

An Automatic Face Identification System Using Flexible Appearance Models

A. Lanitis, C.J.Taylor and T.F.Cootes

Dpt. of Medical Biophysics, University of Manchester

email: lan@wiau.mb.man.ac.uk

We describe the use of flexible models for representing the shape and grey-level appearance of human faces. These models are controlled by a small number of parameters which can be used to code the overall appearance of a face for image compression and classification purposes. The model parameters control both inter-class and within-class variation. Discriminant analysis techniques are employed to enhance the effect of those parameters affecting inter-class variation, which are useful for classification. We have performed experiments on face coding and reconstruction and automatic face identification. Good recognition rates are obtained even when significant variation in lighting, expression and 3D viewpoint, is allowed.

1 Introduction

Human faces display significant variation in appearance due to changes in expression, 3D orientation, lighting conditions, hairstyles and so on. A successful automatic face identification system should be capable of suppressing the effect of these factors allowing any face image to be rendered expression-free with standardised 3D orientation and lighting. We describe how the variations in shape and grey-level appearance in face images can be modelled, and present results for a fully automatic face identification system which tolerates changes in expression, viewpoint and lighting.

We have previously described a shape classification system based on *multiple* shape models [6], one for each class in a database. In this paper we describe the use of a *single* shape model for face classification. The flexible shape model employed is a Point Distribution Model(PDM) [3] which serves two functions: the automatic location of facial characteristics and the classification of faces based on shape. This model is used in conjunction with one or more flexible grey-level models so that each face can be described by a set of shape model parameters and a set of grey-level model parameters, resulting in an efficient representation for coding and identification.

For our experiments a database containing face images from 30 individuals was used. The details of the data set are shown in table 1. In addition to the normal test set, a second test set containing 3 face images per person was used. For this set, subjects were asked to disguise themselves by hiding part of their face, to provide a rigorous test of the robustness of the identification system. Typical images from the data sets are shown in figure 1.

		Conditions	Training Images	Test Images
Number of subjects(classes)	30			
Training images per subject	10			
Test images per subject	10	Lighting conditions	fixed	variable
Male subjects	23	3D movements	yes	yes
Female subjects	7	Expression	variable	variable
Ethnic origin	mixed	Distance from camera	variable	variable
Minimum age of subjects	17	Spectacles	no	yes
Maximum age of subjects	45	Beards/Moustaches	yes	yes
Time period between capturing training/test images	3 - 54 weeks	Hairstyle changes	no	yes
		Background	fixed	variable

Table 1: Details of the face data set used for our experiments

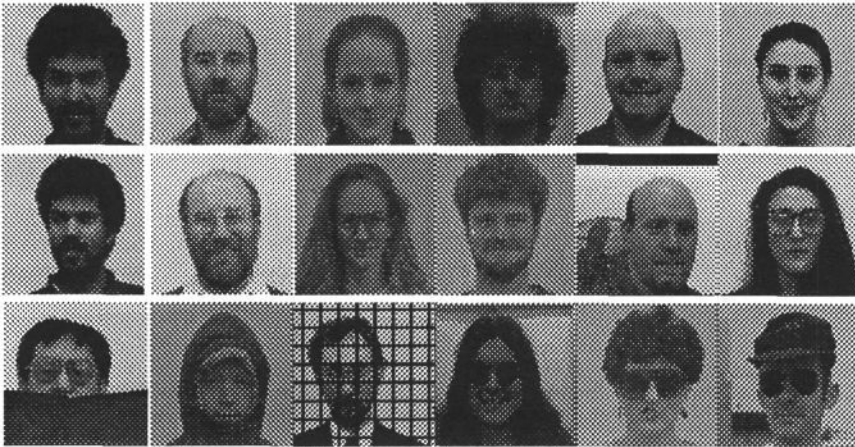


Fig. 1: Examples of images used in our experiments. Training images, test images and difficult test images are shown in top, middle and bottom rows respectively.

2 Background

Face identification techniques can be divided into two main categories: those employing geometrical features and those using grey-level information. Techniques based on geometrical features [2 , 5] use a number of dimensional measurements, or the locations of a number of control points. Based on these features, faces are identified. However, geometrical features are expression and 3D orientation dependent, so that explicit methods of normalisation must be employed. In most cases this involves the normalization of geometrical distances with respect to distances invariant to expression, such as the distance between the eyes. Craw [5] describes an attempt to minimise the effects of position and scale, by using least squares minimisation of the Euclidean distances between 5 control points located on a test image, and the corresponding control points located on the average face shape. This approach was only tested on images which did not display significant variation in 3D orientation.

Pentland [9] describes how principal component analysis of the grey-levels of face images can be used to create a set of eigenfaces. Any face can be approximated for coding and identification purposes by a weighted sum of eigenfaces. During the eigenvalue decomposition, no shape normalisation takes place and for the identification system no shape information is employed. Up to 96% correct classification was reported when this approach was tested on a database containing images from 16 different individuals. Craw [5] describes a similar method, with the difference that for his experiments the shapes of faces were normalised in order to ensure that only grey-level variations were modelled. Faces were deformed to the mean shape and principal component analysis applied to obtain shape-free eigenfaces. Using this approach test faces were retrieved correctly from a data base containing 100 images. These results relied on a user locating 59 key points on the test images interactively. When a similar experiment was performed using shape information alone, the results were not as good as those obtained with eigenfaces.

3 Overview

Our approach consists of two main phases: modelling in which flexible models of facial appearance are generated, and identification in which these models are used for classifying images. Figure 2 shows the main elements of our approach and indicates information flow within the system.

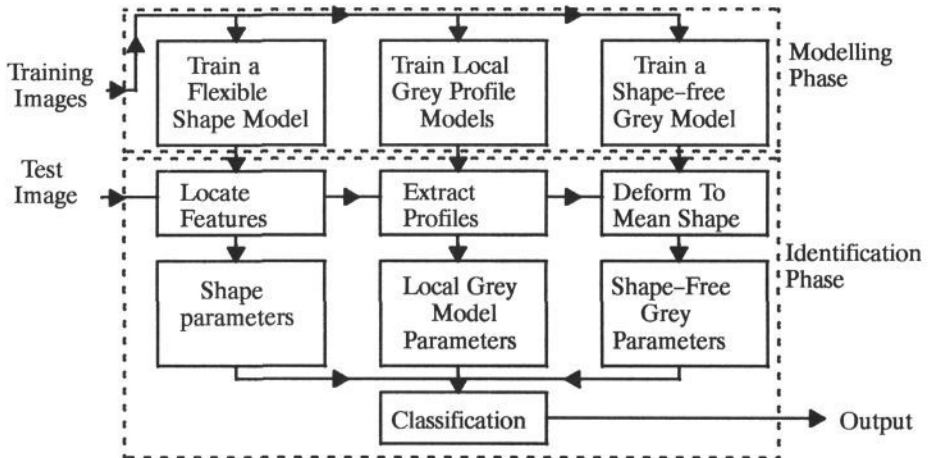


Figure 2: Block diagram of the face identification system

3.1 Modelling

We model the shape variation of human faces using a single PDM [3]. To train this model a number of training examples from each subject is used, so the shape variation reflects both *within-class* and *inter-class* variation. Within-class variation is the variation observed in face images of a particular individual, like changes in expression, 3D orientation and lighting conditions. Inter-class variation is responsible for the differences between human faces that make individuals distinctive, such as the absolute distances between features. Ideally we wish to use just inter-class information for classification so that the effects of expression and 3D orientation are minimised. We isolate inter-class variation using discriminant analysis techniques.

As previous authors have shown, grey-level information is extremely important in face recognition [2 , 5]. We have augmented our shape models with grey-level information using two different approaches. In the first we build a flexible model of the appearance of shape-free faces, obtained by deforming each face in the training set to have the same shape as the mean face. In the second approach we use a large number of local profile models one at each point of the shape model, rather than a single model. Each model point always corresponds to the same facial feature so there is no need to apply a deformation algorithm. If part of a face is occluded grey-level profile models for the points in that area are excluded from the classification procedure, allowing more robust identification. Shape and grey-level models can be used together to describe the overall appearance of each face in terms of its model parameters; collectively we refer to the model parameters as *appearance parameters*.

3.2 Identification

When a test image is presented to our system, facial characteristics are located automatically using an Active Shape Model(ASM) [4] based on the PDM obtained during training. The resulting model points are transformed into shape model parameters. Grey-level information at each model point is collected and transformed to local grey-level model parameters. Then the face is deformed to the mean shape and the grey-level appearance is transformed into the parameters of the shape-free grey-level model. Based on the set of appearance parameters and their known distribution for each class, the new face can be assigned to one of the classes. During the classification procedure discriminant analysis techniques are employed in order to ensure that appearance parameters responsible for within-class variation, are suppressed in favour of those controlling inter-class variation.

4 Flexible Models

All the models used in the system are of the same mathematical form. A flexible model [3] is generated from a set of training examples. Each training example(X_i) is represented by a number of variables.

$$X_i = (x_1, x_2, \dots, x_N) \quad (1)$$

Where x_k is the k th variable.

For instance, in constructing a shape model the x_i represent the co-ordinates of landmark points, expressed in a standard frame of reference. The average example(\bar{X}) is calculated and the deviation of each example from the mean is established. A principal component analysis of the covariance matrix of deviations reveals the main modes of variation. Any training example can be approximated using equation 2 .

$$X_i = \bar{X} + P\mathbf{b} \quad (2)$$

Where P is a matrix of eigenvectors of the covariance of deviations and \mathbf{b} is a vector of eigenvector weights (these are referred as *Model Parameters*)

Equation 3 can be used to transform training examples to model parameters.

$$\mathbf{b} = P^T(X_i - \bar{X}) \quad (3)$$

Usually variables vary in correlated ways so that the number of model parameters needed to describe the training set adequately is much less than the original number of variables. The same method can be used to train both shape and grey-level models. For shape models (PDM's) the variables are point co-ordinates and for grey-level models the variables are based on grey-level intensities.

5 Building a Face Shape Model

We have built a PDM representing human faces using 152 points and 160 training examples (8 examples from 20 individuals in the data set). Typical training examples, the mean shape, and the locations of the model points are shown in figures 3,4 and 5 respectively. The model can describe the shape of any face in the training set using

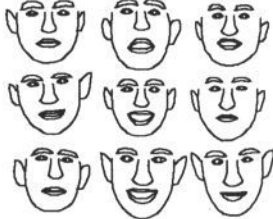


Fig 3: Typical training shapes Fig 4: The average shape Fig 5: Model points located

16 parameters; the effect of the first 4 is shown in figure 6. These parameters are used as shape classification variables. However, figure 6 shows that some of the most significant modes of variation account only for within-class variation. For example the first 3 modes just change the 3D orientation of the model. A way of isolating parameters responsible for inter-class variation must be employed so that classification based on shape parameters is expression and 3D orientation independent.

6 Separating Inter-Class and Within-Class Variation

Discriminant analysis [8] addresses the problem of separating a number of classes, given a number of variables for the training examples for each class. This is achieved by enhancing the effect of inter-class variation variables. In its simplest form discriminant analysis uses the Mahalanobis distance(D_i) measure, as shown in equation 4.

$$D_i^2 = (\mathbf{b} - \bar{\mathbf{b}}_i) C^{-1} (\mathbf{b} - \bar{\mathbf{b}}_i) \quad (4)$$

Where \mathbf{b} is a multivariate observation, $\bar{\mathbf{b}}_i$ is the centroid of a multivariate distribution

for a particular class and C is the common covariance matrix for all classes. A given multivariate observation is assigned to the class that minimises the Mahalanobis distance between the observation and the centroid of that class. Since Mahalanobis distance uses covariance information, it has the ability to suppress the effect of parameters responsible for within-class variation.

Another approach to the problem is to calculate Canonical Discriminant Functions [8]. These functions combine the existing variables linearly (see equation 5) in order to generate a new set of inter-class variation variables, the so called *discriminant parameters*.

$$\mathbf{d} = \mathbf{A}\mathbf{b} \quad (5)$$

Where \mathbf{d} is a vector of discriminant parameters and \mathbf{A} is a matrix containing a weight for each variable.

We have applied the second approach to the face model. Canonical discriminant functions were calculated for the shape parameters obtained for the set of training images. Only 6 discriminant variables were needed to explain most of the inter-class variation. To visualise the effect of inter-class variation on the model, all discriminant variables were set to zero and one was activated at a time. At each step the corresponding shape parameters were calculated using equation 5, and used to reconstruct the shape. Figure 7 shows the effect of the first 4 parameters. Unlike the raw shape parameters (see figure 6) the discriminant parameters deform the model only to generate shapes representing different individuals. No changes in 3D orientation or expression are activated by discriminant parameters, suggesting that if only inter-class variation parameters are utilised for classification, a system independent of expression and viewpoint can be developed.

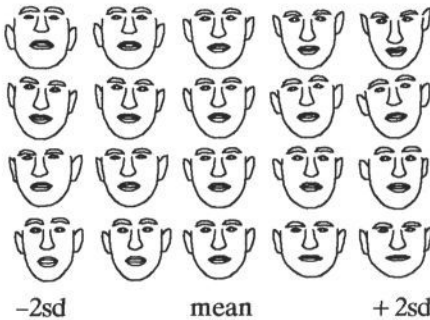


Fig 6: The effect of the main modes of shape variation

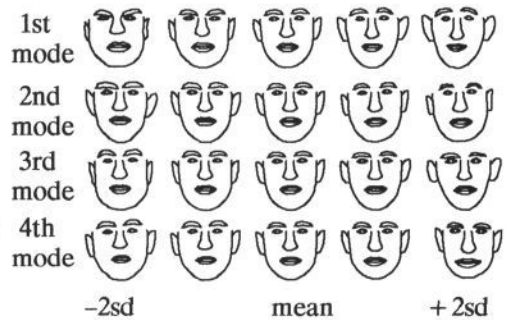


Fig 7: The effect of the main discriminant modes of shape variation

7 Building a Shape-Free Grey-Level Model

7.1 Deforming the Face Images

We wish to model grey-level appearance independently of shape. To do this we deform each face image to the mean shape in such a way that changes in grey-level intensities are kept to a minimum. For this purpose we have used a technique developed by Bookstein [1], based on *thin plate splines*. This allows an image to be deformed so that landmarks on the original image are moved to overlap a set of target landmarks, in such a way that changes in the the grey-level environment around each landmark are kept to minimum. Fourteen landmarks were used for deforming the face images. The position of these landmarks on a face outline and on the average face outline are shown in figure 8. All landmarks are part of the shape model, implying that once the shape model has been fitted all landmarks are located automatically. In figure 9 examples of face images before and after deformation are shown.

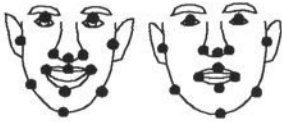


Fig 8: The landmarks used for deforming face images



Fig 9: Examples of original face images (left) and the respective shape-free face images (right)

7.2 Training The Flexible Model

We deform training images to the mean shape and extract grey-level intensities within the face area, as shown in figure 10. Each training example is represented by a vector containing the grey-level at each pixel in the patch (10656 pixels). The method described in section 4 was used to train a flexible grey-level model for our database. Only 12 variables were needed to explain 95% of the variation. Each variable is responsible for a specific mode of variation, as shown in figure 11. For example, there are modes that control the lighting conditions (1st mode), the addition of beards and moustaches (2nd and 4th modes) and the change in expression (3rd mode). Both inter-class and within-class variation are modelled; discriminant analysis techniques can be used to separate them. A problem with this model is that the first mode of variation represents 80% of the grey-level variation on its own so that other possible modes of variation are swamped. A second model using normalised rather than absolute grey-levels was also trained. Grey-levels representing each example were normalised, with respect to the average intensity in the grey-level patch. This model needed 67 variables to explain 95% of the variability. For classification and reconstruction purposes the normalised grey-level model was used.



Fig 10: Extracted grey-level patches used for training the shape-free grey-level model

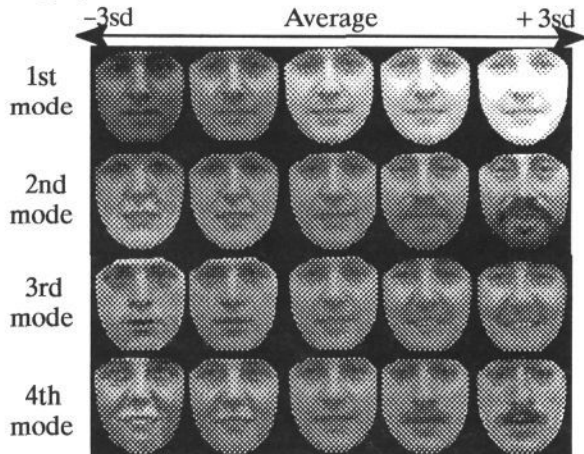


Fig 11: The main modes of grey-level variation

8 Coding, Reconstructing And Deforming Faces

Any face can be described in terms of its flexible shape and flexible grey-level model parameters. This representation provides an efficient coding scheme. For a given face image the shape is extracted and converted to shape parameters. Then the face is deformed to the mean shape and the shape-free grey-level appearance of the face is converted to grey-level model parameters. Using this method faces are coded using just 83 parameters (67 for the grey-level appearance and 16 for the shape). Given a set of shape and grey-level parameters, equation 2 can be employed to reconstruct

the grey-level appearance and the shape of a face. However, the reconstructed grey-levels correspond to the average shape face. The deformation algorithm described in section 7.1 can again be used, with the difference that now we deform *from* the mean shape *to* the shape corresponding to the shape parameters given. An example of automatic coding and reconstruction of a face image is shown in figure 12.

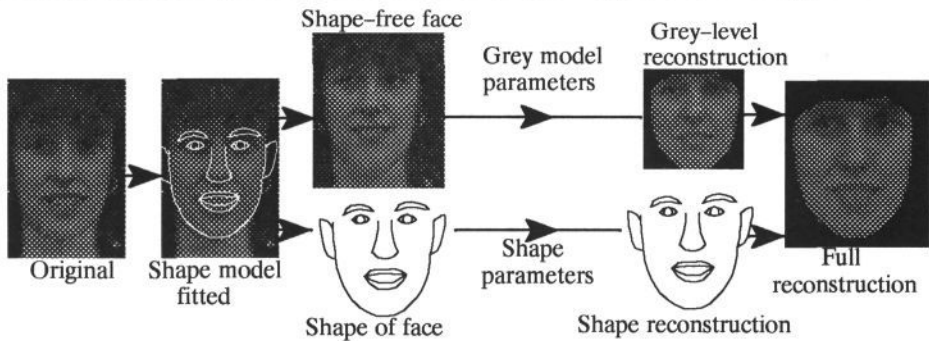


Fig 12: Coding and reconstruction using shape and grey-level models.

Model parameters can be modified to standardise the expression, 3D orientation and lighting conditions. This involves suppressing parameters responsible for within-class variation, using the canonical discriminant functions. The new reconstruction is a within-class variation-free face. A demonstration of this is given in figure 13, where the shape parameters were modified to produce a 3D orientation standardisation for the given face.



Fig. 13: Producing a within-class shape variation-free face image (a) original image (b) shape model fitted (c) within-class variation-free shape overlaid (d) deformed image (e) the resulting within-class variation-free face

9 Building Local Grey-Level Models

Local grey-level profile models can be used to model the grey-level appearance of faces, in the vicinity of each model point. During training, shape model points are overlaid on the corresponding training images and grey-level profiles in a direction perpendicular to the boundary, extracted, as shown in figure 14. The appearance of the extracted grey-level profiles is dependent on the scale of the face in the image. To account for this the length of training profiles was normalised with respect to the scale of the face in the image. Each profile was represented by 10 grey values, with the sampling interval varied to achieve length normalization (see figure 14). To minimise the effects of variable lighting conditions normalised derivatives of the profile data were used rather than absolute grey-levels. A flexible model was built for the profile at each model point, using the method described in section 4. Most of these models needed 4 model parameters to explain 95% of the variation.

By using multiple local models occlusions on face images can be detected. Once grey profiles are extracted, the corresponding model parameters are calculated and the

profiles reconstructed. If at a particular point the residual between the original and the reconstruction is greater than a threshold that point is considered as occluded and the grey-level information at that point is excluded from the classification procedure.

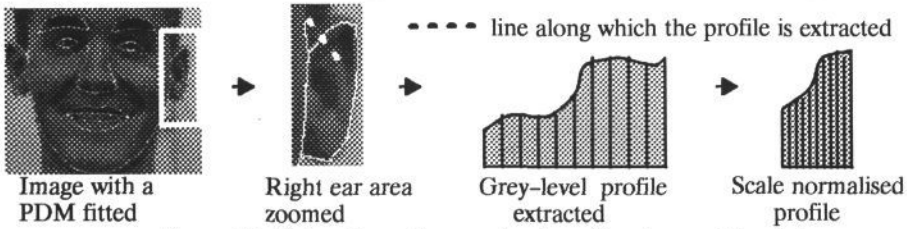


Figure 14: Extraction of a grey-level profile at a model point

10 Using Flexible Models For Identification

Automatic identification of faces can be performed using the flexible shape and grey-level models described above. When a new face image is presented to the system the flexible shape model is used for locating the facial characteristics automatically, using a multi resolution Active Shape Model (ASM) search [4]. In our current system the user indicates the approximate position of the nose and the shape model is overlaid on the image; deformation and Euclidean transformations of the model are applied in an iterative scheme, until the model is fitted to the shape of the face presented. Examples of the original placement and final fitting of the model are shown in figure 15. It should be noted that it is possible to locate facial features by searching the whole



Figure 15: Examples of shape mode fitting

image using global optimization techniques [7] rather than having to indicate the approximate location of the face. Once the model is fitted the resulting model points are expressed in a standard frame of reference and converted to model parameters using equation 3. (In this case X_i and \bar{X} contain the co-ordinates of the shape model points and of the average shape model, respectively). The resulting model parameters (b) are used for shape classification.

At each located shape model point, grey-level information from a profile perpendicular to the boundary is extracted. The length of these profiles is normalised in a manner similar to that applied during training (see section 9). The profile data for each point is converted to model parameters. Detection of occluded points follows; only model parameters for non-occluded points are used to perform classification based on the local grey-level models.

Based on the ASM fit, the landmarks used for deforming face images (see figure 8) are located. Target landmarks lie on the average shape; it is trivial to overlay the average face shape, and locate the target landmarks for the given image. These landmarks are used for deforming the face image to the mean face and shape-free grey-level model parameters are calculated; these parameters are also used for classification.

Face classification is achieved by using the set of appearance parameters calculated for the image presented. The multivariate distribution of these parameters, for each class is established during training, and a new image is assigned to the class that minimises the Mahalanobis distance between the centroid of that class and the calcu-

lated appearance parameters. Classification can be performed using the 3 types of model parameter individually, or in combination.

11 Introducing New Individuals To The System

To introduce a new individual to our identification system a small number of training images are required. Model points on training images are located by fitting the shape model automatically. Once model points are located the whole set of appearance parameters can be calculated. Based on the appearance parameters for all training images an approximation to their distribution for the new individual is established. It is essential that the location of model points is accurate because the derivation of all appearance parameters is critically depended on them. To ensure the accurate location of model points shape model fitting on training images is supervised so that appearance parameters based on inaccurately located points are not used.

12 Results

Initially classification experiments using images from 20 individuals from our database were performed; the results are shown in table 2. All experiments were performed using a Sparc 10 workstation. The two methods involving grey-level information achieve much better results, than the method based on shape alone. These results are consistent with those reported by other researchers [2 , 5], suggesting that the grey-level appearance of faces is much more important than shape for identification purposes. However, shape has an important classification role since classification rates improve significantly when shape information is included. Using grey-level profile models proved to be the most insensitive to occlusions. Any of the combinations of the measures produced very good results. For real time applications the method combining shape and grey-level profile information proved to be the best, whereas if increased classification accuracy is needed the shape-free grey-level model parameters should be included. Automatic training for 10 more individuals was performed; classification results for all 30 individuals are shown in table 3.

13 Discussion

We have presented an automatic face identification system. Both shape and grey-level information are incorporated into the system in such a way that the effects of variations in expression, 3D orientation and lighting are minimised. By using both grey-level and shape information we ensure that the identification information available in faces is fully exploited. We have tested the system using a database in which

Method	Classification time	Normal test set (200 test images)		Difficult test set (60 images)	
		Correct classifications	Correct class within best 3	Correct classifications	Correct class within best 3
Shape model	4 sec	70.0%	82.0%	23.3%	46.7%
Shape-free grey model	30 sec	84.5%	93.0%	33.3%	48.3%
Local grey-level models	4 sec	84.0%	93.5%	40%	61.7%
Shape + Shape-free model	30 sec	94.0%	99.0%	33.3%	58.3%
Shape + Local models	4 sec	91.5%	97%	33.3%	58.3%
All 3 methods	30 sec	95.5%	99.0%	43.3%	71.6%

Table 2: Results for the identification experiments

		Normal test set (300 test images)		Difficult test set (90 images)	
Method	Classification time	Correct classifications	Correct class within best 3	Correct classifications	Correct class within best 3
All 3 methods	30 sec	92.0%	94.0%	48.0%	67.0%

Table 3: Results for the identification experiments on the complete database

test images display significant appearance variation; the results obtained are very promising. The number of correct classifications for the 'disguised' test set is not high, but it should be noted that this test set contains face images that even humans find very difficult to identify. The identification procedure is fully automatic; errors may be caused either by failure in locating landmarks accurately, or by failure of the classification algorithm. We do not distinguish between the two cases, since we believe that locating facial characteristics automatically is an important aspect of an integrated face recognition system. Ten more individuals were added automatically to the original system without seriously affecting the classification performance. The modelling techniques described in this paper can be used for other applications, such as face coding and reconstruction, and automatic interpretation of facial expressions and movements. In the future we plan to make comparative tests between our recognition system, and systems developed by other researchers. The database we used for our experiments is publicly available and has already been supplied to other researchers active in the field. Anybody wishing to experiment using our database should contact the authors.

14 Acknowledgments

Andreas Lanitis is funded by a University of Manchester research studentship and an ORS award. We would like to thank members of the Department of Medical Biophysics, especially Dr. A. Hill and Dr. A. Ratter for their help and advice. Discussions at the 13th Leeds Statistical Workshop proved to be very helpful in implementing the image deformation algorithm. Finally we thank all those who kindly volunteered to provide face images for our experiments.

15 References

- [1] F.L. Bookstein. Principal Warps: Thin-Plate Splines and the Decomposition of Deformations. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 11, no 6, pp 567-585, 1989.
- [2] R. Brunelli and T. Poggio. Face Recognition: Features versus Templates. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 15, no 10, pp 1042-1052, 1993.
- [3] T.F. Cootes, A. Hill, C.J. Taylor, J. Haslam. The Use of Active Shape Models For Locating Structures in Medical Images. Image Vision and Computing Vol 12, no 6, pp 355-366, 1994.
- [4] T.F. Cootes, C.J. Taylor and A. Lanitis. Active Shape Models: Evaluation of a Multi-Resolution Method For Improving Image Search. To appear in this volume, 1994.
- [5] I. Craw and P. Cameron. Face Recognition by Computer. Proc of British Machine Vision Conference 1992, pp 489-507, eds. David Hogg and Roger Boyle, Springer Verlag, 1992.
- [6] A. Lanitis, C.J. Taylor and T.F. Cootes. A Generic System For Classifying Variable Objects Using Flexible Template Matching. Proc. of British Machine Vision Conference 1993, vol 1, pp 329-338, ed. J. Illingworth, BMVA Press, 1993.
- [7] A. Lanitis, Automatic Location of Facial Characteristics using Genetic Algorithms. Wolfon Image Analysis Unit Internal Report, June 1993.
- [8] B.F.J. Manly. Multivariate Statistical Methods, a Primer. Chapman and Hall, 1986.
- [9] M. Turk and A. Pentland. Eigenfaces for Recognition. Journal of Cognitive Neuroscience, Vol 3, no 1, pp 71-86, 1991.