Appears in the *Int'l Joint Conf. on Neural Networks*, **IJCNN'99**, Renaissance Hotel, Washington, DC, July 10-16, 1999.

An Integrated Shape and Intensity Coding Scheme for Face Recognition

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

Abstract

This paper introduces a new face coding scheme which employs an Enhanced Fisher Classifier (EFC) operating on integrated shape and intensity features. The dimensionalities of the shape and the intensity image spaces are first reduced using Principal Component Analysis (PCA), constrained by the EFC for enhanced generalization. The reduced shape and the intensity features are then integrated through a normalization procedure to form integrated features. Experiments using 600 face images from the FERET database of varying illumination and corresponding to 200 subjects, whose facial expression can vary, show the feasibility of the new face coding scheme. In particular, the EFC achieves 98.5% recognition rate using only 25 features. Our experiments also show that the integrated shape and intensity features carry the most discriminating information followed in order by textures, shape vectors, masked images, and shape images.

1 Introduction

Learning to recognize visual objects, such as human faces, requires the ability to derive salient features from the raw input data in order to reduce the amount of data used for classification and simultaneously provide enhanced discriminatory power [3], [12]. Recently shape and intensity (texture or 'shape-free' image) have become prominent for encoding face images [4], [2], [15], [8]. 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 scheme which employs an Enhanced Fisher Classifier (EFC) operating on integrated shape and intensity features. Experimental results, using 600 face images from the FERET database of varying illumination and corresponding to 200 subjects whose facial expression can vary, show that (i) the integrated shape and intensity 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, EFC, performs the best among the eigenfaces method [14] using L_1 or L_2 distance measure, and the Mahalanobis distance classifiers using a common covariance matrix for all classes [8], [5] or a pooled within-class covariance matrix [6].

2 An Integrated Shape and Intensity Coding Scheme

Shape encodes the feature geometry of the face and it is derived by manual annotation of the facial image, while texture (intensity information) 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 the EFC for enhanced generalization, derives low dimensional shape and intensity features, respectively. The low dimensional features are integrated using a normalization procedure to form integrated features accounting for both shape and intensity information. The integrated features are then processed by EFC for face recognition.

2.1 Shape and Intensity Information

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,

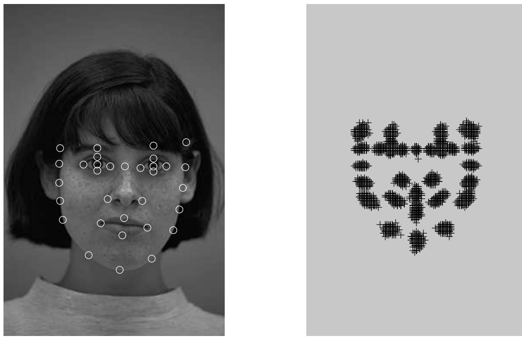

Figure 1. 32 control points representing the shape of a face image (left) and the aligned shapes (right)

bridge of nose, nose, mouth, and the contour of the face. The shapes of all the training images are aligned with respect to one another using translation, rotation and scaling. 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. Examples of textures are shown in Fig. 2.

Figure 2. Textures from the training set

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). Choosing only a subset of principal components one then derives lower dimensional shape and texture features, Y_1 and Y_2 (see next section for detail). The low dimensional features are then integrated using the following normalization procedure to form integrated features encoding both shape and intensity information.

$$Y = \left(\frac{Y_1^t}{\| Y_1 \|} \quad \frac{Y_2^t}{\| Y_2 \|} \right)^t \tag{1}$$

For comparison, we also compute the shape images, images undergoing the same alignment procedure as the shapes do, while preserving the intensity information within the shape outlines only. Examples of shape images are shown in Fig. 3.

Figure 3. Shape images from the training set

2.2 An Enhanced Fisher Classifier (EFC)

Fisher Linear Discrimimant (FLD) is a popular discriminant criterion which measures the between-class scatter Σ_b normalized by the within-class scatter Σ_w [7]. FLD derives a projection matrix Ψ that maximizes the ratio $|\Psi^t \Sigma_b \Psi| \,/\, |\Psi^t \Sigma_w \Psi|$. This ratio is maximized when Ψ consists of the eigenvectors of the matrix $\Sigma_w^{-1}\Sigma_b$ [7].

$$\Sigma_w^{-1}\Sigma_b \Psi = \Psi \Delta \tag{2}$$

where Ψ and Δ are the eigenvector and eigenvalue matrices of $\Sigma_w^{-1}\Sigma_b$, respectively.

The standard FLD based methods such as Fisherfaces [1] and the MDF method [13] apply first PCA for dimensionality reduction and then discriminant analysis for face recognition. Relevant questions concerning PCA are usually related to the number of principal components used and how it affects performance. Regarding discriminant analysis one has to understand the reasons for overfitting and how to avoid it. The answers to those two questions are closely related. One can actually show that using more principal components may lead to decreased performance (for recognition). The explanation for this behavior is that the trailing eigenvalues correspond to high-frequency components and mainly encode noise. As a result, when these trailing but small valued eigenvalues are used to define the reduced PCA subspace, the FLD procedure has to fit for noise as well and as a consequence overfitting takes place.

In order to improve the generalization capability of FLD based classifiers, this paper introduces an Enhanced Fisher Classifier (EFC) which decomposes the FLD procedure into a simultaneous diagonalization of the two within- and between-class scatter matrices. The simultaneous diagonalization is stepwisely equivalent to two operations as pointed out by Fukunaga [7]: (i) transform the original data so that the within-class scatter matrix is whitened, and (ii) apply PCA on the between-class scatter matrix corresponding to the transformed data. As the eigenvalues of the within-class scatter matrix appear in the denominator during whitening, the small (trailing) eigenvalues cause the whitening step to fit for misleading variations and would thus generalize poorly when exposed to new data. For enhanced performance EFC should thus maintain a proper balance between the need that the selected eigenvalues (corresponding to the principal components for the original image space) account for most of the spectral energy of the raw data and the requirement that the eigenvalues of the within-class scatter matrix (in the reduced PCA subspace) are not too small.

The choice of the number of principal components (m) for dimensionality reduction addresses both the energy needs and the eigenvalues magnitude requirement. The eigenvalue spectrum of the covariance matrix supplies a good indicator for meeting the energy need, while the eigenvalue spectrum of the within-class scatter matrix in the reduced PCA subspace should facilitate choosing the number of principal components to meet the magnitude requirement. In particular, the stepwise FLD procedure derives the eigenvalues and eigenvectors of $\Sigma_w^{-1}\Sigma_b$ as the result of the simultaneous diagonalization of Σ_w and Σ_b. First whiten the within-class scatter matrix

$$\Sigma_w \Xi = \Xi \Gamma \quad and \quad \Xi^t \Xi = I \tag{3}$$

$$\Gamma^{-1/2} \Xi^t \Sigma_w \Xi \Gamma^{-1/2} = I \tag{4}$$

The eigenvalue spectrum of the within-class scatter matrix in the reduced PCA subspace can be derived by Eq. 3, and different spectra are obtained corresponding to different number of principal components utilized. Now one has to simultaneously optimize the behavior of the trailing eigenvalues in the reduced PCA space (Eq. 3) with the energy criteria for the original image space. One needs to find two sets of principal components for shape and texture, respectively, before forming the integrated shape and intensity features (see Eq. 1).

After the integrated vector (Eq. 1) is derived, EFC proceeds to compute the between-class scatter matrix as follows:

$$\Gamma^{-1/2} \Xi^t \Sigma_b \Xi \Gamma^{-1/2} = K_b \tag{5}$$

Diagonalize the new between-class scatter matrix K_b

$$K_b \Theta = \Theta \Delta \quad and \quad \Delta^t \Delta = I \tag{6}$$

The overall transformation matrix (after Eq. 1) of EFC is now defined as follows:

$$T = \Xi \Gamma^{-1/2} \Theta \tag{7}$$

3 Experiments

The efficiency of the novel face recognition procedure, EFC, is assessed using the FERET facial database [11]. 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. 2 images are randomly chosen from the 3 images available for each subject for training, while the remaining one is used for testing.

First a comparative assessment has been carried out regarding the input representations and classification methods. The kinds of input include shape images (Fig. 3), masked images (Fig. 4), shape vectors (Fig. 1), textures (shape-free images) (Fig. 2), and integrated shape and intensity features as derived using Eq. 1. Shape, shape images and textures are derived as described in the second section, 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 integrated 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 the previous section and detailed by Liu and Wechsler [9]. The classifiers used are L1 and L2, corresponding to the eigenfaces method [14] 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 [8], [5] or a pooled within-class covariance matrix [6].

Figure 4. Masked images from the training set

Fig. 5, Fig. 6, Fig. 7, Fig. 8 and Fig. 9 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. 5 to Fig. 9 that (i) the integrated shape and intensity 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.

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

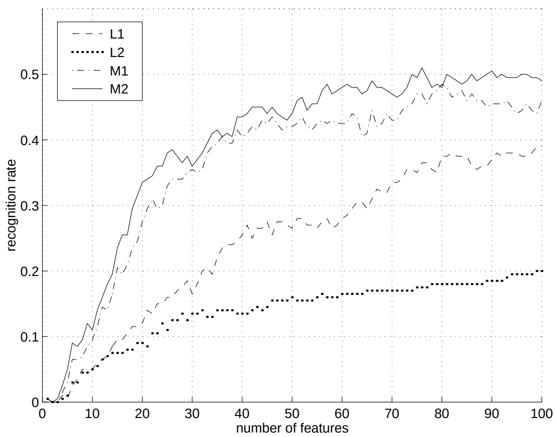

Figure 5. Recognition performance using shape images

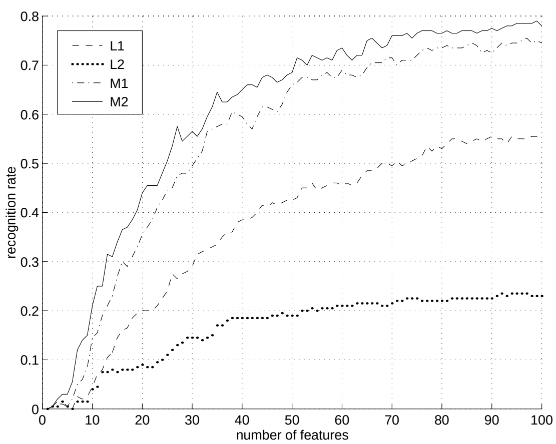

Figure 6. Recognition performance using masked images

nent 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 [10].

As we determined that the best face recognition performance is achieved by the M2 classifier using integrated shape and intensity features, we compared then M2 against EFC using the same integrated shape and intensity features. One can see from Fig. 10 that EFC consistently yields better performance than the M2 classifier. In particular, EFC achieves 98.5% recognition rate using only 25 features.

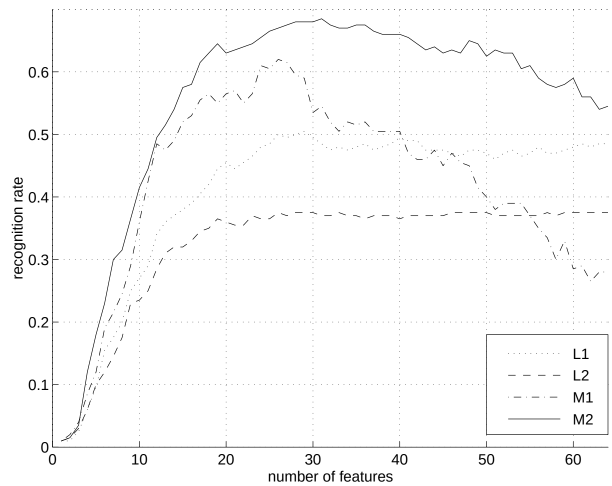

Figure 7. Recognition performance using the shape vectors

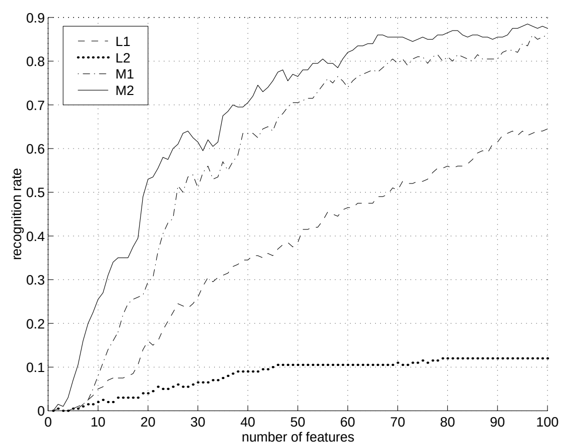

Figure 8. Recognition performance using textures

4 Conclusions

We introduced in this paper a new face coding and recognition method which employs the Enhanced Fisher Classifier (EFC) using integrated shape and intensity information. Experimental results, using 600 images from the FERET database of varying illumination and corresponding to 200

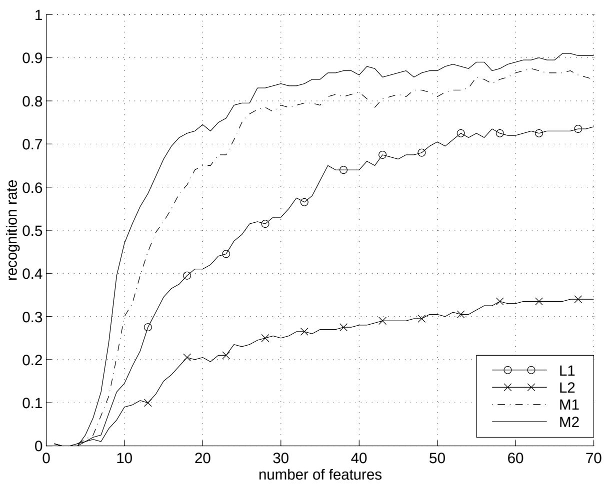

Figure 9. Recognition performance using integrated shape and texture

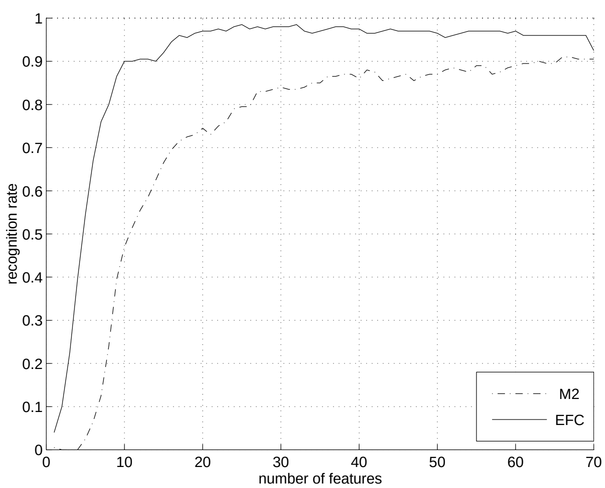

Figure 10. Comparative recognition performance for the EFC and M2 classifiers using integrated shape and intensity features

subjects whose facial expression can vary, show that our new face coding and recognition method performs the best among the eigenfaces method [14] using L_1 or L_2 distance measure, and the Mahalanobis distance classifiers using a common covariance matrix for all classes [8], [5] or a pooled within-class covariance matrix [6]. In particular, EFM achieves 98.5% recognition rate using only 25 features. Our experiments also show that the integrated shape and intensity features carry the most discriminant information followed in order by textures, shape vectors, masked images, and shape images.

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 specific 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, Artificial 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] 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.

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

[8] 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.

[9] 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.

[10] 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.

[11] 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.

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

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

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

[15] 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.