Markov fields for recognition derived from facial texture error

N.P.Costen, T.F.Cootes and C.J.Taylor
University of Manchester
Division of Imaging Science and Biomedical Engineering,
Stopford Building, Oxford Road,
Manchester M13 9PT, U.K.

Abstract

When attempting to code faces for modelling or recognition, estimates of dimensions are typically obtained from an ensemble. These tend to be significantly sub-optimal. Each face contains both predictable and non-predictable qualities; only the predictable aspects are useful for defining coding systems for other faces. Additional information, not coded via the ensemble, is still available. We show that this information can be extracted and described via random Markov fields, and that this can be used to distinguish between images. The distances between images are robust to parameter setting, and can be combined with those derived via ensemble-based techniques to enhance recognition.

1 Introduction

A number of observations concerning the processes required for efficient face recognition can be divined by the consideration of the human psychological literature. One such observation stems from investigations into the nature of the variation which causes faces to differ one another; specifically it appears to divide into two aspects : *general familiarity* information, which is predictable from other faces and *memorability* information, which is not predictable [1].

Memorability information reflects small, discrete, easily verbalised features, for example skin blemishes or warts. Such information has essentially infinite dimensionality and will exhibit fortuitous correlations between faces. Thus it can disproportionately reduce both the specificity and the generality of a set of codes. Within a Principal Components setting, familiarity weighs on the early, high variance eigenvectors, while memorability correlates with the later, low-variance eigenvectors [2]. This division can be located by analysing the nature of the information coded by the Principal Components to ensure that all of the dimensions are themselves acceptable as faces. The memorability information, which will now be categorised as 'error', the difference between the original face and the version of the face coded on the Principal Components, also has information which can be useful for recognition. Advantages for adding the error magnitude into an identity decision have been shown by others [7].

We show that it is possible to provide succinct characterisations of the errors in terms of one-dimensional random Markov fields. These allow a description of the patterns of

BMVC 2001 doi:10.5244/C.15.50

error-value as a raster scan-line is followed across the face image. These Markov fields allow modest but appreciable levels of face recognition when the errors of probe images are assessed as possible samples generated by the Markov fields. This performance is, within limits, robust to changes in the parameters associated with the models, and can be combined with the output of a PCA-based recognition system to improve overall performance.

An additional problem associated with coding errors is that it is still necessary to account for all the variance. In particular, intra-person variation must be both included within the PCA, and excluded from the dimensions on which identity differences are measured. This can be achieved by dividing a large ensemble which includes variation of a large range of types (typically identity, expression, pose and lighting) into subsets which vary predominately on one individual type of variation. We then adopt a recoding strategy, which allows the construction of optimal non-orthogonal sub-spaces, describing the various types of variation. A further PCA can then be performed on the means of normalised identity-codes to provide a final set of dimensions and the error concerned with this set of dimensions included in the Markov field.

2 Background

Facial coding requires the approximation of a manifold, or high dimensional surface, on which any face can be said to lie. This allows accurate coding, recognition and reproduction of previously unseen examples. A number of previous studies [3, 4, 5] have suggested that using a *shape-free* coding provides a ready means of doing this, at least the when the range of pose-angle is relatively small, perhaps $\pm 20^o$ [6]. Here, the correspondence problem between faces is first solved by finding a pre-selected set of distinctive points (corners of eyes or mouths, for example) which are present in all faces. This is typically performed by hand during training. Those pixels thus defined as being part of the face can be warped to a fixed shape by standard grey-level interpolation techniques, ensuring that the image-wise and face-wise coordinates of a given image are equivalent. If a rigid transformation to remove scale, location and orientation effects is performed on the point-locations, they can then be treated in the same way as the grey-levels, as again identical values for corresponding points on different faces will have the same meaning.

Although these operations will linearise the space, allowing interpolation between pairs of faces, they do not give an estimate of the dimensions. Thus, the acceptability as a face of an object cannot be measured; this reduces recognition[3]. In addition, redundancies between feature-point location and grey-level values cannot be described.

Both these problems can be addressed by Principal Components Analysis. Given a set of N vectors $\mathbf{q_i}$ (either the pixel grey-levels, or the feature-point locations) sampled from the images, the covariance matrix $\mathbf{C}$ of the images is calculated,

$$\mathbf{C} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{q_i} - \bar{\mathbf{q}})(\mathbf{q_i} - \bar{\mathbf{q}})^T, \tag{1}$$

and orthogonal unit eigenvectors $\mathbf{\Phi}$ and a vector of eigenvalues λ are extracted from $\mathbf{C}$. This allows an estimate of the dimensions and range of the face-space. The weights $\mathbf{w_i}$ of a face can then be found,

$$\mathbf{w_i} = \mathbf{\Phi}^T(\mathbf{q_i} - \bar{\mathbf{q}}) \tag{2}$$

and the projected version $\mathbf{q}_\mathbf{i}'$ of the face,

$$\mathbf{q}_\mathbf{i}' = \mathbf{\Phi}\mathbf{w}_\mathbf{i} + \bar{\mathbf{q}}. \tag{3}$$

Since the columns of the matrix $\mathbf{\Phi}$ are orthogonal (and typically ordered by declining magnitude of λ_j) the similarity between $\mathbf{q}_\mathbf{i}$ and the projected version, $\mathbf{q}_\mathbf{i}'$ can be controlled by truncating $\mathbf{\Phi}$, and with it $\mathbf{w}$.

Redundancies between shape and grey-levels are removed by performing separate PCAs upon the shape and grey-levels, before the weights of the ensemble are combined to form single vectors on which second PCA is performed [4]. This 'appearance model' allows the description of the face in terms of true, expected variation – the distortions needed to move from one to another [7]. However, it will potentially code the entire variation between the faces which form our ensemble, including both the general and specific variance. The followings studies concern the analysis of the errors, the difference between $\mathbf{q}_\mathbf{i}$ and the projected version, $\mathbf{q}_\mathbf{i}'$, once suitable truncations of the various $\mathbf{\Phi}$ have been calculated.

3 Appearance Model Construction

For testing purposes, an ensemble of 314 facial images was used. This comprised 218 different individuals (the image to individual mapping was known), and was sub-divided into groups varying on facial pose, expression and lighting. Males and females were present in approximately equal proportions, and the individuals were drawn from a range of ages and ethnic groups. All the images had a uniform set of 68 landmarks found manually. A triangulation was applied to the points, bilinear interpolation used to warp the images to a standard shape and size which would yield a fixed number of pixels, which can be varied at the experimenter's will.

The number of dimensions was reduced with regard to both the shape and the region parameter, as described in [10]; this produced a compact, optimally descriptive model. An example of the effects of coding a sample using this model, giving both the approximation and error term (scaling the zero-mean error to fill the full range of the image-greyscales) is shown in Figure 1. As can be seen, while the approximation is a reasonable representation, the error image contains a considerable amount of useful information, in particular the nose-ring.

4 Identity-space Calculation

The parameters derived from the appearance model will describe both inter- and intra-personal variation. It is possible to calculate a inter-personal sub-space from this [11], but any errors with respect to this space will predominately describe intra-personal variation, and so could be added to the Markov fields. Thus we use a recoding algorithm to take account of the multiple possible explanations of the coding of a given face and normalise them before submitting to the final dimensions. If $\mathbf{M}$ is the matrix formed by concatenating $\mathbf{\Phi}^{(\mathbf{j}=\mathbf{1},\mathbf{2}\ldots)}$ and $\mathbf{D}$ is the diagonal matrix of $\lambda^{(\mathbf{j}=\mathbf{1},\mathbf{2}\ldots)}$,

$$\mathbf{w} = (\mathbf{D}\mathbf{M}^T\mathbf{M} + \mathbf{I})^{-1}\mathbf{D}\mathbf{M}^T(\mathbf{q} - \bar{\mathbf{q}}) \tag{4}$$

Figure 1: An example of a face approximated by the appearance model; for the left the original image, the approximated version, and the difference between the two.

and this also gives a projected version of the face

$$\mathbf{q}' = (\mathbf{D}\mathbf{M}^T)^{-1}(\mathbf{D}\mathbf{M}^T\mathbf{M} + \mathbf{I})\mathbf{w} + \bar{\mathbf{q}} \tag{5}$$

with $w_l = 0$ for those subspaces not required in the new version.

The appearance-model weights were obtained (using Equation 2) for each image used to build the truncated model. Separate PCAs were then performed upon the sets of the weights. The covariance matrices for the identity and lighting subspaces were calculated using Equation 1 while the pose and expression subspaces used

$$\mathbf{C}_W = \frac{1}{n_o n_p} \sum_{i=1}^{n_p} \sum_{k=1}^{n_o} (\mathbf{q}_{ki} - \bar{\mathbf{q}}_i)(\mathbf{q}_{ki} - \bar{\mathbf{q}}_i)^T \tag{6}$$

where n_o is the number of observations per individual, n_p is the number of individuals, and $\bar{\mathbf{q}}_i$ is the mean of individual i. Although all the eigenvectors implied by the identity, lighting and expression sets were used, only the two most variable from the pose set were extracted.

The eigenvectors were combined to form $\mathbf{M}$ and Equations 4 and 5 used to give the projection $\mathbf{q}'_\mathbf{j}$ of face $\mathbf{q}$ for subspace j. This procedure looses useful variation. For example, the identity component of the expression and pose images was unlikely to be coded precisely by the identity set alone. Thus the full projection $\mathbf{q}'$ was calculated, and recoded image $\mathbf{r_j}$ included an apportioned error component:

$$\mathbf{r_j} = \mathbf{q}'_\mathbf{j} + \frac{(\mathbf{q}' - \mathbf{q}) \sum_{k=1}^{N_j} \lambda_k^{(j)}}{\sum_{j=1}^{n_s} \sum_{k=1}^{N_j} \lambda_k^{(j)}}. \tag{7}$$

This yielded four ensembles, each of 314 images. A further four PCAs were performed on the recoded ensembles (all using Equation 1), extracting the same number of components as on the previous PCA for the lighting, pose and expression subspaces, plus all the non-zero components for the identity sub-space. Combined, these formed a new estimate of $\mathbf{M}$, and Equations 4, 5 and 7 were applied to give a third-level estimate and so forth. Convergence was assessed by measuring the Mahalanobis distance between

the projections of the images the various spaces. The algorithm continued until successive iterations produced the same pattern of distances; in practice this was almost always achieved by the third iteration.

The identity-only codes of all the images were then obtained using Equations 4 and 5 and a final PCA applied to the between-person covariance matrix

$$\mathbf{C}_B = \frac{1}{n_p}\sum_{i=1}^{n_p}(\bar{\mathbf{q}}_i - \bar{\mathbf{q}})(\bar{\mathbf{q}}_i - \bar{\mathbf{q}})^T \tag{8}$$

so rotating the identity-dimensions to correct in differences in the number of images per person and emphasise between-person variation. This model had 184 dimensions.

5 Markov field description

When seeking to describe the errors in approximating faces, a major aim is to derive a reasonably compact description of the errors which can then be compared one with another. There are a number of methods which could be used. One could, for example, use the magnitude of the error [7], but this will fail to capture the particular pattern of errors present. Alternatively, one could attempt to locate points of particular salience in the image (using an algorithm such as is applied in [12]). This was attempted, but proved to be extremely sensitive to noise.

Thus it was decided to characterise the data by means of Markov random fields [9], constructing one per person in the gallery. These describe the frequency with which an observation which can be placed in one category is followed by some other category, and thus required that the error for the texture-region be described as a one-dimensional vector of approximately 6,000 elements. This was achieved by following the raster-scan used by the image-processing software. Following a two-dimensional path across the image might also be possible. Only the error of the texture was used, as truncation procedure should ensure that the error on the shape is truly noise.

Since the values at each pixel were effectively unbounded double-precision units, it was necessary to supply arbitrary categories into which to place the values. The negative effects of this were minimised by setting the category boundaries to equalise number of values per category across the images which make up the gallery. This is rather akin to using a Mahalanois distance to scale a set of axes to describe a distribution of data more effectively, and both helped ensure that all the transitions were possible, and also reduced disadvantage of using exposed Markov fields rather than full hidden models. It is also necessary to calculate a probability of starting the system off from a given state. This is usually taken directly from the distribution of states derived from the first item of the samples. However, as the whole test-image forms a single sample, the distributions over the whole of each sample were used. An example of the representation of a gallery image, using five Markov states is given in Tables 1 and 2.

The probability that given test sample (a probe-face) can be produced by a model (ie is the same person as a given face) was assessed using the Forward-Backward Procedure [9]. Thus we need to set two parameters: how many states the models have, and the minimum probability of state-transitions. This latter is needed because it is quite likely that a probe image will include a state that is not present in the gallery models. Since the Forward-Backward Procedure estimates the appropriateness of a sample to the model

State	Range	Initial probability
1	$i < -0.0420$	0.221
2	$-0.0422 \leq i < -0.0056$	0.189
3	$-0.0058 \leq i < -0.0020$	0.180
4	$-0.0020 \leq i < 0.0012$	0.169
5	$i \geq 0.0012$	0.230

Table 1: Pixel value ranges and initial probabilities for the states of a typical gallery image. i is the raw pixel grey-level.

by calculating the probability of passing through the sequence of states implied by the sample, any sample including a state which was not present in the training data would otherwise automatically be assigned a probability of zero.

There were two sources of error variation with which models could be constructed and tested. The first was the difference between the image as approximated by the appearance model and the original image (of course, working on the region vector only); the second was the difference between the complete normalised identity description as provided by the recoding algorithm, and its projection through the identity space. This latter error was in fact a set of parameters on the combined appearance model and so was projected through both the combined and region principal components before being added to the appearance model error. The values shown in Tables 1 and 2 derive from a image which combines both types of error, with the PCA identity-space limited to 80 dimensions, and minimum state-transition probability of $P = 0.001$.

	First state				
Second State	1	2	3	4	5
1	0.774	0.183	0.034	0.004	0.005
2	0.199	0.506	0.244	0.034	0.017
3	0.048	0.225	0.486	0.208	0.033
4	0.017	0.043	0.228	0.503	0.208
5	0.002	0.012	0.024	0.161	0.801

Table 2: State-transition probabilities for a typical gallery image.

6 Results

An effective method of performing face recognition should be relatively unaffected by the various parameters which it is necessary to set. To this end, recognition was tested on a set of 22 individuals, disjoint from the ensemble. Each person was present in seven different images, three of which were selected as gallery images, and four as probes. The images were different occasions, and showed significant variation in pose, lighting and expression. All the correspondences required were found by hand; since the interest here is the variation in performance under different coding schemes, automatic correspondence, such as those from [8] were not needed.

6.1 PCA-based recognition

The first test necessary is to determine the correct number of dimensions on the PCA-identity space to use. Although the space is derived from images which are both truncated and normalised, it is still probable that a number of dimensions code memorability information. If this is so, the fortuitous weighings on the later dimensions will reduce performance.

A pooled, within-person covariance matrix was derived from the gallery. This allowed

$$d^2_{i \to k} = (\mathbf{w_k} - \bar{\mathbf{w}}_\mathbf{i})^T \mathbf{C}_W^{-1} (\mathbf{w_k} - \bar{\mathbf{w}}_\mathbf{i}), \tag{9}$$

where $1 \leq k \leq (n_o \times n_p)$ to give Mahalanobis distances from the probes to the mean images of the gallery. A recognition was scored when the smallest d had the same identity for i and k. Figure 2 shows the effects of adding low-eigenvalue dimensions to $\mathbf{w}$ on the number of individuals correctly identified for both the identity-model as described above, and one built directly from the appearance-model parameters. For comparison, Figure 3 shows the effects of the same manipulations on the hit-rate (the number of individuals correctly recognised) when the false-alarm and miss-rates are identical and perfect performance would have a rate of 1. Clearly, the recoded, normalised dimensions overall best with about 80 dimensions, but the un-recoded dimensions show a larger degree of separation (but not a better ordering) with rather larger numbers of dimensions.

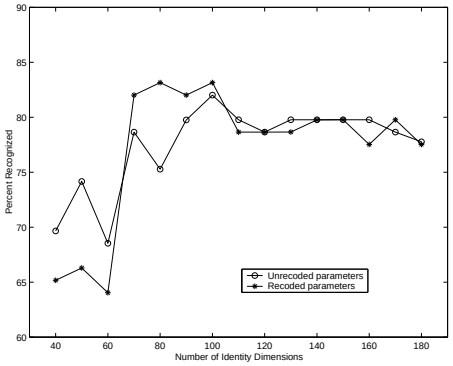

Figure 2: Effects on percentage recognised of varying the number of dimensions used on identity-only PCAs.

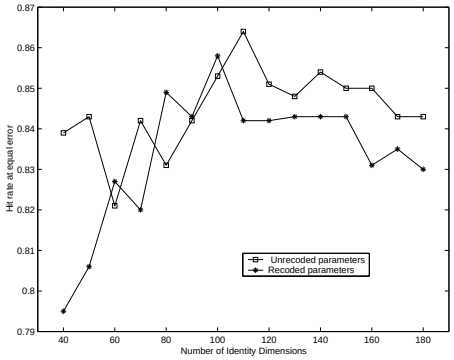

Figure 3: Effects on equal error hit rate of varying the number of dimensions used on identity-only PCAs.

6.2 Markov field recognition

6.2.1 Number of PCA dimensions

The Markov fields work on the complement of the PCAs; they are built from the identity-specific information which has not been included in the PCA-code. Thus it seems reasonable to investigate the effects of varying number of dimensions of recoded identity-PCA used to remove familiarity information. Figure 4 shows the effects on the hit rate at equal errors; clearly the performance is relatively unaffected by the number of dimensions (y-axis is highly magnified in this graph). The number of states used was fixed at 11, and

the minimum state-transition probability at $P = 0.001$. The percentage correct shows a similar pattern, with noticeable peak at 80 dimensions, where 30% of the images are correctly recognised. The error-images here are a mixture of the appearance-model region error and the identity-space error, which is being varied here. If only the appearance model error is used (so all 184 identity-dimensions are included in the model) the figures would be 0.55 and 18%. Clearly recognition is possible, but is significantly worse than identity-PCA, even when more information is present.

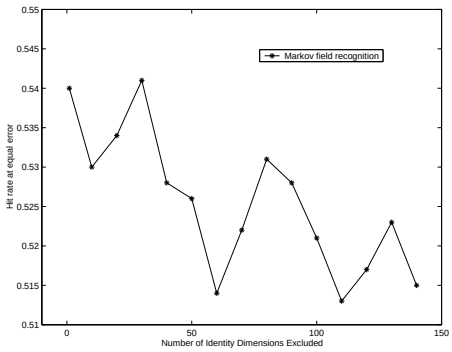

Figure 4: Effects on equal error hit rate via Markov fields of varying the number of dimensions used for identity-only PCAs.

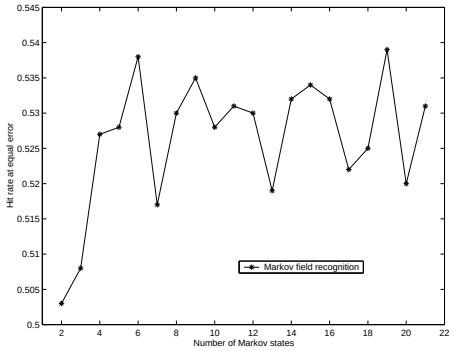

Figure 5: Effects on equal error hit rate of varying the number of states in the Markov fields.

6.2.2 Number of Markov field states

Figure 5 shows the effects of varying the number of states into which the errors are classified for hit rate at equal error . The images were coded on 80 identity-PCA dimensions, with a minimum transition probability of $P = 0.001$. Clearly, assuming there are more than approximately 4 states, the recognition rate is relatively stable.

6.2.3 Markov field state-transition minimum probability

The effects of varying the lower bound to the state-transition probabilities is shown in Figure 6; again the images were coded on 80 identity-PCA dimensions, and 11 state-models were used. Here, there appears to be a critical value of approximately $P = 0.008$; as long as the value is less than this, performance is relatively stable.

6.2.4 Combined PCA and Markov-field recognition

Since the PCA and Markov field identity parameters are based upon orthogonal variance, it should be possible to combine them into a single measure. Given the different ways of calculating similarity, the combination was carried out at the level of the distances. The two distances were normalised, so that the maximum distance from each probe image for both measures was unity. The hit rates for equal errors, with as a comparison, the PCA alone are shown in Figure 7. Clearly, the combined version has an advantage; this is

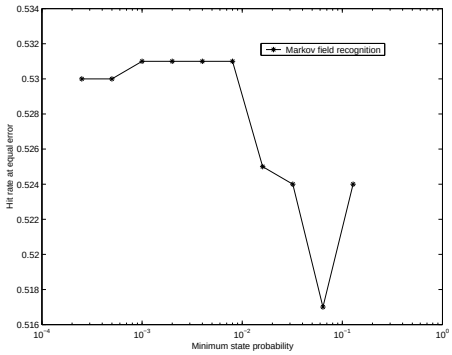

Figure 6: Effects on equal error hit rate of varying the minimum probability of state-transitions in Markov fields.

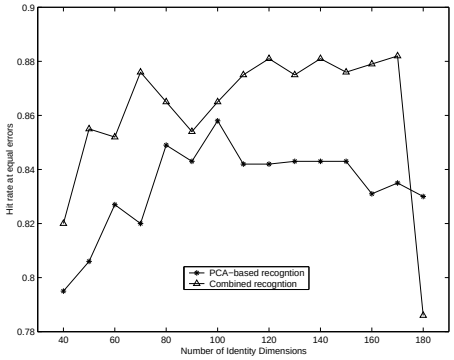

Figure 7: Effects on equal error hit rate of varying the number of PCA-identity dimensions in PCA and combined recognition.

especially true as the dimensional cut-off moves away from the 80-dimension maximum for PCA-based recognition. Percentage recognised follows a similar path.

7 Conclusions

Once faces have been accurately coded, a major problem is that only a sub-set of the codes should be used for manipulations or measurement. Although some of a given set of codes will respond to both generic variation, and so useful when considering faces not in the ensemble, the others will describe specific variation, which should not be used in this way. Nevertheless, this information will discriminate between faces, and should perhaps be used.

We have shown that the analog of this information which is present in the test images, described via a Markov field technique does contain noticeable levels of usable information; this is especially true when the representation error with regard to the overall description of the image (that from the appearance model) is supplemented by error information which is at once contained within the span of the appearance model, is not due to intra-personal variation, and still would have a negative effect on performance if it were included within a PCA-based identity measure. This is made possible by the 'recoding' algorithm, which supplies the lowest energy explanation of the various overlapping explanations of a particular facial configuration. The Markov field measures are relatively consistent when the two major parameters involved, the number of states and the minimum transition probability, are varied, ensuring these are not overly critical.

The PCA and Markov-based measures can be combined the improve performance, and again reduce the necessity of setting particular cut-offs on the number of dimensions in the PCA. It should also be noted that performance could further enhanced by using a dimension-generality corrected LDA [11], and that by making use of the errors, performance will be enhanced in situations where an automatic correspondence-finder will produce relatively inaccurate results.

References

[1] J. R. Vokey and J. D. Read Familiarity, memorability, and the effect of typicality on the recognition of faces. *Memory and Cogntion*,vol 20, pages 291–302, 1992.

[2] A. J. O'Toole, K. A. Deffenbacher, D. Valentin and H. Abdi Structural aspects of face recognition and the other race effect. *Memory and Cogntion*,vol 22, pages 208–224, 1994.

[3] N. P. Costen, I. G. Craw, G. J. Robertson, and S. Akamatsu. Automatic face recognition: What representation ? *European Conference on Computer Vision, Vol 1*, pages 504–513, 1996.

[4] G. J. Edwards, A. Lanitis, C. J. Taylor, and T. F. Cootes. Modelling the variability in face images. *2nd Face and Gesture*, pages 328–333, 1996.

[5] N. P. Costen, I. G. Craw, T. Kato, G. Robertson, and S. Akamatsu. Manifold caricatures: On the psychological consistency of computer face recognition. *2nd Face and Gesture*, pages 4–10, 1996.

[6] T. Poggio and D. Beymer. Learning networks for face analysis and synthesis. *Face and Gesture*, pages 160–165, 1995.

[7] B. Moghaddam, W. Wahid, and A. Pentland. Beyond eigenfaces: Probabilistic matching for face recognition. *3rd Face and Gesture*, pages 30–35, 1998.

[8] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active Appearance Models. *European Conference on Computer Vision, Vol 2*, pages 484–498, 1998.

[9] L. R. Rabiner. A Tutorial on Hidden Markov Models and Selected Applications in Speach Recognition. *Proc. IEEE*, vol 77, pages 257–285, 1988.

[10] N. P. Costen, T. F. Cootes and C. J. Taylor. Compensating for ensemble-specific effects when building facial models. *Proc. British Machine Vision Conference*, pages 62–71, 2000.

[11] C. Liu and H. Wechsler A Shape- and Texture-Based Enhanced Fisher Classifier for Face Recognition. *IEEE Trans. Image Processing*, vol 10, pages 598–608, 2001.

[12] K. N. Walker, T. F. Cootes and C. J. Taylor. Locating Salient Object Features. *Proc. British Machine Vision Conference*, pages 557-566, 1998.