IEEE Trans. Image Processing, vol. 10, no. 4, pp. 598-608, 2001.

A Shape and Texture Based Enhanced Fisher Classifier for Face Recognition

Chengjun Liu [*], *Member, IEEE*, and Harry Wechsler [†], *Fellow, IEEE*

Abstract— This paper introduces a new face coding and recognition method, the Enhanced Fisher Classifier (EFC), which employs the enhanced Fisher linear discriminant model (EFM) on integrated shape and texture features. Shape encodes the feature geometry of a face while texture provides a normalized shape-free image. The dimensionalities of the shape and the texture spaces are first reduced using principal component analysis, constrained by the EFM for enhanced generalization. The corresponding reduced shape and texture features are then combined through a normalization procedure to form the integrated features that are processed by the EFM for face recognition. Experimental results, using 600 face images corresponding to 200 subjects of varying illumination and facial expressions, show that (i) the integrated shape and texture features carry the most discriminating information followed in order by textures, masked images, and shape images; (ii) the new coding and face recognition method, EFC, performs the best among the Eigenfaces method using L_1 or L_2 distance measure, and the Mahalanobis distance classifiers using a common covariance matrix for all classes or a pooled within-class covariance matrix. In particular, EFC achieves 98.5% recognition accuracy using only 25 features.

[*]Chengjun Liu is with the Department of Mathematics and Computer Science, University of Missouri, St. Louis, MO 63121. E-mail: cliu@cs.umsl.edu.

[†]Harry Wechsler is with the Department of Computer Science, George Mason University, Fairfax, VA 22030. E-mail: wechsler@cs.gmu.edu.

Index Terms — Face recognition, Enhanced Fisher Classifier (EFC), Principal Component Analysis (PCA), Fisher Linear Discriminant (FLD), Enhanced FLD Model (EFM), shape and texture

1. Introduction

A successful face recognition methodology depends heavily on the particular choice of the features used by the (pattern) classifier [4], [20]. 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 [6], [2], [23], [13]. 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 [6].

This paper introduces a new face coding and recognition method that employs the Enhanced Fisher Classifier (EFC) operating on integrated shape and texture features, and assesses comparatively the types of input for face representation against some popular face recognition methods. 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 combined through a normalization procedure to form **the integrated shape and texture features**. The dimensionality reduction procedure, constrained by the Enhanced FLD (Fisher Linear Discriminant) Model (EFM) for better generalization [17], maintains a proper balance between the spectral energy needs of PCA for adequate representation, and the Fisher Linear Discriminant (FLD) requirements that the eigenvalues of the within-class covariance matrix should not include small trailing values as they tend to encode noise and appear in the denominator. The other two types of input assessed in this paper are the shape images and the masked images. **Shape images** undergo the same alignment procedure as the shapes do, but preserve the intensity information within the contours of the faces. **Masked images** are derived by first using the centers of two eyes as control points for alignment, and then placing a mask on them.

Experimental results, using 600 face images corresponding to 200 subjects of varying illumination and facial expression, show that (i) the integrated shape and texture features carry the most discriminating information followed in order by textures, masked images, and shape images; (ii) our new face coding and recognition method, EFC, performs the best among the Eigenfaces method [22] using L_1 or L_2 distance measure, and the Mahalanobis distance based methods using a common covariance matrix for all classes [13], [7] or a pooled within-class covariance matrix [10]. In particular, EFC 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 [16]. Psychophysical 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" [9]. Principal Component Analysis (PCA) [12] is the method behind the Eigenfaces coding scheme [22] 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 [14], [16].

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) [21] 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 [17], [8]. One can show that the MDF space is, however, superior to the PCA space for face recognition, only when the training images are representative of the range of face (class) variations; otherwise, the performance difference between the PCA and MDF spaces is not significant [21].

To further improve PCA stand-alone methods, both new face representation approaches and new classifiers are emerging. Beymer [2] introduced a vectorized image representation consisting of shape and texture. Vetter and Poggio [23] used such a vectorized face representation for image synthesis from a single example view. Craw, et al [7] and Lanitis, Taylor and Cootes [13] developed Mahalanobis distance classifiers for face recognition using the shape and texture 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 giving more weight to components corresponding to smaller eigenvalues [7]. Note that the weighting procedure does not differentiate the between-class scatter from the within-class scatter and it 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 [10] 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, Enhanced Fisher Classifier (EFC), which employs the Enhanced FLD Model (EFM) [17] on integrated shape and texture features. The shape encodes the feature geometry of the face and it is derived by manual annotation of the facial image, while the texture provides a normalized (shape-free) face image that is obtained by warping the original facial image to the mean shape (the average of the aligned shapes of all the training images). To reduce the dimensionality of the original shape and texture spaces, PCA, constrained by the EFM for enhanced generalization, derives low dimensional shape and texture features, respectively. The low dimensional features are then combined using a normalization procedure to form integrated features accounting for both shape and texture information. Finally, the integrated features are processed by the 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 that 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, while for an image the vector is formed by concatenating the rows or the columns of the image which may be normalized to have zero mean and unit norm. The covariance matrix of X is defined as follows:

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

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 $\Phi = [\phi_1 \phi_2 \dots \phi_N] \in \mathbb{R}^{N \times N}$ is an orthonormal eigenvector matrix and $\Lambda = diag\{\lambda_1, \lambda_2, \dots, \lambda_N\} \in \mathbb{R}^{N \times N}$ a diagonal eigenvalue matrix with diagonal elements in decreasing order ($\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$). $\phi_1, \phi_2, \dots, \phi_N$ and $\lambda_1, \lambda_2, \dots, \lambda_N$ 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 important property of PCA is its optimal signal reconstruction in the sense of minimum Mean Square Error (MSE) when only a subset of principal components are used to represent the original signal. Following this property, an immediate application of PCA is the dimensionality reduction:

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

where $P = [\phi_1 \phi_2 \ldots \phi_m]$, $m < N$, and $P \in \mathbb{R}^{N \times m}$. 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 that are derived by manual annotation. The control points, shown in Fig. 1, underscore the important shape features such as the eyebrows, eyes, bridge of nose, nose, mouth, and the contour of the face. Shapes of all the training images are aligned with respect to one another by applying the rotation, translation and scaling transformations. In particular, the alignment procedure performs the following operations: (i) rotate the shape of each training image so that the centers of the eyes are on a horizontal line; (ii) translate the shape to its own centroid; (iii) iteratively scale the shape according to the previous average shape and a fixed reference size whose role is to prevent endless shrinking in size. The aligned shapes of the training images are shown in Fig. 1. The figure plots, using plus signs, the distribution of the aligned shapes of the training images, and it shows the variation in the location of the control points across all the images.

Fig. 1 goes here

Texture is a normalized (shape-free) face image that is warped to the mean shape, the average of the aligned shapes of the training images as shown in Fig. 1. The warping transformation first implements a triangulation procedure to partition a face into a number of small triangular regions, and then performs an affine transformation for every small triangular region to warp the original face image to the mean shape. After the warping transformation, all the textures (shape-free images) have the same face contour as the mean shape does. The triangles used in our experiments are shown in Fig. 2. In contrast to the alignment procedure, which uses all the control points, the triangulation procedure does not utilize all the control points. One point, the * point in Fig. 2, actually is not a control point. It is the average of the upper-left and upper-right control points.

Using only some of the control points for the triangulation procedure reduces the number of small triangles, helps to increase processing speed and improves texture quality (smoothness). Examples of textures are shown in Fig. 3.

Fig.s 2 and 3 go here

Let $X_1 \in \mathbb{R}^{N_1}$ and $X_2 \in \mathbb{R}^{N_2}$ represent the shape and the texture of a face image, where N_1 and N_2 are the dimensionalities of the shape and the texture spaces, respectively. The shape consists of the coordinates of the control points (see Fig. 1). Using Eq. 1 and Eq. 2 one can derive the covariance matrices, $\Sigma_{X_1} \in \mathbb{R}^{N_1 \times N_1}$ and $\Sigma_{X_2} \in \mathbb{R}^{N_2 \times N_2}$, and the eigenvector matrices, $\Phi_1 \in \mathbb{R}^{N_1 \times N_1}$ and $\Phi_2 \in \mathbb{R}^{N_2 \times N_2}$, of the shape and the texture, respectively. To improve the generalization performance of the EFC classifier, one should choose only a subset of principal components to derive the lower dimensional shape and texture features, $Y_1 \in \mathbb{R}^{m_1}$ and $Y_2 \in \mathbb{R}^{m_2}$ (using Eq. 3), where m_1 and m_2 are the dimensionalities of the reduced shape space and the reduced texture space, respectively. 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 the high-frequency components and mainly encode noise. As a result, when these trailing but small valued eigenvalues are used to define the reduced PCA space, the subsequent FLD procedure has to fit for noise as well and as a consequence overfitting takes place (see Sect. 3.3).

The remaining question now is how to determine the dimensionalities for the reduced shape space (m_1) and the reduced texture space (m_2), respectively. Note that the goal of using Eq. 3 for dimensionality reduction is two-fold. On the one hand, we hope to lose as little representative information of the original data as possible in the transformation from the high dimensional space to the low dimensional one. On the other hand, in the reduced space the small trailing eigenvalues of the within-class covariance matrix are excluded so that we can obtain more robust subsequent FLD projection (see Sect. 3.3). As a result, the eigenvalue spectrum of the within-class covariance matrix of the training data supplies useful information regarding the choice for the dimensionality of the space (see Sect. 3.3).

The low dimensional features are then integrated using the following normalization procedure to form integrated features encoding both shape and texture information:

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

where $Z \in \mathbb{R}^{m_1+m_2}$, and m_1 and m_2 are the dimensionalities of the reduced shape space and the reduced texture space, respectively. Note that a concatenated shape and texture vector has been earlier suggested by Cootes et al [5]. They use a diagonal weight matrix to combine the shape and texture features. Rather than calculating a weight matrix, we implement a simple normalization procedure to commensurate the lower dimensional shape and texture features as shown in Eq. 4. The rational behind such a simple normalization is that the shapes and textures are viewed as equal important discriminating information. To achieve such a goal, the shape and texture features are normalized to have unit norms, respectively, before they are concatenated to form an augmented feature vector. Note that these lower dimensional shape and texture features derived by PCA, Y_1 and Y_2, are constrained by the EFM for enhanced generalization performance (see Sect. 3.3).

For comparison purposes (see Sect. 4) we also compute the shape images and masked images. Shape images undergo the same alignment procedure as the shapes do, but preserve the intensity information within the contours of the faces only. Examples of shape images are shown in Fig. 4. Note that there are two major differences between shape images and textures: (i) Different shape images usually have different contours, while all the textures share the same contour (due to warping) specified by the mean shape. As a result, the average image of the shape images has a blurred contour, while the one of the texture images has a sharp contour as shown in Fig. 5. (ii) Each shape image undergoes only one alignment transformation, while each texture undergoes different affine transformations (warping) corresponding to the different small triangular regions partitioned by a triangulation procedure as shown in Fig. 2. 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.

Examples of masked images are shown in Fig. 6.

Fig.s 4, 5 and 6 go here

3.3. An Enhanced FLD Model (EFM)

We present in this section an Enhanced FLD Model (EFM) that determines the dimensionalities for the reduced shape space (m_1) and the reduced texture space (m_2), respectively. FLD is a popular discriminant criterion which measures the between-class scatter normalized by the within-class scatter [12]. Let Y be a random vector representing the lower dimensional shape or texture feature (Eq. 3) or the integrated shape and texture feature (Eq. 4). Let $\omega_1, \omega_2, \ldots, \omega_L$ and $N_1, N_2, \ldots, N_L$ denote the classes and the number of images within each class, respectively. Let $M_1, M_2, \ldots, M_L$ and M be the means of the classes and the grand mean. The within- and between-class covariance matrices Σ_w and Σ_b are defined as follows [12]:

$$\Sigma_w = \sum_{i=1}^{L} P(\omega_i) E\{(Y - M_i)(Y - M_i)^t | \omega_i\} \tag{5}$$

$$\Sigma_b = \sum_{i=1}^{L} P(\omega_i)(M_i - M)(M_i - M)^t \tag{6}$$

where $P(\omega_i)$ is *a priori* probability and $E(\cdot)$ denotes the expectation operator.

FLD derives a projection matrix Ψ that maximizes the ratio $|\Psi^t \Sigma_b \Psi| \, / \, |\Psi^t \Sigma_w \Psi|$ [1]. This ratio is maximized when Ψ consists of the eigenvectors of the matrix $\Sigma_w^{-1}\Sigma_b$ [21]:

$$\Sigma_w^{-1}\Sigma_b \Psi = \Psi \Delta \tag{7}$$

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

FLD is behind several face recognition methods [21], [1], [11], [17]. As the original image space is high dimensional, most methods first perform dimensionality reduction using PCA, as it is the case with the Fisherfaces method suggested by Belhumeur, Hespanha, and Kriegman [1]. Using similar arguments, Swets and Weng [21] point out that the Eigenfaces method only derives the Most Expressive Features (MEF). Such PCA inspired features do not necessarily provide for good discrimination. As FLD is capable of distinguishing the within- and the between-class scatters, subsequent FLD projections are used to build the Most Discriminating Features (MDF) classification space. The MDF space is, however, superior to the MEF space for face recognition only when the training images are representative of the range of face (class) variations; otherwise, the performance difference between the MEF and MDF is not significant [21]. The drawback of FLD is that it requires large sample sizes for good generalization [17].

In order to improve the generalization capability of FLD, the EFM method decomposes the FLD procedure into a simultaneous diagonalization of the two within- and between-class covariance matrices [17]. The simultaneous diagonalization is stepwise equivalent to two operations as pointed out by Fukunaga [12]: whitening the within-class covariance matrix and applying PCA on the between-class covariance matrix using the transformed data. The stepwise operation shows that during whitening the eigenvalues of the within-class covariance matrix appear in the denominator. As the small (trailing) eigenvalues tend to capture noise [17], they cause the whitening step to fit for misleading variations and thus generalize poorly when exposed to new data. As a result, for enhanced performance EFM should preserve a proper balance between **the energy criterion —** 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, for representational adequacy, and **the magnitude criterion** — the requirement that the eigenvalues of the within-class covariance matrix (in the reduced PCA space) are not too small, for better generalization.

The choice of the range of principal components (m) for dimensionality reduction (see Eq. 3) reflects both the energy need and the magnitude requirement. The eigenvalue spectrum of the covariance matrix (see Eq. 2) supplies a good indicator for meeting the energy criterion, and one needs further to derive the eigenvalue spectrum of the within-class covariance matrix in the reduced PCA space to facilitate the choice of the range of principal components so that the magnitude criterion is met. Toward that end, one should carry out the stepwise FLD process. 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 covariance matrix:

$$\Sigma_w \Xi = \Xi \Gamma \quad and \quad \Xi^t \Xi = I \tag{8}$$

$$\Gamma^{-1/2} \Xi^t \Sigma_w \Xi \Gamma^{-1/2} = I \tag{9}$$

where Ξ, Γ are the eigenvector and the diagonal eigenvalue matrices of Σ_w.

The eigenvalue spectrum of the within-class covariance matrix in the reduced PCA space can be derived by Eq. 8, and different spectra are obtained corresponding to different number of principal components utilized (see Eq. 3 and Eq. 5). Now one has to simultaneously optimize the behavior of the trailing eigenvalues in the reduced PCA space (Eq. 8) with the energy criteria for the original image space (Eq. 2). Following this procedure one can derive the lower dimensional shape and texture features, $Y_1 \in \mathbb{R}^{m_1}$ and $Y_2 \in \mathbb{R}^{m_2}$, respectively. The integrated shape and texture vector is then derived using Eq. 4.

Now, let the Y in Eq. 5 represent the integrated shape and texture feature Z (Eq. 4), and calculate the within- and between-class covariance matrices Σ_w and Σ_b (Eq. 5 and Eq. 6). EFM first diagonalizes the within-class covariance matrix Σ_w using Eq. 8 and 9. Note that now the Ξ and Γ are the eigenvector and the eigenvalue matrices corresponding to the integrated shape and texture features. EFM proceeds then to compute the between-class covariance matrix as follows:

$$\Gamma^{-1/2} \Xi^t \Sigma_b \Xi \Gamma^{-1/2} = K_b \tag{10}$$

Diagonalize the new between-class covariance matrix K_b:

$$K_b \Theta = \Theta \Delta \quad and \quad \Theta^t \Theta = I \tag{11}$$

where Θ, Δ are the eigenvector and the diagonal eigenvalue matrices of K_b.

The overall transformation matrix of EFM is now defined as follows:

$$T = \Xi \Gamma^{-1/2} \Theta \tag{12}$$

3.4. Face Recognition

The EFC method employs the EFM on the integrated shape and texture features. When an unknown face image is presented to the EFC classifier, the shape and the texture of the image are first calculated as explained in Sect. 3.2, and the integrated feature, Z, is then derived using Eq. 4. Let T be the overall transformation matrix of the EFM as defined by Eq. 12. The new feature U of the unknown face image is derived as follows:

$$U = T^t Z \tag{13}$$

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:

$$\|U - U_k^0\| = \min_j \|U - U_j^0\|, \quad U \in \omega_k \tag{14}$$

The unknown face is classified to class ω_k to whom the feature U (of the unknown face) is the nearest neighbor.

4. Experiments

The classification accuracy of the novel face recognition procedure introduced in this paper is assessed using the FERET facial database [19]. The experiments involve 600 face images corresponding to 200 subjects such that each subject has three images of size 256×384 with 256 gray scale levels. Since the images are acquired during different photo sessions, both the lighting conditions and the facial expressions may vary (see Fig. 7). Two images are randomly chosen from the three images available for each subject for training, while the remaining image is used for testing. Note that for a face recognition problem, usually there are a large number of faces (classes), but only a few training examples per face [18].

Fig. 7 goes here

The EFC method includes the following major procedures: texture computation (warping), PCA dimensionality reductions for both shape and texture, EFM implementation (Sect. 3.3), and recognition (Sect. 3.4). The computational complexity of the method mainly falls into two parts: the PCA and EFM procedures. If we use a singular value decomposition (SVD) algorithm to implement these two procedures, then according to [3], the SVD of matrix of size $N \times N$ has the complexity of $O(N^3)$.

The experiments include (i) a comparative assessment of the face recognition performance regarding the types of input for face representation; and (ii) the comparative performance of our EFC method against some popular face recognition methods using the most discriminative type of face representation determined by (i). The kinds of input include shape images (Fig. 4), masked images (Fig. 6), textures (shape-free images) (Fig. 3), and the integrated shape and texture features as derived using Eq. 4. The popular face recognition methods against which we compared our method are the Mahalanobis distance classifier utilizing a pooled within-class covariance matrix [10] (we call this method M2), the Mahalanobis distance classifier utilizing a common covariance matrix for all classes [13], [7] (M1), and the Eigenfaces method [22] applying L_1 distance measure (L1) and L_2 distance measure (L2), respectively.

Shape images, textures and masked images are derived as described in Sect. 3.2, and their sizes are 74×73, 70×69 and 120×130, respectively. The dimension of the integrated shape and texture features 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 for enhanced generalization performance that is discussed in Sect. 3.3. Fig. 8 shows the face recognition performance regarding the different types of input for face representation using the M2 method. The curves show that the integrated shape and texture features carry the most discriminating information followed in order by the textures, the masked images, and the shape images. The same conclusion holds true when the other three methods, M1, L1 and L2, are applied. For detailed information, please see [15].

The reason that the shape images perform the worst is that the shape images do not share the same contour while the masked images and the textures do. As a result, the variances on the contours for the shape images are quite large, indicating that the contours are chosen as important discriminative features. Fig. 9 and 10 show the first 30 basis vectors of the shape images and textures, respectively. One can see from these figures that while shape images emphasize the contours as discriminative features, the textures tend to pick internal face features for classification. The reason that the masked images perform worse than the textures do is that masked images discard shape information (while textures use such information, i.e. the mean shape, for warping) and do not perform image warping.

Fig.s 8, 9 and 10 go here

As we determined that the integrated shape and texture features carry the most discriminating information, we compared then our EFC method against the Mahalanobis distance based methods, M1 and M2, and the Eigenfaces method utilizing different distance measures, L1 and L2, using the same integrated shape and texture features. Fig. 11 shows the comparative face recognition performance of our new EFC method against the M1, M2, L1, and L2 methods. One can see from this figure that EFC consistently yields the best performance followed in order by the Mahalanobis distance based methods, M1 and M2, and the Eigenfaces method utilizing different distance measures, L1 and L2.

The reason that the Mahalanobis distance classifiers perform better than the Eigenfaces method is that the Mahalanobis distance measure counteracts the fact that L_1 or L_2 distance measure in the PCA space weights preferentially for low frequencies. Such behavior should be expected to be even more pronounced 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 scatters, the M2 classifier differentiates between these two scatters. Again, the reasonable expectation that M2 should perform better than M1 is validated by our experiments.

EFC method, which considers both classification accuracy and generalization performance, yields the best recognition results. In particular, EFC achieves 98.5% recognition rate using only 25 features compared to 79.5% for the M2 classifier using the same number of features.

Fig. 11 goes here

5. Conclusions

We introduced in this paper a new face coding and recognition method, Enhanced Fisher Classifier (EFC), which employs the Enhanced FLD Model (EFM) on integrated shape and texture features. Our experiments show that the integrated shape and texture features carry the most discriminating information followed in order by textures, masked images, and shape images. Experimental results, using 600 images corresponding to 200 subjects of varying illumination and facial expression, show that our new face coding and recognition method performs the best among the Eigenfaces method [22] using L_1 or L_2 distance measure, and the Mahalanobis distance classifiers using a common covariance matrix for all classes [13], [7] or a pooled within-class covariance matrix [10]. In particular, the EFC achieves 98.5% recognition rate using only 25 features.

Acknowledgments: The authors would like to thank the anonymous reviewers for their critical and constructive comments and suggestions. The authors also want to express their gratitude to Jonathon Phillips and Hyeonjoon Moon who supplied us the mask images, and to Albert Pujol who helped us for the manual annotation. 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.

References

[1] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 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] T.F. Chan, "An improved algorithm for computing the singular value decomposition," *ACM Trans. Math. Software*, vol. 8, pp. 72–83, 1982.

[4] R. Chellappa, C.L. Wilson, and S. Sirohey, "Human and machine recognition of faces: A survey," *Proc. IEEE*, vol. 83, no. 5, pp. 705–740, 1995.

[5] T.F. Cootes, G.J. Edwards, and C.J. Taylor, "Active appearance models," in *Proc. the Fifth European Conference on Computer Vision*, University of Freiburg, Germany, June 2-6, 1998.

[6] I Craw and P. Cameron, "Face recognition by computer," in *Proc. British Machine Vision Conference*, 1992, pp. 498–507.

[7] I. Craw, N. Costen, T. Kato, and S. Akamatsu, "How should we represent faces for automatic recognition?," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 8, pp. 725–736, 1999.

[8] G. Donato, M.S. Bartlett, J.C. Hager, P. Ekman, and T.J. Sejnowski, "Classifying facial actions," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 10, pp. 974–989, 1999.

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

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

[11] K. Etemad and R. Chellappa, "Discriminant analysis for recognition of human face images," *J. Opt. Soc. Am. A*, vol. 14, pp. 1724–1733, 1997.

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

[13] A. Lanitis, C.J. Taylor, and T.F. Cootes, "Automatic interpretation and coding of face images using flexible models," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 743–756, 1997.

[14] C. Liu and H. Wechsler, "Comparative assessment of independent component analysis (ICA) for face recognition," in *Proc. Second Int'l Conference on Audio- and Video-based Biometric Person Authentication*, Washington D. C., March 22-24, 1999.

[15] C. Liu and H. Wechsler, "Face recognition using shape and texture," in *Proc. Computer Vision and Pattern Recognition*, Fort Collins, CO, June 23-25, 1999.

[16] C. Liu and H. Wechsler, "Evolutionary pursuit and its application to face recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 6, pp. 570–582, 2000.

[17] C. Liu and H. Wechsler, "Robust coding schemes for indexing and retrieval from large face databases," *IEEE Trans. on Image Processing*, vol. 9, no. 1, pp. 132–137, 2000.

[18] P.J. Phillips, "Matching pursuit filters applied to face identification," *IEEE Trans. Image Processing*, vol. 7, no. 8, pp. 1150–1164, 1998.

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

[20] A. Samal and P.A. Iyengar, “Automatic recognition and analysis of human faces and facial expression: A survey,” *Pattern Recognition*, vol. 25, no. 1, pp. 65–77, 1992.

[21] D.L. Swets and J. Weng, “Using discriminant eigenfeatures for image retrieval,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 831–836, 1996.

[22] M. Turk and A. Pentland, “Eigenfaces for recognition,” *Journal of Cognitive Neuroscience*, vol. 13, no. 1, pp. 71–86, 1991.

[23] T. Vetter and T. Poggio, “Linear object classes and image synthesis from a single example image,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 733–742, 1997.

List of Figures

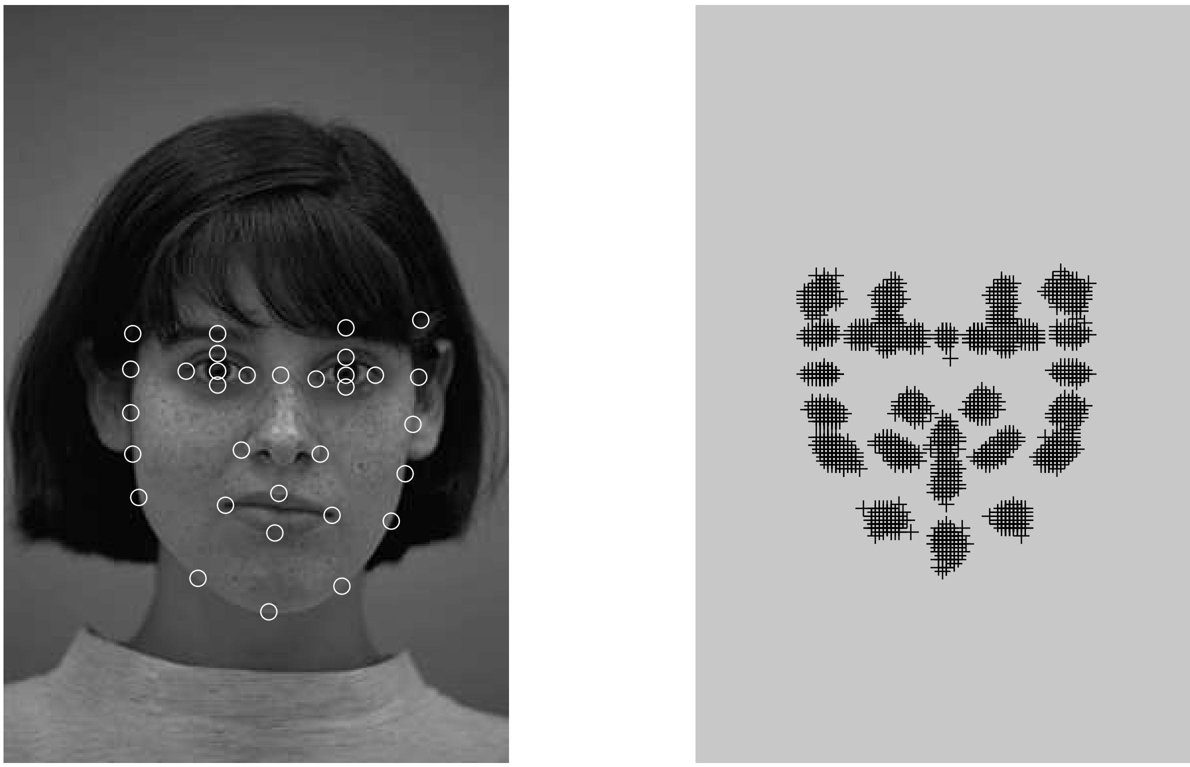

Figure 1. Left: the shape of a face image represented by 32 control points (white circles). Right: the distribution of the aligned shapes of the training images (plus signs). It shows the variation in the location of the control points across all the images.

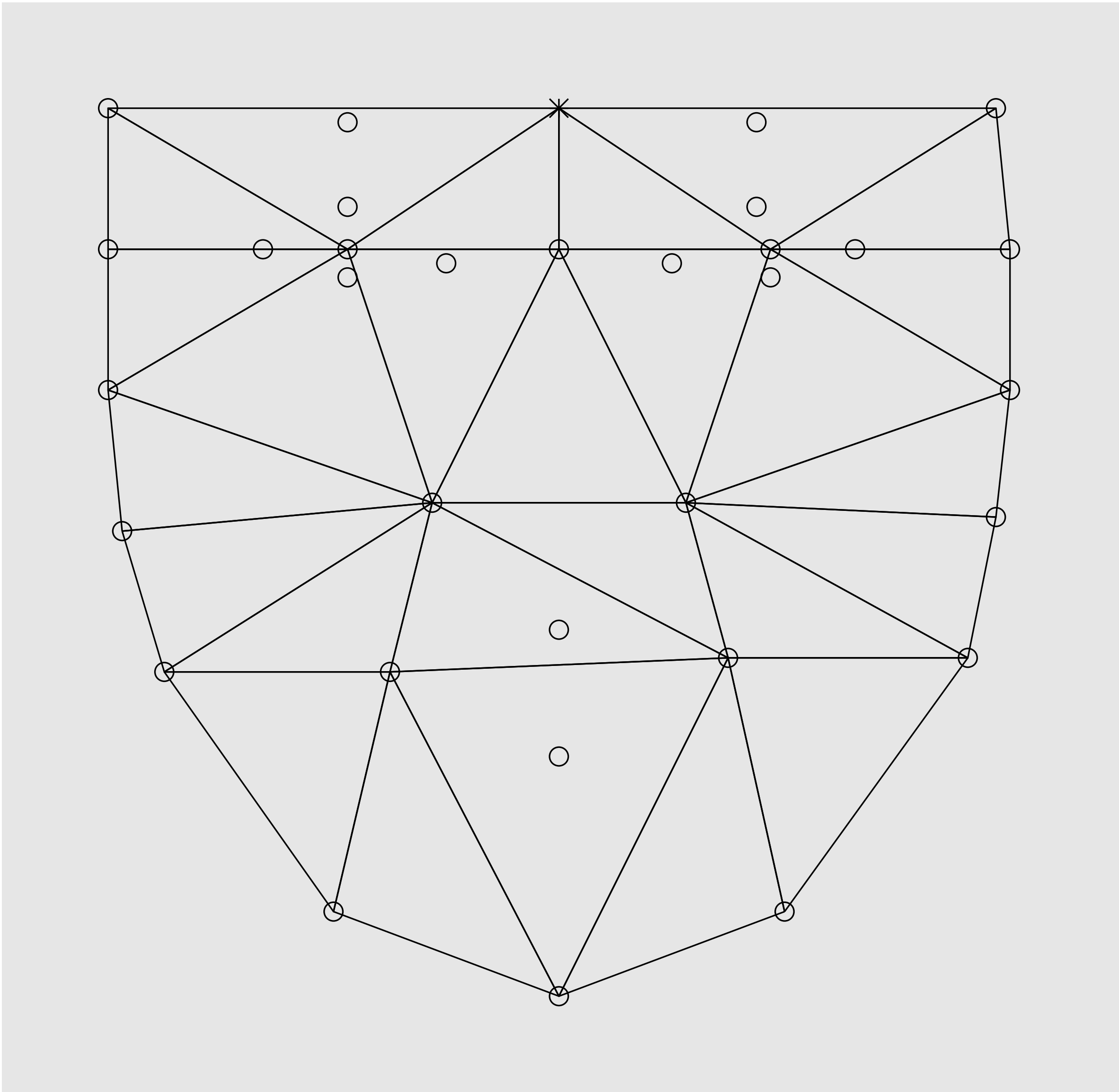

Figure 2. The triangulation procedure produces a set of small triangular regions that are utilized to derive textures by warping the individual images to the mean shape. Note that this triangulation procedure does not utilize all the control points. The * point, not being a control point, actually is the average of the upper-left and upper-right control points.

Figure 3. Some example textures from the training set. Textures are computed by warping the individual images to the mean shape using affine transformations over the small triangular regions derived by the triangulation procedure.

Figure 4. Some example shape images from the training set. Shape images undergo the same alignment procedure as the shapes do, but preserve the intensity information within the contours of the faces only.

Figure 5. Left: the mean face of the shape images. Right: the mean face of the textures. Note that the average image of the shape images has a blurred contour, while the one of the textures has a sharp contour.

Figure 6. Some example masked images from the training set. 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.

Figure 7. Some example FERET images used in our experiments. Since the images are acquired during different photo sessions, both the lighting conditions and the facial expressions may vary.

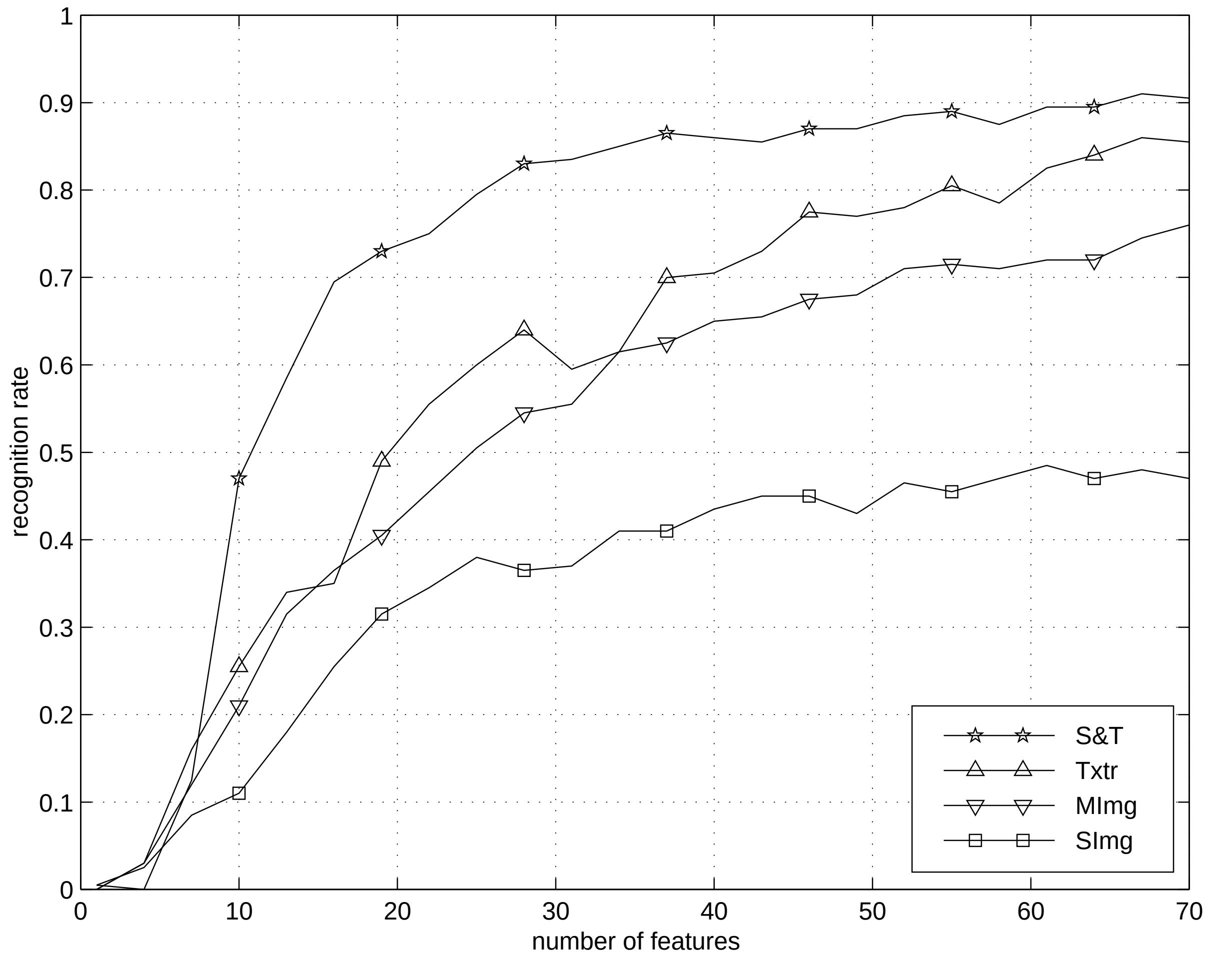

Figure 8. Comparative recognition performance (M2 method) using the integrated shape and texture features (S&T), the textures (Txtr), the masked images (MImg), and the shape images (SImg), respectively. The curves show that the integrated shape and texture features carry the most discriminating information followed in order by the textures, the masked images, and the shape images.

Figure 9. The first 30 basis vectors of the shape images. The absolute values on the contours are relatively large, indicating that these contours are used as important discriminative features.

Figure 10. The first 30 basis vectors of the textures. The textures tend to pick internal face features as important discrimination information.

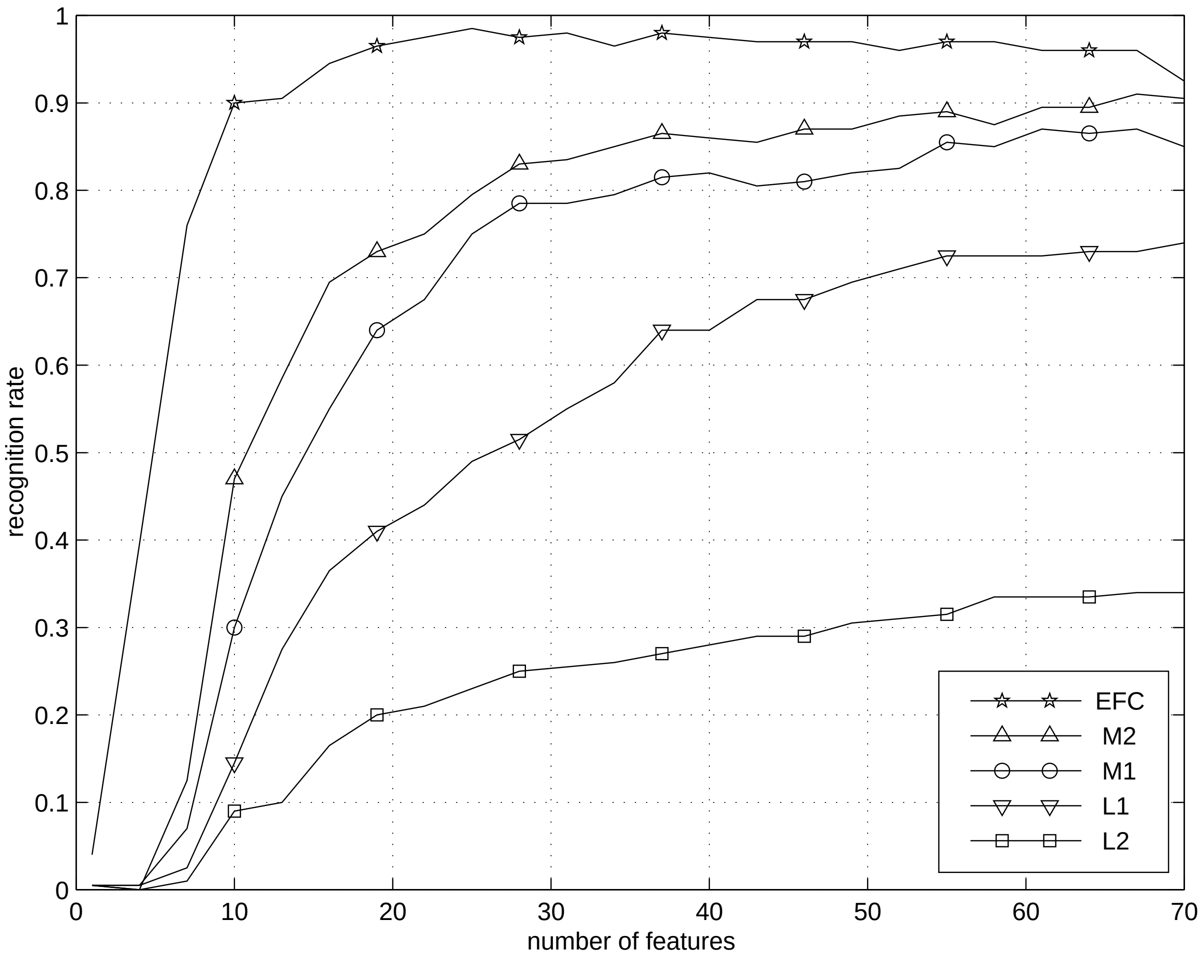

Figure 11. Comparative recognition performance of the new EFC method, the Mahalanobis distance classifier with a pooled within-class covariance matrix (M2), the Mahalanobis distance classifier with a common covariance matrix for all classes (M1), and the Eigenfaces method using L_1 distance measure (L1) and L_2 distance measure (L2), respectively. The input information for all the above methods is the integrated shape and texture features. The curves show that the new coding and face recognition method, EFC, achieves the best classification accuracy.

Chengjun Liu received the Ph.D. from George Mason University in 1999, and he is presently an Assistant Professor of Mathematics and Computer Science at University of Missouri - St. Louis. His research interests are in Computer Vision, Pattern Recognition, Image Processing, Evolutionary Computation, and Neural Computation. His recent research has been concerned with the development of novel and robust models for image retrieval and face recognition based upon statistical and machine learning concepts. The class of new models includes the Probabilistic Reasoning Models (PRM), which combines the optimal representation criterion (PCA) with the optimal classification criterion (Bayes classifier), the Enhanced FLD (Fisher Linear Discriminant) Models (EFM), which improves upon the generalization ability of FLD and the Enhanced Independent Component Analysis (EICA), both using sensitivity analysis on the spectral range of eigenvalues. He has also pursued the development of novel evolutionary methods leading to the development of the Evolutionary Pursuit (EP) method for pattern recognition in general, and face recognition in particular.

Harry Wechsler received the Ph.D. in Computer Science from the University of California, Irvine, in 1975, and he is presently Professor of Computer Science at George Mason University. His research, in the field of intelligent systems, has been in the areas of PERCEPTION: Computer Vision (CV), Automatic Target Recognition (ATR), Signal and Image Processing (SIP), MACHINE INTELLIGENCE: Pattern Recognition (PR), Neural Networks (NN), and Data Mining, EVOLUTIONARY COMPUTATION: Genetic Algorithms (GAs) and Animats, and HUMAN-COMPUTER INTELLIGENT INTERACTION (HCII) : Face and Hand Gesture Recognition, Biometrics, Video Tracking and Surveillance, and Interpretation of Human Activity. He was Director for the NATO Advanced Study Institutes (ASI) on ”Active Perception and Robot Vision” (Maratea, Italy, 1989), ”From Statistics to Neural Networks” (Les Arcs, France, 1993) and ”Face Recognition: From Theory to Applications” (Stirling, UK, 1997), and he has served as co-Chair for the International Conference on Pattern Recognition held in Vienna, Austria, in 1996. He authored over 200 scientific papers, his book ”Computational Vision” was published by Academic Press in 1990, and he was the editor for ”Neural Networks for Perception” (Vol 1 & 2), published by Academic Press in 1991. He was elected as an IEEE Fellow in 1992 and as an Int. Association of

Pattern Recognition (IAPR) Fellow in 1998.