

Combining Point Distribution Models with Shape Models Based on Finite Element Analysis

T.F.Cootes and C.J.Taylor

Department of Medical Biophysics
University of Manchester
email: bim@wiau.mb.man.ac.uk

Abstract

This paper describes a method of combining two approaches to modelling flexible objects. Modal Analysis using Finite Element Methods (FEMs) generates a set of vibrational modes for a *single* shape. Point Distribution Models (PDMs) generate a statistical model of shape and shape variation from a *set* of example shapes. A new approach is described which generates vibrational modes when few example shapes are available and changes smoothly to using more statistical modes of variation when a large data set is presented. Results are given for both synthetic and real examples. Experiments using the models for image search show that the combined version performs better than either the PDM or FEM models alone.

1 Introduction

This paper addresses the problem of generating shape models which can represent image structure and be used to locate examples of such structure. Often the structures to be located can vary in shape, either because they are flexible or articulated, or because natural variation is present. The shape models must thus be flexible to allow for deformations of the structure. There have been many approaches to this problem (see below) but of particular interest are Finite Element Methods (Pentland and Sclaroff [1,2], Karaolani *et al.* [3], Nastar and Ayache [4]) and Point Distribution Models (Cootes *et al.* [6]) which have a number of complementary properties.

Finite Element Methods take a *single* instance of a shape and treat it as if it were made of a flexible material. The techniques of Modal Analysis give a set of linear deformations of the shape equivalent to the modes of vibration of the original shape. A flexible model consisting of the original shape and these modes can be used for image search. However, the modes are somewhat arbitrary and may not be representative of the real variations which occur in a class of shapes.

Point Distribution Models are statistical models generated from a *set* of example shapes. The mean shape and a number of modes of variation can be calculated from this set. Each mode describes one way in which the shapes in the training set tend to vary from the mean. By analysing sets of real examples a compact description of shape variation can be formed. Such models best represent the class of shapes of interest when a large number of examples is available. However, obtaining such a set can be difficult, and a PDM trained on a small set may not allow enough variation to adequately span the space of plausible shapes.

The aim of this work is to develop a technique which combines the properties of both of the above approaches. In particular the new technique must

- Give a set of 'vibrational' modes when presented with a single example shape,
- Give modes similar to a PDM when presented with a large number of shapes,

- Give modes which combine vibrational modes and statistical modes of variation when presented with a ‘moderate’ number of shapes.

Such a modelling technique would be useful when more than one example of a shape was available, but not enough to train a PDM properly. In particular, when one is ‘bootstrapping’ a PDM, using the existing model to locate examples in new images which are then to be incorporated into the model, such a system is essential.

This paper describes a technique which satisfies the above criteria. We concentrate on the formulation in 2-D, but the extension to 3-D is relatively straightforward. Examples of models, both synthetic and real are given and we demonstrate how the combined model can perform better during image search than either FEMs or PDMs alone.

The work we describe here combines two of the most promising approaches to modelling shape and shape variation. A review of other approaches is given in [6].

2 Finite Element Models

A number of authors have presented work concerning flexible shape models based on Finite Element Methods [3,4]. We will concentrate on the work of Pentland and Sclaroff [1,2] because they allow point-to-point correspondences to be made which can be useful when generating training sets for PDMs.

An elastic body can be represented as a set of n nodes, mass matrix \mathbf{M} and a stiffness matrix \mathbf{K} . In two dimensions these are both $2n \times 2n$. Modal Analysis allows calculation of a set of vibrational modes by solving the generalised eigenproblem

$$\mathbf{K}\Phi = \mathbf{M}\Phi\Omega^2 \quad (1)$$

where $\Phi = (\phi_1 | \phi_2 | \dots | \phi_n)$ is a $2n \times 2n$ matrix of eigenvectors representing the modes and $\Omega^2 = \text{diag}(\omega_1^2, \omega_2^2, \dots, \omega_{2n}^2)$ a diagonal matrix of eigenvalues associated with each eigenvector (ω_i is the frequency of the mode represented by ϕ_i).

If $\hat{\mathbf{x}} = (x_1, x_2, \dots, x_n, y_1, \dots, y_n)^T$ is a vector representing the original points, then a new shape can be generated using

$$\mathbf{x} = \hat{\mathbf{x}} + \Phi\mathbf{u} \quad (2)$$

where \mathbf{u} is a set of parameters.

The energy of deformation in the i^{th} mode is proportional to ω_i^2 .

For the experiments below we assume that the structure can be modelled as a set of unit masses mutually interconnected by springs with constant stiffness and rest length equal to the distance between the points. In this case \mathbf{M} is the identity and the stiffness matrix \mathbf{K} is easily calculated. The modes are then the eigenvectors of the symmetric matrix \mathbf{K} .

3 Point Distribution Models

Point Distribution Models are statistical models of shape and shape variation generated from sets of examples. Each example consists of a set of n labelled points.

Suppose we have m examples $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{in}, y_{i1}, \dots, y_{in})^T$ ($i = 1..m$) which have been translated, rotated and scaled to overlap as much as possible (using a Procrustes technique, see [6]).

We calculate the mean shape,

$$\hat{\mathbf{x}} = \frac{1}{m} \sum_{i=1}^{i=m} \mathbf{x}_i \quad (3)$$

and the covariance about the mean

$$\mathbf{S}_m = \frac{1}{m} \sum_{i=1}^{i=m} (\mathbf{x}_i - \hat{\mathbf{x}})(\mathbf{x}_i - \hat{\mathbf{x}})^T \quad (4)$$

$$\mathbf{S}_m = \left(\frac{1}{m} \sum_{i=1}^{i=m} \mathbf{x}_i \mathbf{x}_i^T \right) - \hat{\mathbf{x}} \hat{\mathbf{x}}^T \quad (5)$$

The t eigenvectors \mathbf{p}_i of \mathbf{S}_m corresponding to the largest t eigenvalues λ_i give a set of basis vectors, or *modes of variation*, for a deformable model.

A new example can be calculated using

$$\mathbf{x} = \hat{\mathbf{x}} + \mathbf{P}\mathbf{b} \quad (6)$$

where $\mathbf{P} = (\mathbf{p}_1 | \mathbf{p}_2 | \dots | \mathbf{p}_t)$ and $\mathbf{b} = (b_1 \ b_2 \ \dots \ b_t)^T$ is a set of t shape parameters.

4 A New Combined Model

The Basic Idea

There is a clear analogy between (2) and (6); they are both linear models. This suggests there may be some way to combine the two approaches. If we have just one example shape, we cannot build a PDM and our only option is the FEM approach. If, however, we have two examples, we can build a PDM, but it would only have a single mode of variation, which would linearly interpolate between the two examples. It would have no 'sensible' way of modelling any other distortion. The Finite Element Methods can generate a whole set of modes which allow smooth deformation of a single shape, but it is not clear how one would take two shapes into account.

One approach to combining FEM and PDMs is as follows. We calculate the modes of vibration of both shapes, and use them to generate a large number of new examples by randomly selecting model parameters \mathbf{u} using some suitable distribution. We then train a PDM on this new large set of examples. The PDM would thus incorporate a mixture of the modes of vibration of each original example with the original mode of variation that interpolates between the two shapes.

Such a strategy would be applicable for any number of original example shapes. However, we would have to reduce the magnitude of the allowed vibrational modes as the number of examples increased to avoid incorporating spurious modes. For instance, suppose one was drawing examples from a set of rectangles which all had the same width but different heights. Given a single example one only has the vibrational modes, which will allow both width and height to vary. Given a second example there is some evidence to suggest that the width is fixed, and as more examples are added our confidence in a fixed width increases. However, unless we limit the magnitude of the vibrations allowed, each new example will have modes with varying width, and the final model will be inappropriate.

It would, of course, be time consuming and error-inducing to actually generate large numbers of vibrationally deformed examples. Fortunately the effect can be achieved with a little matrix algebra.

The Mathematics

Consider the Finite Element model based on the i^{th} example, \mathbf{x}_i .

$$\mathbf{x} = \mathbf{x}_i + \Phi_i \mathbf{u} \quad (7)$$

where $\Phi_i = (\phi_{i1} | \phi_{i2} | \dots | \phi_{is})$ is a $2n \times s$ matrix of s eigenvectors representing the modes of vibration about \mathbf{x}_i .

Suppose we were to generate a set of examples by selecting the values for \mathbf{u} from a distribution with mean $\mathbf{0}$ and covariance matrix \mathbf{S}_u . The distribution of $\Phi_i \mathbf{u}$ would have a covariance of

$$\Phi_i \mathbf{S}_u \Phi_i^T \quad (8)$$

about the original example \mathbf{x}_i , and the distribution of \mathbf{x} would have a covariance of

$$\mathbf{C}_i = \Phi_i \mathbf{S}_u \Phi_i^T + \mathbf{x}_i \mathbf{x}_i^T \quad (9)$$

If we allow the elements of \mathbf{u} to be independent and normally distributed about zero with a variance $\alpha \lambda_j$, (where α is a constant and λ_j is as given below) then

$$\mathbf{S}_u = \alpha \Lambda \quad (10)$$

where $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_s)$ is an $s \times s$ diagonal matrix.

The covariance of \mathbf{x} about the origin is

$$\mathbf{C}_i = \alpha \Phi_i \Lambda \Phi_i^T + \mathbf{x}_i \mathbf{x}_i^T \quad (11)$$

If the frequency associated with the j^{th} mode is ω_j then we will choose

$$\lambda_j = \omega_j^{-2} \quad (12)$$

This gives a distribution which has large variation in the low frequency, low deformation modes, and a small variation in the high frequency, high deformation modes¹.

One can justify the choice of λ_i by considering the strain energy required to deform the original shape \mathbf{x}_i into a new example \mathbf{x} . The contribution to the total from the j^{th} mode is

$$E_j = \frac{1}{2} u_j^2 \omega_j^2 \quad (13)$$

The form of equation (12) ensures that the energy is spread equally amongst all the modes.

The constant α controls the magnitude of the deformations, and is discussed below.

1. In simple shapes the first 3 modes can be translation and linearised rotation, and the first 3 eigenvalues can be close to zero. In this case these first modes could be ignored and the calculation could proceed with the remaining $s-3$. However, in practice with any moderately complex shape the near zero frequencies do not appear to arise, and all the modes can be used.

Combining Examples

To calculate the covariance about the origin of a set of examples drawn from distributions about several original shapes, we simply average the individual covariances :

$$S' = \frac{1}{m} \sum_{i=1}^m C_i = \frac{1}{m} \sum_{i=1}^m (\alpha \Phi_i \Lambda \Phi_i^T + \mathbf{x}_i \mathbf{x}_i^T) \quad (14)$$

$$S' = \frac{1}{m} \sum_{i=1}^m (\alpha \Phi_i \Lambda \Phi_i^T) + \frac{1}{m} \sum_{i=1}^m \mathbf{x}_i \mathbf{x}_i^T \quad (15)$$

substituting in from (5) gives

$$S' = \frac{1}{m} \sum_{i=1}^m (\alpha \Phi_i \Lambda \Phi_i^T) + S_m + \hat{\mathbf{x}} \hat{\mathbf{x}}^T \quad (16)$$

where $\hat{\mathbf{x}}$ is the mean of the original examples, and S_m is the covariance of those examples about the mean.

The covariance about the mean, $\hat{\mathbf{x}}$, is then

$$S = S' - \hat{\mathbf{x}} \hat{\mathbf{x}}^T = S_m + \alpha \left(\frac{1}{m} \sum_{i=1}^m \Phi_i \Lambda \Phi_i^T \right) \quad (17)$$

We can build a model by calculating the eigenvectors and eigenvalues of this matrix, and proceed as normal for the PDM.

When $\alpha = 0$ (ie the magnitude of the vibrations is set to zero) (17) gives $S = S_m$ and we get exactly the result one would obtain from a PDM.

When we allow non-zero vibrations of each training example ($\alpha > 0$) the eigenvectors of S will include the effects of the vibrations. One effect of this is to increase the number of non-zero eigenvectors when fewer than $2n$ examples are available. If $m < 2n$ examples are presented, S_m will have at most $m-1$ non-zero eigenvalues giving allowable modes of variation for a PDM. However the addition of the terms due to the vibrations to the S matrix means it can have many more non-zero eigenvalues, leading to more allowable modes.

Choice of Proportion of FEM Variation to Include (α)

As the number of examples increases we wish to rely more on the statistics of the real data and less on effects caused by the 'artificial' modes of vibration. To achieve this we must reduce α as the number of samples m increases.

The size of α can be considered as a measure of our uncertainty about the completeness of the model. Having seen a large number of examples we become more certain that our model can deal with all examples from a class of structures by linear combinations of those already seen, without recourse to other deformations. Thus in this case α will be small. Statistically the variance of the estimate of a parameter tends to decrease as m^{-1} . A possible relationship between α and the number of training examples is then

$$\alpha = a_1/m \quad (18)$$

An alternative approach is to use ‘bootstrap’ or ‘jackknife’ techniques [5] to estimate the error in our estimate of the variation in the training set. We could then choose α to allow for this uncertainty.

5 Examples of the Method

Simple Synthetic Example

Two sets of 16 points were generated, one forming a square, the other a rectangle with aspect ratio 0.5. Figure 1 shows the modes corresponding to the four smallest eigenvectors of the FEM governing equation for the square – these are the modes of vibration of the shape, Φ_1 . Figure 2 shows the first 4 modes of vibration of the rectangle.

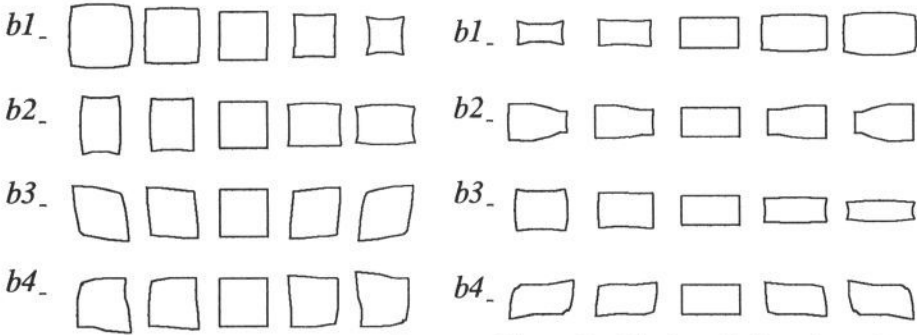


Figure 1 : Modes of vibration of a square

Figure 2 : Modes of vibration of a rectangle

Figure 3 shows the modes of variation generated from the eigenvectors of the combined covariance matrix (given by Eq. 17). This demonstrates that the principal mode is now the mode which changes the aspect ratio (which would be the one and only mode of a PDM trained on the two shapes).

A Real Example : A Face

Figures 4 and 5 show the first three modes of vibration for two examples of faces, one with a neutral expression, the other smiling. 169 points are used to represent the shape. In each case the variation caused by the modes is exaggerated to highlight the form of the deformation – in practice one would apply tighter limits to the amplitude of each mode. Figure 6 shows the modes of the model trained using both examples. The first mode describes the main variations between the two examples (the smiling) and subsequent modes combine variation from the two sets of FEM modes.

6 Comparing Performance of Models for Locating Image Structures

The flexible models generated by the approach described above are identical in form to PDMs, and so can be used for image search in both the Active Shape Model (ASM) framework [6–7] and the Genetic Algorithm approach described by Hill *et al* [8]. Here we compare PDMs, FEMs and the combined models when used for local optimisation in an ASM.

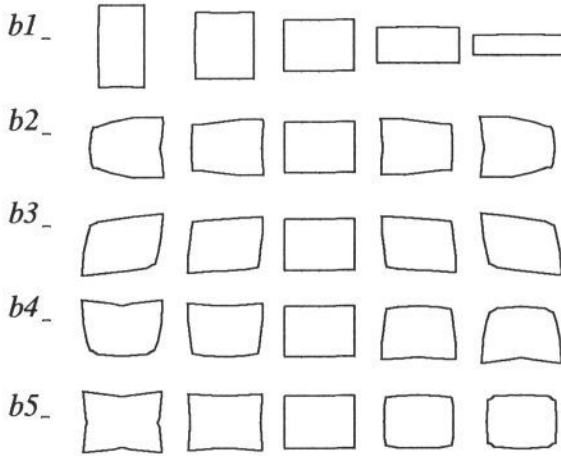


Figure 3 : Modes of Variation of combined model trained using square and rectangle. Notice that the mode changing the aspect ratio has been promoted to the most significant.

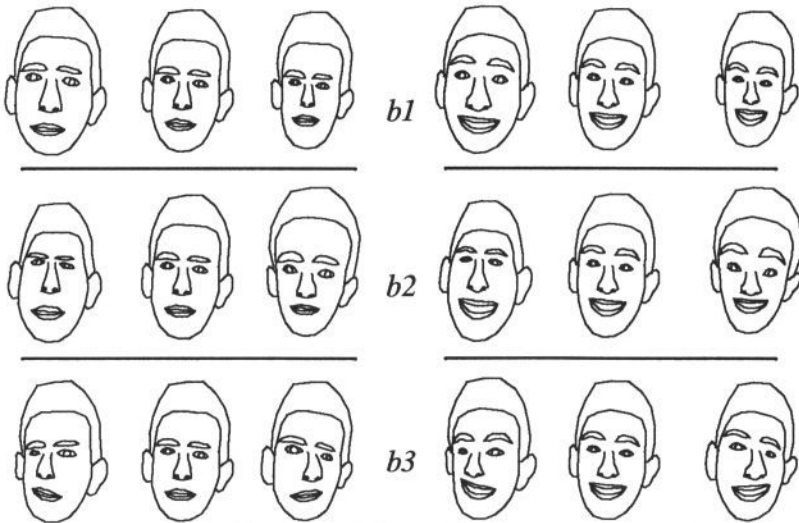


Figure 4 : FEM modes for neutral face Figure 5 : FEM modes for smiling face

For an ASM we require a shape model describing the locations of landmark points, and a set of models of the grey-levels in regions around each point. Starting from an initial rough estimate the parameters of a model instance are iteratively refined so that each model points moves towards a nearby location suggested by its grey-level model. This allows fast and robust location of structures in unseen images. In the experiments below we use a multi-resolution implementation [7]. Early iterations are performed on a low resolution image; as the search progresses the model automatically moves to higher and higher resolutions until convergence is reached.

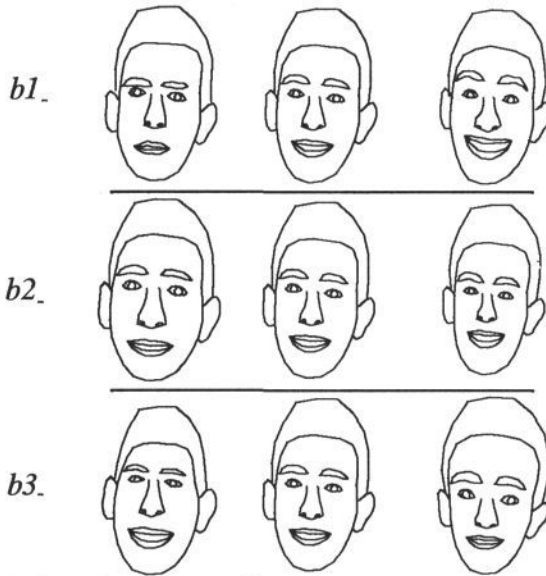


Figure 6 : Modes of variation of model trained on above two face examples

A set of 80 labelled images of faces (8 ‘mugshots’ of 10 different people) was split into a training set and a test set, with 4 images of each person in each set². Shape models were trained by selecting sets of examples from the training set at random. A set of grey-level models was generated by using all 40 examples in the full training set. The performance of each shape model was tested by running an ASM on every test image and measuring the average distance of the model points from the known points originally labelled by hand. Each model was started with a scale of 70% of that of the known face, and offset from the known centre by $(-10, -10)$ (the faces have a width of about 300 pixels). During the search the shape variation was constrained by applying a limit of $\pm 2.5\sqrt{\lambda_i}$ to each shape parameter, b_i , equivalent to 2.5 standard deviations from the mean.

Figure 7 shows average displacement plotted against ma (the number of training examples \times the amount of FEM variation included in the covariance matrix). When $a = 0$ we have a pure PDM model. The figure shows that as a is increased the performance first improves, then worsens. The best performance for a given number of examples, m , occurs when $ma \approx 0.5$. This suggests that equation 18 is appropriate for choosing a , and that for the face example $a_1 = 0.5$. The difference between the fit at $a = 0$ and the minimum value gives the potential improvement given by adding extra FEM variation to the model. The magnitude of this improvement drops as the number of training examples increases.

Figure 8 compares the average performance of different models training with various numbers of examples. The value of a for the combined model was calculated

2. The images are a subset of a database of face images collected and labelled by Andreas Lanitis. This is to be made publicly available – contact Andreas on email lan@wiau.mb.man.ac.uk

using equation 18 with $\alpha_1 = 0.5$. This clearly demonstrates that the combined models give superior performance when trained with fewer than 16 examples. When 16 or more examples are available, the PDM is sufficiently well representative of the full class of shapes that adding extra variability gives no further improvement.

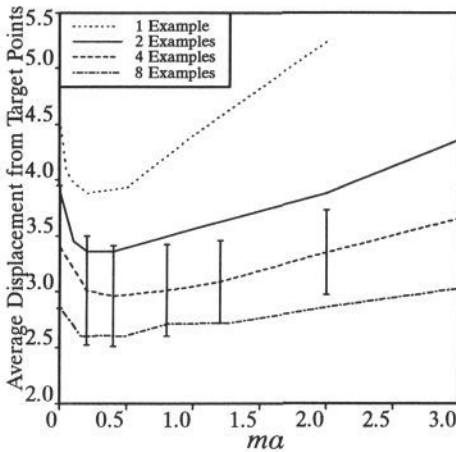


Figure 7 : Relationship between average fit, the number of examples used to train the model, m , and the proportion of FEM variation included, α .

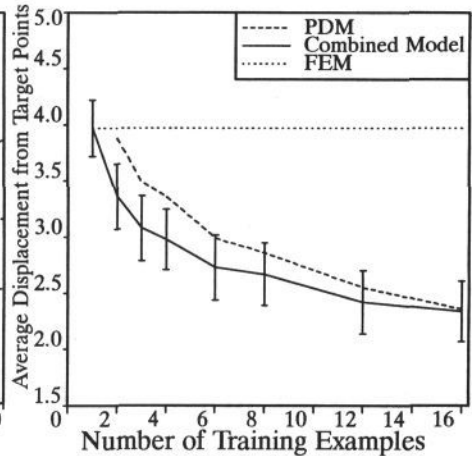


Figure 8 : Relationship between performance of ASMs and the number of training examples for different shape models.

7 Discussion and Conclusions

Related Approaches to Combining PDMs and FEMs

Martin, Pentland & Kikinis [9] have combined FEM and PDM models in order to classify normal and pathological shape deformations of brain structures in 3D MR scans. Their model of variability is to add the two types of modes together :

$$\mathbf{x} = \hat{\mathbf{x}} + \Phi \mathbf{u} + \mathbf{Pb} \quad (19)$$

They use the shape of the head to calculate the physical mode parameters (\mathbf{u}) then use the deformation so defined to warp the structure of interest toward the mean. The \mathbf{b} parameters are then simply calculated from any residual differences from the mean, and can be used to discriminate between various pathologies. Equation 19 is not satisfactory for our purposes because the two sets of modes are not independent – there will be redundancy in the model. We wish to find the best single model and use it for a one stage image search.

The New Combined Model

The approach described above allows us to combine the shape variability determined statistically from a training set with that generated artificially by building a ‘physical’ model of the objects of interest. When only a single example is available only FEM modes are used. As more examples are added, the statistical description of the shape variation becomes more complete, and less additional FEM variation is required. In the limit the statistical PDM modes are sufficient to span all the variation occurring in a class of shapes. Although the formulation is for FEM based models, any

method which generates linear modes of variation from a single example could be used.

The calculation of the FEM modes for every example in the training set can prove time consuming. It may be sufficient just to calculate the modes of the mean of the set, and use this to augment the PDM covariance matrix.

The combined model requires a single parameter, α , describing the amount of FEM based variation to add to the covariance matrix. The experiments above suggest it is reasonable to assume that the optimal value for the parameter is inversely proportional to the number of training examples used. The proportionality constant will be a function of the class of shapes modelled. Investigations are underway to determine guide-lines for choosing this constant given a single example.

The experiments described in Section 6 show that for the face example a PDM trained on 2 or more examples performs better in ASM search than a pure FEM based model. However, when only small numbers (< 16) of training examples are used the model combining both PDM and FEM variation performs better than either alone.

Acknowledgements

Tim Cootes is funded by an SERC Postdoctoral Fellowship. The work was developed from ideas generated during discussions with Jeff Hunter. This paper could not have been written without the support of the members of the Wolfson Image Analysis Unit. The face images were collected and labelled by Andreas Lanitis. The software was written using the RADIAL-2.4 C++ Image Processing Library.

References

- [1] A. Pentland and S. Sclaroff, Closed-Form Solutions for Physically Based Modelling and Recognition, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **13**, 1991, 715-729.
- [2] S.Sclaroff and A.Pentland, Modal Matching For Correspondence and Recognition, MIT Media Lab. Perceptual Computing Section Technical Report No. 201, May 1993.
- [3] P. Karaolani, G.D. Sullivan, K.D. Baker and M.J. Baines, A Finite Element Method for Deformable Models. *Proceedings of the Fifth Alvey Vision Conference, Reading, 1989*, pp. 73-78.
- [4] C. Nastar and N. Ayache, Non-Rigid Motion Analysis in Medical Images : a Physically Based Approach. in *Proceedings of IPMI '93*, (H.H.Barrett, A.F.Gmitro. Ed.s) Springer-Verlag, Berlin 1993, pp.17-32.
- [5] B.Efron and R. Tibshirani, Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statistical Science*, Vol.1, No.1, 1986, pp54-77.
- [6] T.F.Cootes, A.Hill, C.J.Taylor, J.Haslam, The Use of Active Shape Models for Locating Structures in Medical Images. *Image & Vision Computing Vol.12, No.6 July 1994*, 355-366.
- [7] T.F.Cootes, C.J.Taylor, A.Lanitis, Active Shape Models: Evaluation of a Multi-Resolution Approach to Improving Image Search. *BMVC'94*
- [8] A. Hill, T.F. Cootes, C.J. Taylor and K. Lindley. Medical Image Interpretation : A Generic Approach using Deformable Templates. *Medical Informatics (To Appear)*.
- [9] J.Martin, A.Pentland and R.Kikinis, Shape Analysis of Brain Structures Using Physical and Experimental Modes. *M.I.T. Media Lab. Perceptual Computing Technical Report No. 276, January 1994*.