(\mathbf{X}|\phi_A)))$
- Set the estimate of the deformation to be $F_0(\mathbf{X}) = f_A(\mathbf{X})$.
- Generate a new image, $I_1^{(1)}$ by applying the inverse of the transformation $F_0(\mathbf{X})$ to I_1 . This can be done by simply writing the samples $\mathbf{s}_1 = I_1(F_0(\mathbf{X}))$ into the grid positions \mathbf{X} in a new image. $I_1^{(1)}$ is thus the image I_1 projected back into the frame of image I_0
- Find the parameters ϕ_1 of the diffeomorphism $f_1(\mathbf{X}|\phi_1)$ which best maps I_0 onto $I_1^{(1)}$ by minimising $Q(\mathbf{s}_0, I_1^{(1)}(f_1(\mathbf{X}|\phi_1)))$
- Set the estimate of the deformation to be $F_1(\mathbf{X}) = F_0 f_1(\mathbf{X})$.
- Generate a new image, $I_1^{(2)}$ by applying the inverse of the transformation $F_1(\mathbf{X})$
- Repeat, optimising each transformation f_i in turn, updating $F_i = F_{i-1} f_i$ and recomputing the back-warped image $I_1^{(i+1)}$

The computation of the intermediate images $I_1^{(1)}$ is necessary for efficiency. For the optimisations with few parameters, we use Nelder-Mead Downhill Simplex [10]. For the larger grid optimisations we estimate the gradient of the objective function by displacing each node in turn, then perform a line search along that direction (the approach used by Rueckert *et al.* [11]).

The final result is the set of all parameters of the warp $F_n(\mathbf{X}) = f_A f_1 \dots f_n(\mathbf{X})$.

This approach gives us a method of registering a pair of images using a diffeomorphic mapping. In addition, it ensures that since at each stage the warps are defined in the reference frame (image I_0), parameters from different warps to different images can be compared.

4 Statistical Models of Diffeomorphisms

Suppose we have a set of training images. If we choose one as a reference, and choose a suitable class of compositional warps (such as the grid-based deformations), we can use the method described above to find the diffeomorphic deformation from the reference image into every other image. Each such deformation is summarised by the affine parameters, $\phi_{A,i}$ and the parameters controlling the non-rigid warps, concatenated into vectors Φ_i .

Shape is usually defined as the geometric properties of an object invariant to some transformation (commonly similarity or affine). When constructing a statistical shape model from a set of points, Procrustes Analysis [7] is used to remove the effect of such global transformation. In our case the affine transformation is encoded in the parameters ϕ_A . Since we are usually only interested in the non-rigid components, we will discard the affine terms (note however that in some cases we may wish to retain information about size or skew, in which case we can decompose the affine component appropriately).

To build a statistical model, we simply apply Principal Component Analysis (PCA) to the warp parameter vectors Φ_i . Where appropriate we should apply a pre-scaling to the elements of each Φ_i so that they are commensurate. This scaling should be chosen so that changing each parameter causes a similar amount of movement of the resulting warp. For instance, with grid based warps, moving one node in a 3 x 3 grid by one unit has a four times larger effect than moving one node by the same amount in a 6 x 6 grid. Therefore the parameters controlling the coarser grid should be weighted 4 times more than those of the finer grid.

By selecting a suitable subset of the principal components, we generate a compact linear model with a relatively small number of modes;

$$\Phi = \hat{\Phi} + \mathbf{P}\mathbf{b} \quad (3)$$

where $\hat{\Phi}$ is the mean of the training set, \mathbf{P} contains the t most significant eigenvectors and \mathbf{b} is the vector of shape parameters.

This model can be used to generate new diffeomorphic warps by first computing the warp parameters, then plugging these into the composition of warps. To ensure that the result is indeed diffeomorphic, we must apply constraints to the parameters Φ . In the case of grid-based warps, this simply involves testing each parameter to ensure it is within a fixed range (see Appendix A). The linear nature of the model ensures that we can use the approach to interpolate between two known diffeomorphic warps by moving along a straight line in Φ -space. This will not generate any illegal parameter values.

The full model of deformation is thus controlled by the global affine parameters, ϕ_A and the model parameters \mathbf{b} ,

$$W(\mathbf{x}) = f_A(F(\mathbf{x}|\hat{\Phi} + \mathbf{P}\mathbf{b})|\phi_A) \quad (4)$$

Examples of such models are shown below.

4.1 Examples from Sets of Points

Figure 3 shows the first three modes of variation of a model trained on deformations estimated from sets of points annotated on a set of face images. A composition of four

grid-based warps was used, with 2x2, 4x4, 8x8 and 16x16 nodes respectively. 51 faces images taken from the XM2VTS database were used [9]. The nodes at the edges are assumed fixed, so these warps have 2, 18, 98 and 450 degrees of freedom respectively, giving a total of 528. The model explaining 95% of the variation has 33 degrees of freedom.

Figure 3 shows the continuous deformation of space (using an arbitrary grid) and the effect on the points describing the main features in the reference image.

For comparison Figure 4 shows the first 3 modes of a linear statistical shape model trained on the point positions (a PDM) [5] (in this case 21 dimensions are sufficient to capture 95% of the variance in the training set). The diffeomorphic mode has captured the same sorts of variation.

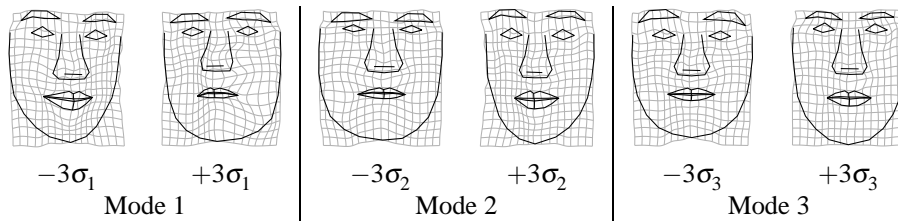


Figure 3: Modes of diffeomorphic shape model trained on sets of face points

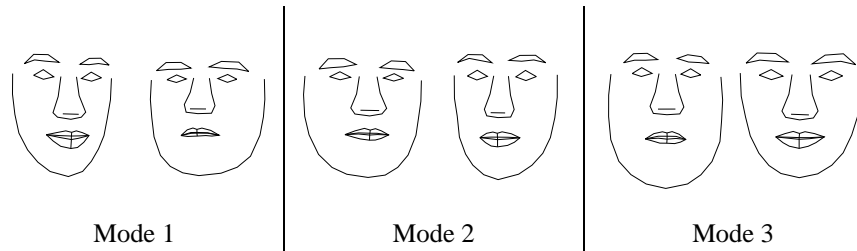


Figure 4: Modes of shape model (PDM) trained on sets of face points

4.2 Examples from Unlabelled Images

We used the non-rigid registration technique described in section 3.2 to find the warps from the first face image to subsequent images. We then built a statistical model of the warp parameters. Figure 5 shows the first three modes of the resulting model of diffeomorphic shape variation. The method has successfully captured similar forms of variation to the models built from annotated landmarks. However, the variation seems a little less 'extreme' (eg around the chin), suggesting that the non-rigid registration has been more conservative when matching.

We can evaluate the quality of the non-rigid registration by projecting the reference points to each target image in turn using the found deformation field. These points can then be compared with manual annotations on the same images. We measure the distance from each point to the equivalent curve on the target image, and obtain a mean error of

1.7 pixels (s.d.0.6 pixels). The faces are about 100 pixels wide. This suggests that the compositional method is giving good results.

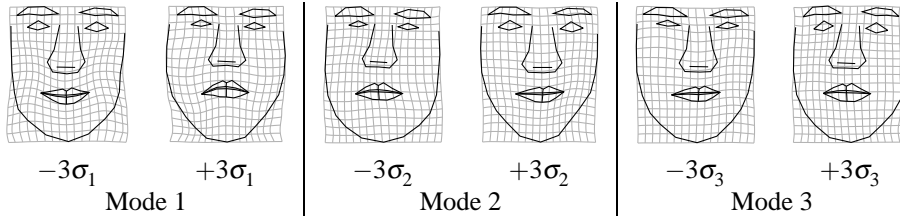


Figure 5: Modes of diffeomorphic shape model trained on warps from non-rigid registration of face images

Figure 6 shows the model resulting from applying the technique to a set of 16 MR brain slices. Again, the method has automatically registered the images and found the main modes of variation. Comparing the registration with manual annotation for the central brain structures, we find a point to curve error with a mean of 1.7 pixels (s.d. 0.5 pixels). The brains are about 150 pixels across.

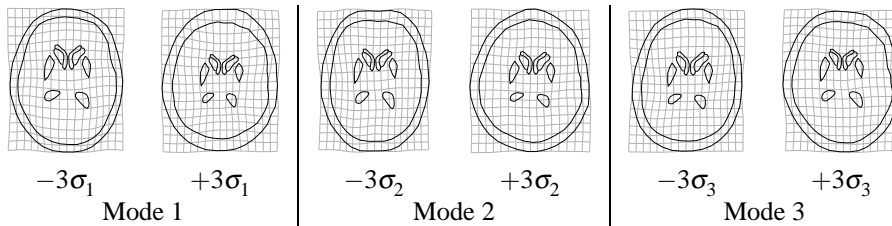


Figure 6: Modes of diffeomorphic shape model trained on warps from non-rigid registration of MR brain slice images

5 Discussion

The key to the models is the method of generating parameterised classes of diffeomorphisms by composing relatively simple basis functions. Though we have presented results with a particular type of grid-based interpolating warp, a wide range of alternatives could be used. One obvious choice would be the B-splines used to such success by Rueckert *et al.*[11], as long as suitable constraints are used to ensure diffeomorphisms [2].

Of course, a different choice of basis functions will lead to a different model. It is hoped that models built with different but suitably expressive bases will be able to generate similar sorts of deformation fields, though this remains an area of active investigation.

The grid-based warps limit the sorts of deformation that the method can generate. This means it cannot necessarily get a perfect match to known corresponding points. However, using sufficiently fine grids should allow the matching to be 'close enough'.

Though results here are presented for 2D examples, the extension to 3D is straightforward.

The compositional non-rigid registration method described in Section 3.2 was shown to give accurate results, and will be useful in its own right. However, it is a greedy algorithm and potentially prone to local minima. In further work we will investigate its performance more and compare it with alternative registration schemes.

The eventual aim is to use the models in image interpretation. The models appear capable of representing shape variation in a compact form, and should allow fast robust model matching to new images, in the same way that linear models of point position have proved very useful. We are exploring using fast gradient approximation techniques such as that exploited by the Active Appearance Model algorithm [4] for driving the image matching with the models. It is hoped that this will lead to more rapid estimation of deformation between images.

6 Conclusions

We have described a method of modelling diffeomorphic shape deformation statistically, leading to a parameterised model capable of synthesising diffeomorphisms efficiently. The resulting models have modes of variation similar to those of linear models of point position, but unlike the latter, explicitly define a full diffeomorphic deformation field. We anticipate that the models will find wide application in the fields of modelling and interpreting deformable objects, particularly in the medical field.

Acknowledgements

The authors would like to thank their collaborators in the MIAS-IRC Project for many fruitful discussions.

A Grid-Based Diffeomorphisms

We consider diffeomorphic warp fields represented by deformations of a grid of control points.

The general approach is that the nodes of a regular grid are displaced by some amount, and that displacement is propagated (or interpolated) to estimate the displacement at intermediate positions.

Consider the case in which we are interested in the region $[0, n+1][0, m+1]$, into which we have placed $n \times m$ nodes at integer positions. If the nodes at (i, j) in the reference frame are displaced by $\mathbf{d}_{i,j}$, then this implies a deformation of the region as

$$\begin{aligned} \mathbf{d}(x, y) = & k(x-i)k(y-j)\mathbf{d}_{i,j} && +k(i+1-x)k(y-j)\mathbf{d}_{i+1,j} \\ & +k(x-i)k(j+1-y)\mathbf{d}_{i,j+1} && +k(i+1-x)k(j+1-y)\mathbf{d}_{i+1,j+1} \end{aligned} \quad (5)$$

where $i \leq x < i+1$ and $j \leq y < j+1$, and $k(r)$ is a kernel function used to interpolate the deformation. Note that if $k(r) = 1 - |r|$, then we have bilinear interpolation. However, this is not diffeomorphic since it is not smooth (there are discontinuities in the gradient at integer boundaries).

For a smooth, invertible mapping, it can be shown that $k(r)$ must satisfy

$$\begin{aligned} k(0) = 1 & & k(-1) = k(1) = 0 \\ k'(-1) = k'(1) = 0 & & k(r) = k(-r) \end{aligned} \quad (6)$$

and that certain limits are placed on the displacements of the nodes.

The simplest polynomial kernel with these properties is $k(r) = (1 - r^2)^2$.

An interesting alternative function is

$$k(r) = 0.5(1 + \cos(\pi r)) \quad (7)$$

The nice thing about the latter function is that $k(r) + k(1 - r) = 1$, which means that if four neighbouring nodes are displaced by the same amount then all the points in the enclosed square are translated by that amount, with no other distortions occurring.

By examining the Jacobian of the deformation (5) we can show that with this trigonometric kernel the deformation is guaranteed to be diffeomorphic if no node is displaced more than $\frac{1}{\pi} \approx 0.3183$ along x or y . Thus we can construct a parameterised diffeomorphism of the form $f(\mathbf{x}|\phi) = \mathbf{d}(x, y)$ where the parameters ϕ are the $2nm$ ordinates of the displacements of the nodes. Note that this can easily be extended to 3D if required.

References

- [1] Fred L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6):567–585, 1989.
- [2] Y. Choi and S. Lee. Injectivity conditions of 2d and 3d uniform cubic b-spline functions. *Graphical Models*, 62(6):411–427, 2000.
- [3] G. Christensen. Consistent linear-elastic transformations for image matching. In *16th Conference on Information Processing in Medical Imaging*, pages 224–237, Visegrád, Hungary, June 1999.
- [4] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In H. Burkhardt and B. Neumann, editors, *5th European Conference on Computer Vision*, volume 2, pages 484–498. Springer, Berlin, 1998.
- [5] T. F. Cootes, C. J. Taylor, D.H. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, January 1995.
- [6] T.F. Cootes, S. Marsland, C.J. Twining, K. Smith, and C.J. Taylor. Groupwise diffeomorphic non-rigid registration for automatic model building. In *8th European Conference on Computer Vision*, volume 4, pages 316–327. Springer, 2004.
- [7] I. Dryden and K. V. Mardia. *The Statistical Analysis of Shape*. Wiley, London, 1998.
- [8] P.T. Fletcher, S. Joshi, C. Lu, and S.M. Pizer. Gaussian distributions on lie groups and their application to statistical shape analysis. In *18th Conference on Information Processing in Medical Imaging*, pages 450–462. Springer-Verlag, 2003.
- [9] K. Messer, J. Matas, J. Kittler, J. Luetin, and G. Maitre. XM2VTSdb: The extended m2vts database. In *Proc. 2nd Conf. on Audio and Video-based Biometric Personal Verification*. Springer Verlag, 1999.
- [10] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in C (2nd Edition)*. Cambridge University Press, 1992.
- [11] D. Rueckert, A.F. Frangi, and J.A. Schnabel. Automatic construction of 3D statistical deformation models using non-rigid registration. In *MICCAI*, pages 77–84, 2001.
- [12] S.M.Pizer, P.T. Fletcher, and S. Joshi ... Deformable m-reps for 3D medical image segmentation. *International Journal of Computer Vision*, 2-3(55):85–106, 2003.
- [13] C. Twining, S. Marsland, and C.J. Taylor. Measuring geodesic distances on the space of bounded diffeomorphisms. In P.L. Rosin and D. Marshall, editors, *13th British Machine Vision Conference*, volume 2, pages 847–856. BMVA Press, September 2002.