Appears in the *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, **CVPR'99**, Fort Collins, Colorado, USA, June 23-25, 1999.

Face Recognition Using Shape and Texture

Chengjun Liu and Harry Wechsler
Department of Computer Science
George Mason University
Fairfax, VA 22030-4444
{cliu, wechsler}@cs.gmu.edu

Abstract

We introduce in this paper a new face coding and recognition method which employs the Enhanced FLD (Fisher Linear Discrimimant) Model (EFM) on integrated shape (vector) and texture ('shape-free' image) information. Shape encodes the feature geometry of a face while texture provides a normalized shape-free image by warping the original face image to the mean shape, i.e., the average of aligned shapes. The dimensionalities of the shape and the texture spaces are first reduced using Principal Component Analysis (PCA). The corresponding but reduced shape and texture features are then integrated through a normalization procedure to form augmented features. The dimensionality reduction procedure, constrained by EFM for enhanced generalization, maintains a proper balance between the spectral energy needs of PCA for adequate representation, and the FLD discrimination requirements, that the eigenvalues of the within-class scatter matrix should not include small trailing values after the dimensionality reduction procedure as they appear in the denominator.

1. Introduction

A successful face recognition methodology depends heavily on the particular choice of the features used by the (pattern) classifier [3], [15]. Feature selection in pattern recognition involves the derivation of salient features from the raw input data in order to reduce the amount of data used for classification and simultaneously provide enhanced discriminatory power. Recently shape and texture ('shape-free' image) have become prominent for encoding face images [4], [2], [18], [9]. Shape and texture coding, usually used in conjunction with norm based coding, is a two-stage process once the face has been located. Coding starts by annotating the face using important internal and face boundary points. Once these control points are located, they are aligned using translation, scaling and rotation transformations as necessary, and a corresponding mean shape is derived. The next stage then triangulates the annotated faces and warps each face to the mean shape. The first stage yields the shape, while the second stage yields the texture and corresponds to what is known as full anticaricature [4].

This paper introduces a new face coding and recognition method which employs the Enhanced FLD (Fisher Linear Discrimimant) Model (EFM) [10] on augmented shape and texture features. Experimental results, using 600 face images of varying illumination and corresponding to 200 subjects whose facial expression can vary, show that (i) the augmented shape and texture features carry the most discriminant information followed in order by textures, shape vectors, masked images, and shape images; (ii) our new face coding and recognition method performs the best among the eigenfaces method [17] using L_1 or L_2 distance measure, and the Mahalanobis distance classifiers using a common covariance matrix for all classes [9], [5] or a pooled within-class covariance matrix [7]. In particular, EFM achieves 98.5% recognition rate using only 25 features.

2. Background

Learning to recognize visual objects, such as human faces, requires the ability to find meaningful patterns in spaces of very high dimensionality [11]. Psychophisical findings indicate, however, that "perceptual tasks such as similarity judgment tend to be performed on a low-dimensional representation of the sensory data. Low dimensionality is especially important for learning, as the number of examples required for attaining a given level of performance grows exponentially with the dimensionality of the underlying representation space" [6]. Principal Component Analysis (PCA) [8] is the method behind the eigenfaces coding scheme [17] whose primary goal is to project the similarity judgment for face recognition in a low-dimensional space. Note, however, that PCA driven coding schemes are optimal and useful only with respect to data compression and decorrelation of low (2nd) order statistics. The recognition aspect is not considered and one should thus not expect optimal performance for tasks such as face recognition when using PCA-like coding schemes.

The Fisher Linear Discriminant (FLD) is a popular discriminant method for the very purpose of achieving high separability between the different patterns in whose classification one is interested. Characteristic of this approach are recent but similar methods such as the Most Discriminating Features (MDF) [16] and the Fisherfaces [1]. The combined use of PCA and FLD like methods is an improvement over PCA methods, but still has its own drawbacks, especially those associated with overfitting and lack of generalization as a result of insufficient training data [10]. To address the overfitting problem Liu and Wechsler [10] introduced Enhanced FLD Models (EFM) to improve on the generalization capability of the standard FLD based classifiers such as Fisherfaces.

To further improve PCA stand-alone methods, both new face representation methods and new classifiers are emerging. Beymer [2] introduced a vectorized image representation consisting of shape and texture. Vetter and Poggio [18] used such a vectorized face representation for image synthesis from a single example view. Craw, Costen and Kato [5] and Lanitis, Taylor and Cootes [9] developed a Mahalanobis distance classifier for face recognition using the same type of shape and texture augmented representation. The Mahalanobis distance is measured with respect to a common covariance matrix for all classes in order to treat variations along all axes as equally significant by weighting components corresponding to smaller eigenvalues more heavily [5]. Note that the weighting procedure does not differentiate the between-class scatter from the within-class scatter and suppresses the former while reducing the latter. To address this issue and to better distinguish the different roles of the two scatters, Edwards, Cootes and Taylor [7] presented yet another Mahalanobis distance classifier by using the pooled within-class covariance matrix.

3. Face Recognition Using Shape and Texture

We introduce now a new face coding and recognition method which employs the Enhanced FLD Model (EFM) on integrated shape and texture information. Shape encodes the feature geometry of the face and it is derived by manual annotation of the facial image, while texture provides a normalized (shape-free) face image which is obtained by warping the original facial image to the mean shape (the average of the aligned shapes corresponding to the training images). To reduce the dimensionality of the original shape and texture spaces, PCA, constrained by EFM, derives low dimensional shape and texture features, respectively. The low dimensional features are integrated using a normalization procedure to form augmented features accounting for both shape and texture information. The augmented features are then processed by EFM for face recognition.

3.1. Principal Component Analysis (PCA)

PCA is a standard decorrelation technique and following its application one derives an orthogonal projection basis which directly leads to dimensionality reduction, and possibly to feature selection. Let $X \in \mathbb{R}^N$ be a random vector representing a shape or an image, where N is the dimensionality of the corresponding shape or image space. For a shape the vector consists of the coordinates of the control points representing the shape, and for an image the vector is formed by concatenating the rows or the columns of the image which may be normalized to have a unit norm and/or an equalized histogram. The covariance matrix of X is defined as

$$\Sigma_X = E\{[X - E(X)][X - E(X)]^t\} \tag{1}$$

where $E(\cdot)$ is the expectation operator, t denotes the transpose operation, and $\Sigma_X \in \mathbb{R}^{N\times N}$. The PCA of a random vector X factorizes the covariance matrix Σ_X into the following form

$$\Sigma_X = \Phi\Lambda\Phi^t \tag{2}$$

where $\Phi = [\phi_1\phi_2\ldots\phi_N] \in \mathbb{R}^{N\times N}$ is an orthonormal eigenvector matrix and $\Lambda = diag\{\lambda_1, \lambda_2, \ldots, \lambda_N\} \in \mathbb{R}^{N\times N}$ a diagonal eigenvalue matrix with diagonal elements in decreasing order ($\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$). $\phi_1, \phi_2, \ldots, \phi_N$ and $\lambda_1, \lambda_2, \ldots, \lambda_N$ are the eigenvectors and the eigenvalues of Σ_X, respectively.

An important property of PCA is decorrelation, i.e., the components of the transformed data, $X' = \Phi^t X$, are decorrelated since the covariance matrix of X' is diagonal, $\Sigma_{X'} = \Lambda$, and the diagonal elements are the variances of the corresponding components. Another property of PCA is its optimal signal reconstruction in the sense of minimum Mean Square Error (MSE) when only a subset of principal components, $P = [\phi_1\phi_2\ldots\phi_m]$ where $m < N$ and $P \in \mathbb{R}^{N\times m}$, are used to represent the original signal. Following this property, an immediate application of PCA is the dimensionality reduction

$$Y = P^t X \tag{3}$$

The lower dimensional vector $Y \in \mathbb{R}^m$ captures the most expressive features of the original data X.

3.2. Shape and Texture

The feature geometry of a face, shape (vector), is represented by a set of control points which are derived by manual annotation. As shown in Fig. 1 the control points underscore important shape features such as eyebrows, eyes, bridge of nose, nose, mouth, and the contour of the face.

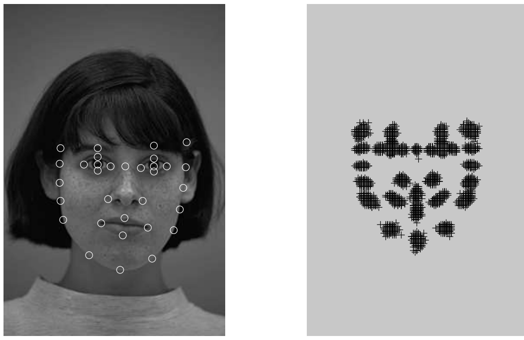

Figure 1. 32 control points representing the shape of a face image (left) and the aligned shapes (right)

The shapes of all the training images are aligned with respect to one another using translation, rotation and scaling. To simplify the alignment procedure we carried it out stepwise as follows: (i) rotate each shape of a training image so that the centers of the eyes are on a horizontal line; (ii) translate each shape to its own centroid; (iii) iteratively scale the shapes according to the so-far-derived average shape and a fixed reference size whose role is to prevent endless shrinking. The aligned shapes of the 400 training images are shown in Fig. 1.

The texture can now be defined as a normalized (shape-free) face image which is obtained by warping the original facial image to the mean shape, the average of the aligned shapes of the training images (see Fig. 1). Warping is carried out using a triangulation procedure and the triangles used in our experiments are shown in Fig. 2. Note that not all the control points are used during the triangulation and one point (the * point in Fig. 2) is the average of the upper-left and upper-right control points. Keeping the number of triangles small helps to increase processing speed and improve texture quality (smoothness). Examples of textures are shown in Fig. 3.

Let X_1 and X_2 represent the shape and the texture, respectively. The shape vector consists of the coordinates of the control points (see Fig. 1). From Eq. 1 and Eq. 2 one derives the covariance matrices, Σ_{X_1} and Σ_{X_2}, and the principal components, Φ_1 and Φ_2, of shape and texture. Choosing only a subset of principal components and using Eq. 3 one then derives lower dimensional shape and texture features, Y_1 and Y_2 (see [10] for detail). The low dimensional features are then integrated using the following normalization procedure to form augmented features encoding both shape and texture information.

$$Y = \left(\frac{Y_1^t}{\| Y_1 \|} \quad \frac{Y_2^t}{\| Y_2 \|} \right)^t \tag{4}$$

For comparison purposes (see Sect. 4) we also compute

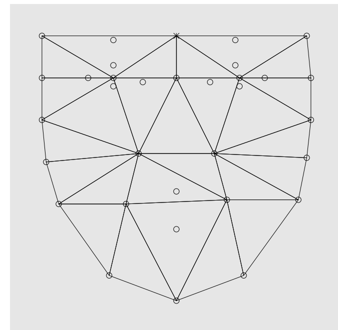

Figure 2. Triangles used for warping an individual image to the mean shape

Figure 3. Textures from the training set

the shape images, images undergoing the same alignment procedure as the shapes do, while preserving the intensity information within the shapes only. Examples of shape images are shown in Fig. 4 and correspond to the same faces shown in Fig. 3.

Figure 4. Shape images from the training set

3.3. Face Recognition

When an unknown face image is presented to the EFM classifier, the shape and the texture (with respect to the mean shape of the training images) of the image are first derived (see Sect. 3.2), and the augmented vector, Y, is then formed (see Eq. 4). Let T be the overall transformation matrix of EFM as defined in [10]. The new feature U of the unknown face image is derived as follows.

$$U = T^t Y \tag{5}$$

Let $U_k^0, k = 1, 2, \ldots, n$, be the prototype for class k, the mean of the training samples of class k after the EFM transformation. The classification rule is then specified as follows.

$$\|U - U_k^0\| = \min_j \|U - U_j^0\|, \qquad U \in \omega_k \tag{6}$$

The unknown face is classified to belong to that class ω_k from which the features of the face, U, yield the minimum Euclidean distance.

4. Experiments

The efficiency of the novel face recognition procedure introduced in this paper, i.e., using EFM based on integrated shape and texture information, is assessed using the FERET facial database [14]. The experiments involve 600 face images of size 256×384 with 256 gray scale levels and corresponding to 200 subjects such that each subject has 3 images. Since the images are acquired during different photo sessions both the lighting conditions and the facial expression may vary (see Fig. 5). 2 images are randomly chosen from the 3 images available for each subject for training, while the remaining one is used for testing.

Figure 5. Some example images used in our experiments

First a comparative assessment has been carried out regarding the input representations and classification methods. The kinds of input include shape images (Fig. 4), masked images (Fig. 6), shape vectors (Fig. 1), textures (shape-free images) (Fig. 3), and augmented shape and texture features as derived using Eq. 4. Shape, shape images and textures are derived as described in Sect. 3.2, and their size is 64, 74×73 and 70×69, respectively. Masked images are derived by first using the centers of two eyes as control points for alignment, and then masking them to yield 120×130 images. The dimension of the augmented shape and texture feature space is 70, which is derived by PCA reducing the shape and the texture spaces to 20 and 50, respectively. The reason for making such choices is discussed in detail by Liu and Wechsler in [10]. The classifiers used are L1 and L2, corresponding to the eigenfaces method [17] using L_1 or L_2 distance measure, and M1 and M2, corresponding to the Mahalanobis distance classifier using a common covariance matrix for all classes [9], [5] or a pooled within-class covariance matrix [7].

Figure 6. Masked images from the training set

Fig. 7, Fig. 8, Fig. 9, Fig. 10 and Fig. 11 show the face recognition performance using different classifiers with respect to different kinds of input. The recognition rate is the percentage of the top response being correct. One can see from Fig. 7 to Fig. 11 that (i) the augmented shape and texture features carry the most discriminant information followed in order by textures, shape vectors, masked images, and shape images; (ii) M2 consistently performs better than M1 which is followed by L1 and L2.

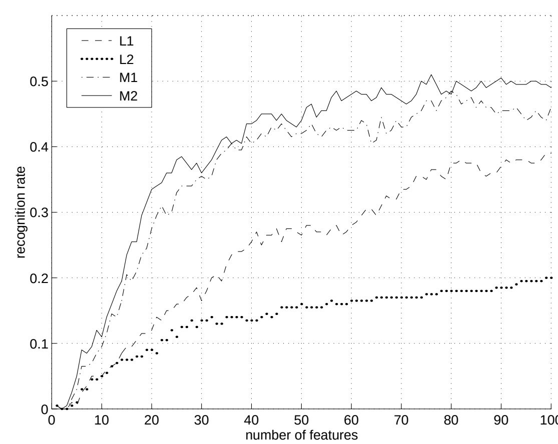

Figure 7. Recognition performance using shape images

The reason that Mahalanobis distance classifiers perform better than eigenfaces is that the Mahalanobis distance mea-

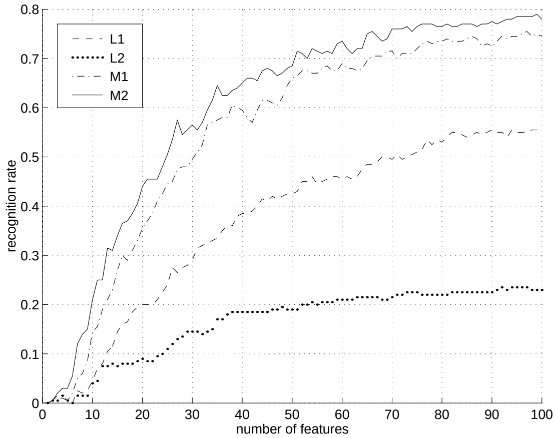

Figure 8. Recognition performance using masked images

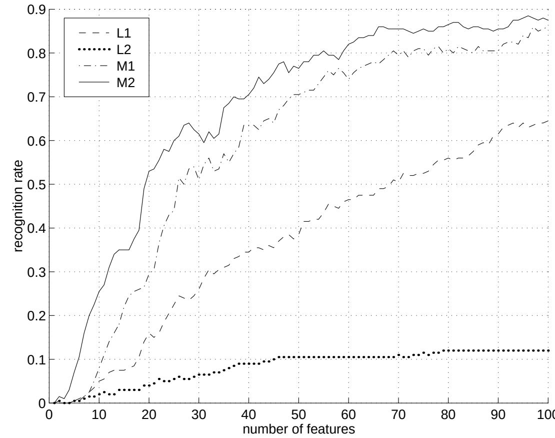

Figure 10. Recognition performance using textures

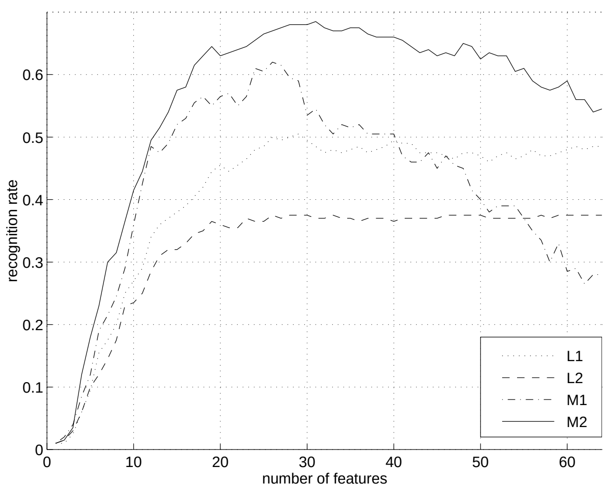

Figure 9. Recognition performance using the shape vectors

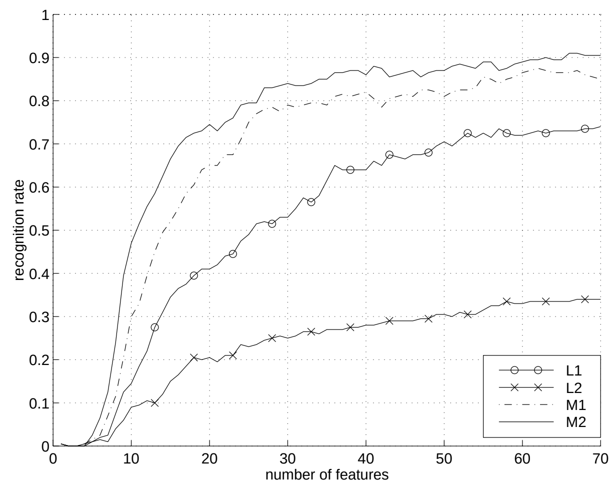

Figure 11. Recognition performance using augmented shape and texture

sure counteracts the fact that L_1 or L_2 distance measure in the PCA subspace weights preferentially for low frequencies. Such behavior should be expected even more prominent when faces are aligned and cropped, as the first several leading eigenvalues encode then mostly for prototypical (norm) representational aspects rather than discrimination information. As the L_2 measure weights more the low frequencies than L_1 does, the L1 classifier should perform better than L2, a conjecture validated by our experiments. While the M1 classifier uses a common covariance matrix for all the classes and it derives eigenvalues which encode for both within- and between-class scatter, the M2 classifier differentiates between the two scatters. Again, the reasonable expectation that M2 should perform better than M1 is validated by our experiments. Actually, Liu and Wechsler have shown that under specific assumption, the M2 classifier approximates the Bayes classifier [12].

As we determined that the best face recognition performance is achieved by the M2 classifier using augmented shape and texture information, we compared then M2 against EFM using the same augmented shape and texture features. One can see from Fig. 12 that EFM consistently yields better performance than the M2 classifier. In particular, EFM achieves 98.5% recognition rate using only 25 features.

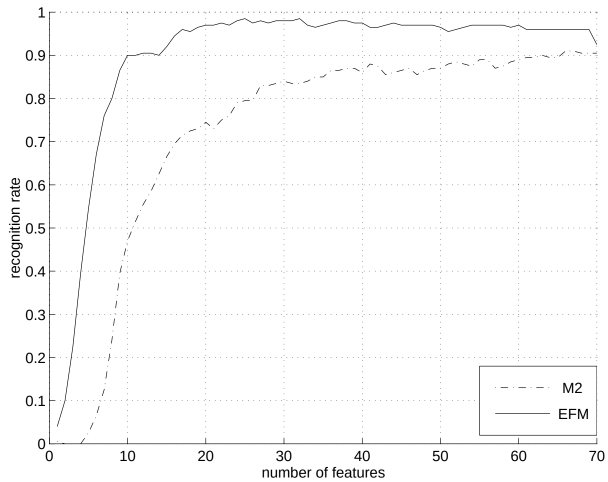

Figure 12. Comparative recognition performance for the EFM and M2 classifiers using augmented shape and texture

5. Conclusions

We introduced in this paper a new face coding and recognition method which employs the EFM using augmented shape and texture information. Our experiments show that the augmented shape and texture features carry the most discriminant information followed in order by textures, shape vectors, masked images, and shape images. Experimental results, using 600 images of varying illumination and corresponding to 200 subjects whose facial expression can vary, show that our new face coding and recognition method performs the best among the eigenfaces method [17] using L_1 or L_2 distance measure, and the Mahalanobis distance classifiers using a common covariance matrix for all classes [9], [5] or a pooled within-class covariance matrix [7]. In particular, EFM achieves 98.5% recognition rate using only 25 features.

Acknowledgments: This work was partially supported by the DoD Counterdrug Technology Development Program, with the U.S. Army Research Laboratory as Technical Agent, under contract DAAL01-97-K-0118. The authors would like to thank Jonathon Phillips and Hyeonjoon Moon for supplying us the mask images, and Albert Pujol for the manual annotation.

References

[1] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using class specifi c linear projection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.

[2] D. Beymer. Vectorizing face images by interleaving shape and texture computations. A.I. memo No. 1537, Artifi cial Intelligence Laboratory, MIT, 1995.

[3] R. Chellappa, C. Wilson, and S. Sirohey. Human and machine recognition of faces: A survey. *Proc. IEEE*, 83(5):705–740, 1995.

[4] I. Craw and P. Cameron. Face recognition by computer. In *Proc. British Machine Vision Conference*, pages 498–507, 1992.

[5] I. Craw, N. Costen, and T. Kato. How should we represent faces for automatic recognition? *to appear in IEEE Trans. Pattern Analysis and Machine Intelligence*.

[6] S. Edelman and N. Intrator. Learning as extraction of low-dimensional representations. In D. Medin, R. Goldstone, and P. Schyns, editors, *Mechanisms of Perceptual Learning*. Academic Press, 1990.

[7] G. Edwards, T. Cootes, and C. Taylor. Face recognition using active appearance models. In *Proc. the Fifth European Conference on Computer Vision*, University of Freiburg, Germany, June 2-6, 1998.

[8] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, second edition, 1991.

[9] A. Lanitis, C. Taylor, and T. Cootes. Automatic interpretation and coding of face images using flexible models. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(7):743–756, 1997.

[10] C. Liu and H. Wechsler. Enhanced Fisher linear discriminant models for face recognition. In *Proc. the 14th International Conference on Pattern Recognition*, Queensland, Australia, August 17-20, 1998.

[11] C. Liu and H. Wechsler. Face recognition using evolutionary pursuit. In *Proc. the Fifth European Conference on Computer Vision*, University of Freiburg, Germany, June 2-6, 1998.

[12] C. Liu and H. Wechsler. Probabilistic reasoning models for face recognition. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Santa Barbara, California, USA, June 23-25, 1998.

[13] B. Moghaddam and A. Pentland. Probabilistic visual learning for object representation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(7):696–710, 1997.

[14] P. Phillips, H. Wechsler, J. Huang, and P. Rauss. The FERET database and evaluation procedure for face recognition algorithms. *Image and Vision Computing*, 16:295–306, 1998.

[15] A. Samal and P. Iyengar. Automatic recognition and analysis of human faces and facial expression: A survey. *Pattern Recognition*, 25(1):65–77, 1992.

[16] D. L. Swets and J. Weng. Using discriminant eigenfeatures for image retrieval. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 18(8):831–836, 1996.

[17] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 13(1):71–86, 1991.

[18] T. Vetter and T. Poggio. Linear object classes and image synthesis from a single example image. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(7):733–742, 1997.