

Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition

Chengjun Liu, *Member, IEEE*, and Harry Wechsler, *Fellow, IEEE*

Abstract—This paper introduces a novel Gabor–Fisher Classifier (GFC) for face recognition. The GFC method, which is robust to changes in illumination and facial expression, applies the Enhanced Fisher linear discriminant Model (EFM) to an augmented Gabor feature vector derived from the Gabor wavelet representation of face images. The novelty of this paper comes from 1) the derivation of an augmented Gabor feature vector, whose dimensionality is further reduced using the EFM by considering both data compression and recognition (generalization) performance; 2) the development of a Gabor–Fisher classifier for multi-class problems; and 3) extensive performance evaluation studies. In particular, we performed comparative studies of different similarity measures applied to various classifiers. We also performed comparative experimental studies of various face recognition schemes, including our novel GFC method, the Gabor wavelet method, the Eigenfaces method, the Fisherfaces method, the EFM method, the combination of Gabor and the Eigenfaces method, and the combination of Gabor and the Fisherfaces method. The feasibility of the new GFC method has been successfully tested on face recognition using 600 FERET frontal face images corresponding to 200 subjects, which were acquired under variable illumination and facial expressions. The novel GFC method achieves 100% accuracy on face recognition using only 62 features.

Index Terms—Eigenfaces, enhanced Fisher linear discriminant model (EFM), face recognition, Fisher linear discriminant (FLD), Gabor–Fisher classifier (GFC), Gabor wavelets.

I. INTRODUCTION

AMONG the most challenging tasks for visual form (“shape”) analysis and object recognition are understanding how people process and recognize each other’s face, and the development of corresponding computational models for automated face recognition. Face recognition is largely motivated by the need for surveillance and security, telecommunication and digital libraries, human-computer intelligent interaction, and smart environments [4], [8], [30], [33].

A good face recognition methodology should consider representation as well as classification issues, and a good representation method should require minimum manual annotations. The Gabor wavelets, whose kernels are similar to

the two-dimensional (2-D) receptive field profiles of the mammalian cortical simple cells, exhibit desirable characteristics of spatial locality and orientation selectivity. The biological relevance and computational properties of Gabor wavelets for image analysis have been described in [5], [6], [18], and [27]. The Gabor wavelet representation facilitates recognition without correspondence (hence, no need for manual annotations) because it captures the local structure corresponding to spatial frequency (scale), spatial localization, and orientation selectivity [34]. As a result, the Gabor wavelet representation of face images should be robust to variations due to illumination and facial expression changes [29], [32], [34].

This paper introduces a novel Gabor–Fisher Classifier (GFC) method for face recognition. The GFC method, which is robust to illumination and facial expression variability, applies the Enhanced Fisher linear discriminant Model (EFM) [23] to an augmented Gabor feature vector derived from the Gabor wavelet representation of face images. To encompass all the features produced by the different Gabor kernels one concatenates the resulting Gabor wavelet features to derive an augmented Gabor feature vector. The dimensionality of the Gabor vector space is then reduced under the eigenvalue selectivity constraint of the EFM method to derive a low-dimensional feature representation with enhanced discrimination power. The feasibility of the new GFC method has been successfully tested on face recognition using a data set from the FERET database, which is a standard testbed for face recognition technologies [31]. Specifically we used 600 FERET frontal face images corresponding to 200 subjects, which were acquired using variable illumination and facial expressions. The effectiveness of the GFC method is shown in terms of both absolute performance indices and comparative performance against some popular face recognition schemes such as the Gabor wavelet method [10], the Eigenfaces method [36], the Fisherfaces method [1], the EFM method [23], the combination of Gabor and the Eigenfaces method, and the combination of Gabor and the Fisherfaces method. In particular, the novel GFC method achieves 100% recognition accuracy using only 62 features.

II. BACKGROUND

Face recognition depends heavily on the particular choice of features used by the classifier [22], [23]. One usually starts with a given set of features and then attempts to derive an optimal subset (under some criteria) of features leading to high classification performance with the expectation that similar performance can also

Manuscript received April 12, 2001; revised November 26, 2001. C. Liu was supported in part by the New Jersey Institute of Technology under SBR Grant 421270. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Christine Guillemot.

C. Liu is with the Department of Computer Science, New Jersey Institute of Technology, Newark, NJ 07102 USA (e-mail: liu@cs.njit.edu).

H. Wechsler is with the Department of Computer Science, George Mason University, Fairfax, VA 22030 USA (e-mail: wechsler@cs.gmu.edu).

Publisher Item Identifier S 1057-7149(02)03548-0.

be displayed on future trials using novel (unseen) test data. Principal component analysis (PCA) is a popular technique used to derive a starting set of features for both face representation and recognition. Kirby and Sirovich [19] showed that any particular face can be (i) economically represented along the eigenpictures coordinate space, and (ii) approximately reconstructed using just a small collection of eigenpictures and their corresponding projections ('coefficients'). Applying PCA technique to face recognition, Turk and Pentland [36] developed a well-known Eigenfaces method. The Eigenfaces method, however, does not consider the classification aspect, as it is based on the optimal representation criterion (PCA) in the sense of mean-square error. To improve the PCA standalone classification performance, one needs to combine further this optimal representation criterion with some discrimination criterion.

One widely used discrimination criterion in the face recognition community is the Fisher linear discriminant (FLD, a.k.a. linear discriminant analysis, or LDA) [16], which defines a projection that makes the within-class scatter small and the between-class scatter large. As a result, FLD derives compact and well-separated clusters. FLD is behind several face recognition methods [1], [12], [23], [35]. As the original image space is high dimensional, most of these methods apply PCA first for dimensionality reduction, as it is the case with the Fisherfaces method due to Belhumeur *et al.* [1]. Subsequent FLD transformation is used then to build the most discriminating features (MDF) space for classification [35]. The drawback of FLD is that it requires large training sample size for good generalization. For a face recognition problem, however, usually there are a large number of faces (classes), but only a few training examples per face. One possible remedy for this drawback is to artificially generate additional data and thus increase the sample size [12]. Yet another remedy is to improve FLD's generalization performance by balancing the need for adequate signal representation and subsequent classification performance using sensitivity analysis on the spectral range of the within-class eigenvalues [23].

Gabor wavelets model quite well the receptive field profiles of cortical simple cells [15]. The Gabor wavelet representation, therefore, captures salient visual properties such as spatial localization, orientation selectivity, spatial frequency characteristic. Lades *et al.* [21] pioneered the use of Gabor wavelets for face recognition using the Dynamic Link Architecture (DLA) framework. The DLA starts by computing the Gabor jets, and then it performs a flexible template comparison between the resulting image decompositions using graph-matching. Wiskott *et al.* [38] have expanded on DLA when they developed a Gabor wavelet based elastic bunch graph matching method to label and recognize human faces. Based on the 2-D Gabor wavelet representation and the labeled elastic graph matching, Lyons *et al.* [25], [26] proposed an algorithm for two-class categorization of gender, race, and facial expression. The algorithm includes two steps: registration of a grid with the face using either labeled elastic graph matching [21], [38] or manual annotation of 34 points on every face image [26]; and categorization based on the features extracted at grid points using linear discriminant analysis (LDA). Donato *et al.* [10] have recently shown through experiments that the Gabor wavelet representation gives better performance than other techniques for classifying facial actions.

A. Gabor Wavelets

Gabor wavelets were introduced to image analysis due to their biological relevance and computational properties [6], [7], [18], [27]. The Gabor wavelets, whose kernels are similar to the 2-D receptive field profiles of the mammalian cortical simple cells, exhibit desirable characteristics of spatial locality and orientation selectivity, and are optimally localized in the space and frequency domains.

The Gabor wavelets (kernels, filters) can be defined as follows [5], [27], [21]:

$$\psi_{\mu,\nu}(z) = \frac{\|k_{\mu,\nu}\|^2}{\sigma^2} e^{(-\|k_{\mu,\nu}\|^2 \|z\|^2 / 2\sigma^2)} \left[e^{ik_{\mu,\nu}z} - e^{-\sigma^2/2} \right] \quad (1)$$

where μ and ν define the orientation and scale of the Gabor kernels, $z = (x, y)$, $\|\cdot\|$ denotes the norm operator, and the wave vector $k_{\mu,\nu}$ is defined as follows:

$$k_{\mu,\nu} = k_\nu e^{i\phi_\mu} \quad (2)$$

where $k_\nu = k_{\max}/f^\nu$ and $\phi_\mu = \pi\mu/8$. k_{\max} is the maximum frequency, and f is the spacing factor between kernels in the frequency domain [21].

The Gabor kernels in (1) are all self-similar since they can be generated from one filter, the mother wavelet, by scaling and rotation via the wave vector $k_{\mu,\nu}$. Each kernel is a product of a Gaussian envelope and a complex plane wave, while the first term in the square brackets in (1) determines the oscillatory part of the kernel and the second term compensates for the DC value. The effect of the DC term becomes negligible when the parameter σ , which determines the ratio of the Gaussian window width to wavelength, has sufficiently large values.

In most cases one would use Gabor wavelets of five different scales, $\nu \in \{0, \dots, 4\}$, and eight orientations, $\mu \in \{0, \dots, 7\}$ [13], [18], [3]. Fig. 1 shows the real part of the Gabor kernels at five scales and eight orientations and their magnitude, with the following parameters: $\sigma = 2\pi$, $k_{\max} = \pi/2$, and $f = \sqrt{2}$. The kernels exhibit desirable characteristics of spatial frequency, spatial locality, and orientation selectivity.

B. Gabor Feature Representation

The Gabor wavelet representation of an image is the convolution of the image with a family of Gabor kernels as defined by (1). Let $I(x, y)$ be the gray level distribution of an image, the convolution of image I and a Gabor kernel $\psi_{\mu,\nu}$ is defined as follows:

$$O_{\mu,\nu}(z) = I(z) * \psi_{\mu,\nu}(z) \quad (3)$$

where $z = (x, y)$, $*$ denotes the convolution operator, and $O_{\mu,\nu}(z)$ is the convolution result corresponding to the Gabor kernel at orientation μ and scale ν . Therefore, the set $\mathcal{S} = \{O_{\mu,\nu}(z) : \mu \in \{0, \dots, 7\}, \nu \in \{0, \dots, 4\}\}$ forms the Gabor wavelet representation of the image $I(z)$.

Applying the convolution theorem, we can derive each $O_{\mu,\nu}(z)$ from (3) via the Fast Fourier Transform (FFT)

$$\mathfrak{F}\{O_{\mu,\nu}(z)\} = \mathfrak{F}\{I(z)\} \mathfrak{F}\{\psi_{\mu,\nu}(z)\} \quad (4)$$

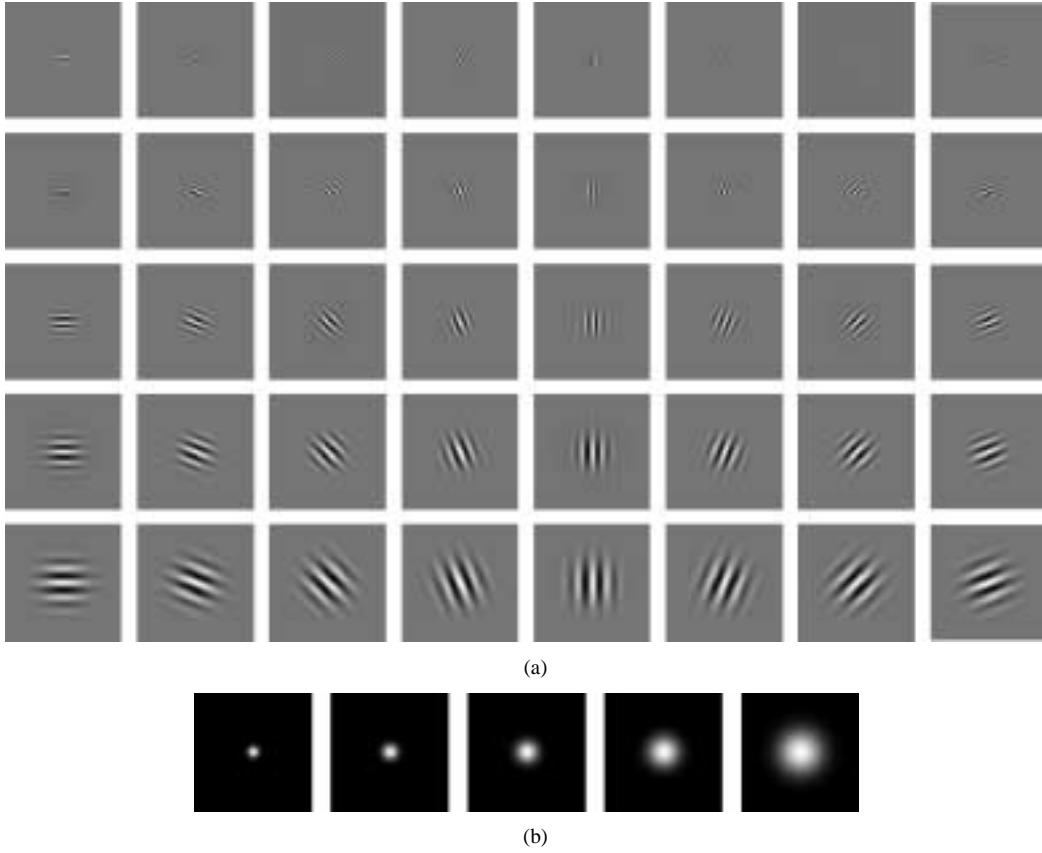


Fig. 1. Gabor Wavelets. (a) The real part of the Gabor kernels at five scales and eight orientations with the following parameters: $\sigma = 2\pi$, $k_{\max} = \pi/2$, and $f = \sqrt{2}$. (b) The magnitude of the Gabor kernels at five different scales. The kernels exhibit desirable characteristics of spatial frequency, spatial locality, and orientation selectivity.

and

$$O_{\mu,\nu}(z) = \Im^{-1}\{\Im\{I(z)\}\Im\{\psi_{\mu,\nu}(z)\}\} \quad (5)$$

where \Im and \Im^{-1} denote the Fourier and inverse Fourier transform, respectively.

Fig. 2 shows the Gabor wavelet representation (the real part and the magnitude) of a sample image. These representation results display scale, locality, and orientation properties corresponding to those displayed by the Gabor wavelets in Fig. 1. To encompass different spatial frequencies (scales), spatial localities, and orientation selectivities, we concatenate all these representation results and derive an augmented feature vector \mathcal{X} . Before the concatenation, we first downsample each $O_{\mu,\nu}(z)$ by a factor ρ to reduce the space dimension, and normalize it to zero mean and unit variance. We then construct a vector out of the $O_{\mu,\nu}(z)$ by concatenating its rows (or columns). Now, let $O_{\mu,\nu}^{(\rho)}$ denote the normalized vector constructed from $O_{\mu,\nu}(z)$ (downsampled by ρ and normalized to zero mean and unit variance), the augmented Gabor feature vector $\mathcal{X}^{(\rho)}$ is then defined as follows:

$$\mathcal{X}^{(\rho)} = \left(O_{0,0}^{(\rho)t} \quad O_{0,1}^{(\rho)t} \quad \dots \quad O_{4,7}^{(\rho)t} \right)^t \quad (6)$$

where t is the transpose operator. The augmented Gabor feature vector thus encompasses all the elements (downsampled and normalized) of the Gabor wavelet representation set, $\mathcal{S} = \{O_{\mu,\nu}(z) : \mu \in \{0, \dots, 7\}, \nu \in \{0, \dots, 4\}\}$, as important

discriminating information. Fig. 3 shows (in image form rather than in vector form) an example of the augmented Gabor feature vector, where the downsampling factor is 64, i.e., $\rho = 64$.

III. GABOR-FISHER CLASSIFIER

We describe in this section our novel Gabor-Fisher Classifier (GFC) method which applies the Enhanced Fisher linear discriminant Model (EFM) [23] to the augmented Gabor feature vector $\mathcal{X}^{(\rho)}$ derived in Section II.B. The dimensionality of the resulting vector space is reduced, using the eigenvalue selectivity constraint of the EFM method, in order to derive low-dimensional features with enhanced discrimination power.

A. Dimensionality Reduction and Discriminant Analysis

The augmented Gabor feature vector introduced in Section II-B resides in a space of very high dimensionality: $\mathcal{X}^{(\rho)} \in \mathbb{R}^N$, where N is the dimensionality of the vector space. 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” [11]. Low-dimensional representations are also important when one considers the intrinsic computational aspect. Principal component analysis, or PCA [17], [9], whose primary goal is to project the high dimensional visual

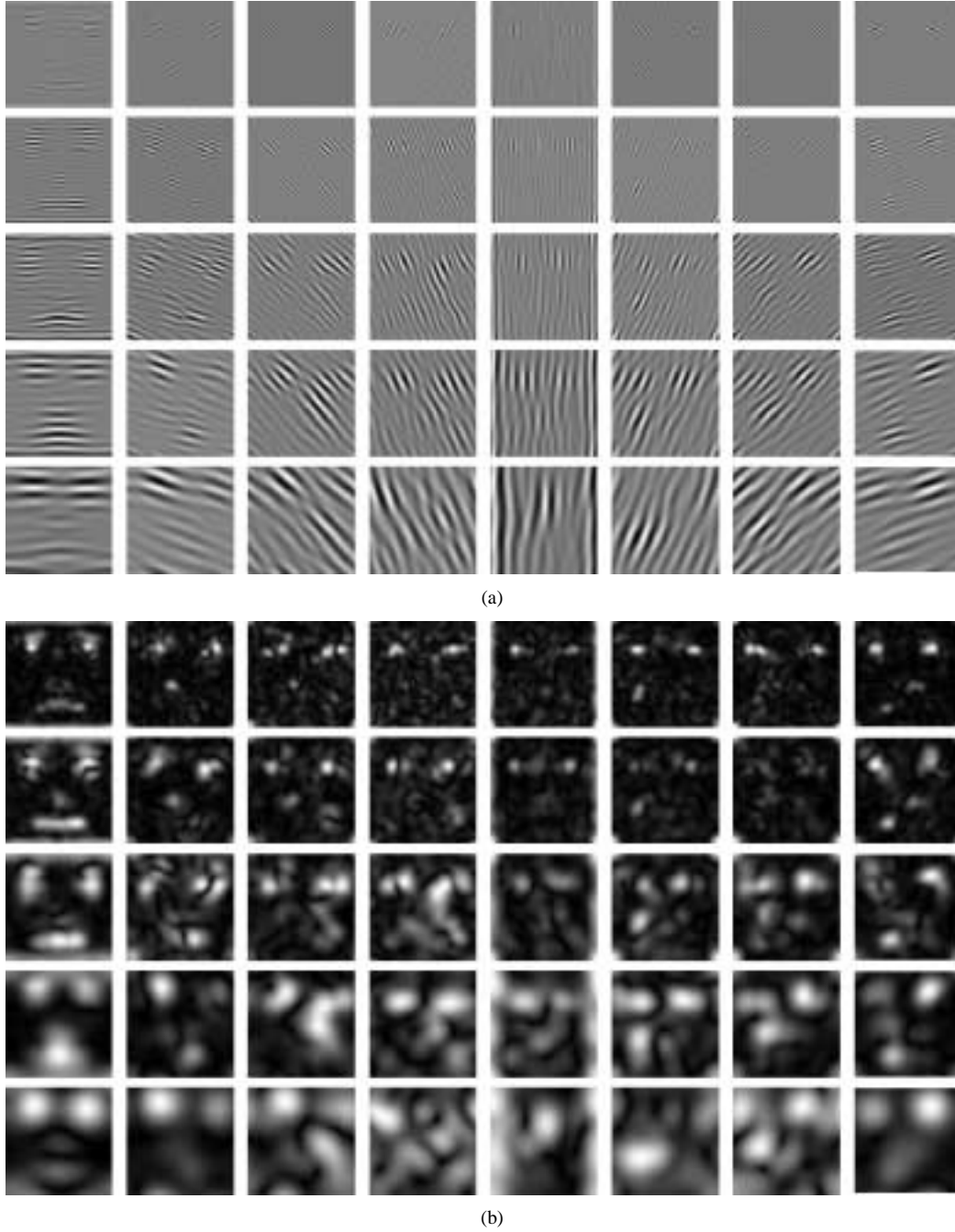


Fig. 2. Gabor wavelet representation (the real part and the magnitude) of a sample image. (a) The real part of the representation and (b) the magnitude of the representation.

stimuli (face images) into a lower dimensional space, is the optimal method for dimensionality reduction in the sense of mean-square error.

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 $\Sigma_{\mathcal{X}^{(\rho)}} \in \mathbb{R}^{N \times N}$ define the covariance matrix of the augmented feature vector $\mathcal{X}^{(\rho)}$

$$\Sigma_{\mathcal{X}^{(\rho)}} = \mathcal{E} \left\{ \left[\mathcal{X}^{(\rho)} - \mathcal{E} \left(\mathcal{X}^{(\rho)} \right) \right] \left[\mathcal{X}^{(\rho)} - \mathcal{E} \left(\mathcal{X}^{(\rho)} \right) \right]^t \right\} \quad (7)$$

where $\mathcal{E}(\cdot)$ is the expectation operator. The PCA of a random vector $\mathcal{X}^{(\rho)}$ factorizes its covariance matrix $\Sigma_{\mathcal{X}^{(\rho)}}$ into the fol-

lowing form:

$$\Sigma_{\mathcal{X}^{(\rho)}} = \Phi \Lambda \Phi^t \text{ with } \Phi = [\phi_1 \phi_2 \dots \phi_N], \quad \Lambda = \text{diag} \{ \lambda_1, \lambda_2, \dots, \lambda_N \} \quad (8)$$

where $\Phi \in \mathbb{R}^{N \times N}$ is an orthogonal eigenvector matrix and $\Lambda \in \mathbb{R}^{N \times N}$ a diagonal eigenvalue matrix with diagonal elements in decreasing order ($\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$).

An important property of PCA is its optimal signal reconstruction in the sense of minimum mean-square error when only a subset of principal components is used to represent the original signal. Following this property, an immediate application

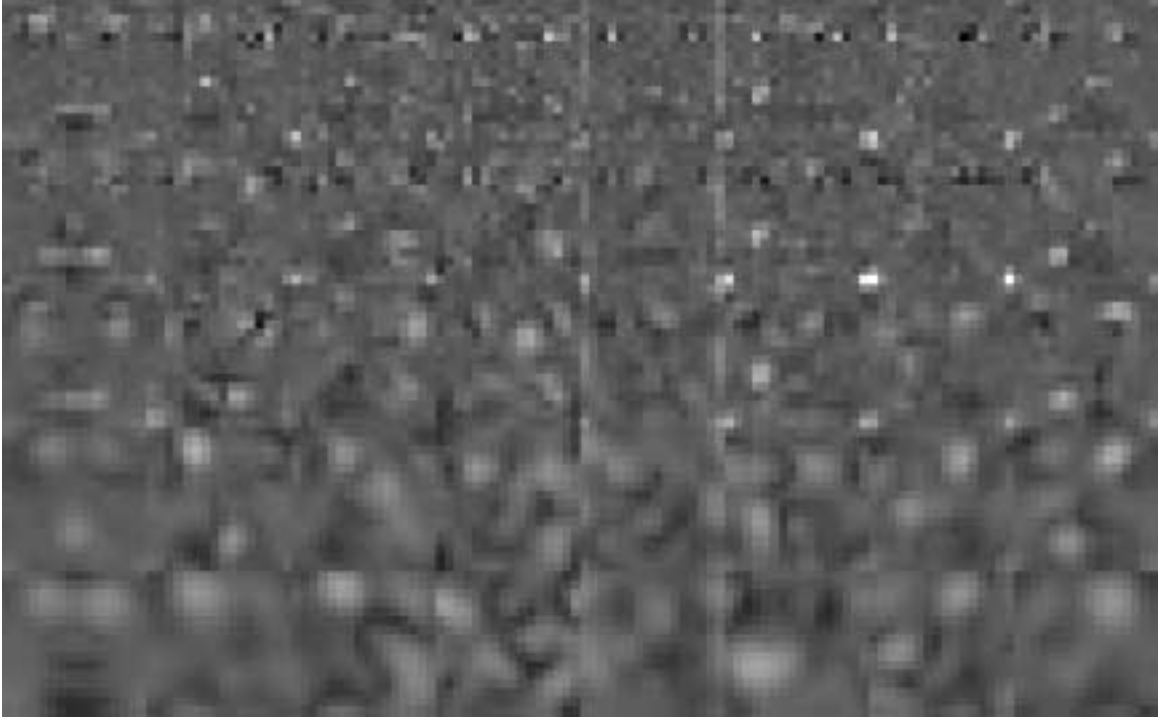


Fig. 3. An example of the augmented Gabor feature vector (in image form rather than in vector form), where the downsampling factor is 64, i.e., $\rho = 64$. In our experiments, the dimensionality of this vector space is as high as 10 240.

of PCA is dimensionality reduction

$$\mathcal{Y}^{(\rho)} = P^t \mathcal{X}^{(\rho)} \quad (9)$$

where $P = [\phi_1 \phi_2 \dots \phi_m]$, $m < N$ and $P \in \mathbb{R}^{N \times m}$. The lower dimensional vector $\mathcal{Y}^{(\rho)} \in \mathbb{R}^m$ captures the most expressive features of the original data $\mathcal{X}^{(\rho)}$.

However, one should be aware that the PCA driven coding schemes are optimal and useful only with respect to data compression and decorrelation of low (second) order statistics. PCA does not take into account the recognition (discrimination) aspect and one should thus not expect optimal performance for tasks such as face recognition when using such PCA-like encoding schemes. To address this obvious shortcoming, one has to reformulate the original problem as one where the search is still for low-dimensional patterns but is now also subject to seeking a high discrimination index, characteristic of separable low-dimensional patterns. One solution that has been proposed to solve this new problem is to use the Fisher linear discriminant (FLD) [16] for the very purpose of achieving high separability between the different patterns in whose classification one is interested. Characteristic of this approach are recent schemes such as the most discriminating features (MDF) method [35] and the Fisherfaces method [1].

FLD is a popular discriminant criterion that measures the between-class scatter normalized by the within-class scatter [17]. Let $\omega_1, \omega_2, \dots, \omega_L$ and N_1, N_2, \dots, N_L denote the classes and the number of images within each class, respectively. Let M_1, M_2, \dots, M_L and M be the means of the classes and the grand mean. The within- and between-class scatter matrices,

Σ_w and Σ_b , are defined as follows:

$$\Sigma_w = \sum_{i=1}^L P(\omega_i) \mathcal{E} \left\{ (\mathcal{Y}^{(\rho)} - M_i)(\mathcal{Y}^{(\rho)} - M_i)^t | \omega_i \right\} \quad (10)$$

and

$$\Sigma_b = \sum_{i=1}^L P(\omega_i) (M_i - M)(M_i - M)^t \quad (11)$$

where $P(\omega_i)$ is a *a priori* probability, $\Sigma_w, \Sigma_b \in \mathbb{R}^{m \times m}$, and L denotes the number of classes.

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 $\Sigma_w^{-1} \Sigma_b$ [35]

$$\Sigma_w^{-1} \Sigma_b \Psi = \Psi \Delta \quad (12)$$

where $\Psi, \Delta \in \mathbb{R}^{m \times m}$ are the eigenvector and eigenvalue matrices of $\Sigma_w^{-1} \Sigma_b$, respectively.

One drawback of FLD is that it requires large training sample size for good generalization. When such requirement is not met, FLD overfits to the training data and thus generalizes poorly to the novel testing data [23].

B. Enhanced Fisher Linear Discriminant Model

The Enhanced Fisher linear discriminant Model (EFM) improves the generalization capability of FLD by decomposing the FLD procedure into a simultaneous diagonalization of the two within- and between-class scatter matrices [23]. The simultaneous diagonalization is stepwisely equivalent to two operations as pointed out by Fukunaga [17]: whitening the within-class scatter matrix and applying PCA on the between-class

scatter matrix using the transformed data. The stepwise operation shows that during whitening the eigenvalues of the within-class scatter matrix appear in the denominator. As the small (trailing) eigenvalues tend to capture noise [23], they cause the whitening step to fit for misleading variations and thus generalize poorly when exposed to new data. To achieve enhanced performance EFM preserves 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, i.e., representational adequacy, and the requirement that the eigenvalues of the within-class scatter matrix (in the reduced PCA space) are not too small, i.e., better generalization.

The choice of the range of principal components (m) for dimensionality reduction [see (9)] takes into account both the spectral energy and the magnitude requirements. The eigenvalue spectrum of the covariance matrix [see (8)] provides a good indicator for meeting the energy criterion; one needs then to derive the eigenvalue spectrum of the within-class scatter matrix in the reduced PCA space to facilitate the choice of the range of principal components so that the magnitude requirement is met. Toward that end, one carries out the stepwise FLD process described earlier. 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:

$$\Sigma_w \Xi = \Xi \Gamma \quad \text{and} \quad \Xi^t \Xi = I \quad (13)$$

$$\Gamma^{-1/2} \Xi^t \Sigma_w \Xi \Gamma^{-1/2} = I \quad (14)$$

where $\Xi, \Gamma \in \mathbb{R}^{m \times m}$ are the eigenvector and the diagonal eigenvalue matrices of Σ_w , respectively.

The eigenvalue spectrum of the within-class scatter matrix in the reduced PCA space can be derived by (13), and different spectra are obtained corresponding to different number of principal components utilized (see (9) and (10)). Now one has to simultaneously optimize the behavior of the trailing eigenvalues in the reduced PCA space (13) with the energy criteria for the original image space (8).

After the feature vector $\mathcal{Y}^{(\rho)}$ (9) is derived, EFM first diagonalizes the within-class scatter matrix Σ_w using (13) and (14). Note that now Ξ and Γ are the eigenvector and the eigenvalue matrices corresponding to the feature vector $\mathcal{Y}^{(\rho)}$. EFM proceeds then to compute the between-class scatter matrix as follows:

$$\Gamma^{-1/2} \Xi^t \Sigma_b \Xi \Gamma^{-1/2} = K_b. \quad (15)$$

Diagonalize now the new between-class scatter matrix K_b

$$K_b \Theta = \Theta \Upsilon \quad \text{and} \quad \Theta^t \Theta = I \quad (16)$$

where $\Theta, \Upsilon \in \mathbb{R}^{m \times m}$ are the eigenvector and the diagonal eigenvalue matrices of K_b , respectively.

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

$$T = \Xi \Gamma^{-1/2} \Theta \quad (17)$$

C. Similarity Measures and Classification Rule for Gabor Feature Based Classification

The GFC applies the EFM method on the (lower dimensional) augmented Gabor feature vector $\mathcal{Y}^{(\rho)}$ derived by (9). When an image is presented to the GFC classifier, the augmented Gabor feature vector of the image is first calculated as detailed in Section II-B, and the lower dimensional feature, $\mathcal{Y}^{(\rho)}$, is derived using (9). The dimensionality of the lower dimensional feature space is determined by the EFM method, which derives further the overall transformation matrix, T , as defined by (17). The new feature vector, $\mathcal{U}^{(\rho)}$, of the image is defined as follows:

$$\mathcal{U}^{(\rho)} = T^t \mathcal{Y}^{(\rho)}. \quad (18)$$

Let $\mathcal{M}_k^0, k = 1, 2, \dots, L$, be the mean of the training samples for class ω_k after the EFM transformation. The GFC method applies, then, the nearest neighbor (to the mean) rule for classification using some similarity (distance) measure δ

$$\delta(\mathcal{U}^{(\rho)}, \mathcal{M}_k^0) = \min_j \delta(\mathcal{U}^{(\rho)}, \mathcal{M}_j^0) \longrightarrow \mathcal{U}^{(\rho)} \in \omega_k. \quad (19)$$

The image feature vector, $\mathcal{U}^{(\rho)}$, is classified as belonging to the class of the closest mean, \mathcal{M}_k^0 , using the similarity measure δ .

The similarity measures used in our experiments to evaluate the efficiency of different representation and recognition methods include L_1 distance measure, δ_{L_1} , L_2 distance measure, δ_{L_2} , Mahalanobis distance measure, δ_{Md} , and cosine similarity measure, δ_{\cos} , which are defined as follows:

$$\delta_{L_1}(\mathcal{X}, \mathcal{Y}) = \sum_i |\mathcal{X}_i - \mathcal{Y}_i| \quad (20)$$

$$\delta_{L_2}(\mathcal{X}, \mathcal{Y}) = (\mathcal{X} - \mathcal{Y})^t (\mathcal{X} - \mathcal{Y}) \quad (21)$$

$$\delta_{Md}(\mathcal{X}, \mathcal{Y}) = (\mathcal{X} - \mathcal{Y})^t \Sigma^{-1} (\mathcal{X} - \mathcal{Y}) \quad (22)$$

$$\delta_{\cos}(\mathcal{X}, \mathcal{Y}) = \frac{-\mathcal{X}^t \mathcal{Y}}{\|\mathcal{X}\| \|\mathcal{Y}\|} \quad (23)$$

where Σ is the covariance matrix, and $\|\cdot\|$ denotes the norm operator. Note that the cosine similarity measure includes a minus sign in (23), because the nearest neighbor (to the mean) rule of (19) applies minimum (distance) measure rather than maximum similarity measure.

IV. EXPERIMENTS

We assessed the feasibility and performance of our novel GFC on the face recognition task, using a data set from the FERET database, which is a standard testbed for face recognition technologies [31], [2]. Specifically we used 600 FERET frontal face images corresponding to 200 subjects, which were acquired under variable illumination and facial expressions. Comparative performance is carried out against some popular face recognition schemes such as the Gabor wavelet method [10], the Eigenfaces method [36], the Fisherfaces method [1], the EFM method [23], the combination of Gabor and the Eigenfaces method, and the combination of Gabor and the Fisherfaces method.

The FERET database used for evaluating face recognition algorithms displays diversity across gender, ethnicity, and age. The image sets were acquired without any restrictions imposed



Fig. 4. Example FERET images used in our experiments (cropped to the size of 128×128 to extract the facial region). Note that the images are acquired during different photo sessions, they display both different lighting conditions and facial expressions. Two images are randomly chosen from the three images available for each subject for training, while the remaining image (unseen during training) is used for testing. In particular, the above figure shows in the top two rows the examples of training images used in our experiments, and in the bottom row the examples of test images.

on facial expression and with at least two frontal images shot at different times during the same photo session. 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. First, the centers of the eyes of an image are manually detected, then rotation and scaling transformations align the centers of the eyes to predefined locations. Finally, the face image is cropped to the size of 128×128 to extract the facial region, which is further normalized to zero mean and unit variance. Fig. 4 shows some example images used in our experiments that are already cropped to the size of 128×128 . Note that as the images were acquired during different photo sessions, they display different illumination characteristics and facial expressions. As two images are randomly chosen for training, while the remaining image (unseen during training) is used for testing (see Fig. 4), the GFC has to cope with both illumination and facial expression variabilities.

For comparison purpose, we first implemented the Eigenfaces method [36], the Fisherfaces method [1], and the EFM method [23] and tested their performance using the original face images as shown in Fig. 4. The comparative face recognition performance of these three methods is shown in Fig. 5, and one can see from the figure that the EFM method performs better than the Fisherfaces method followed by the Eigenfaces method. Both the EFM method and the Fisherfaces method apply the L_2 distance measure, while the Eigenfaces method applies the Mahalanobis distance measure. For the Eigenfaces method, the Mahalanobis distance measure performs better than the L_1 distance measure, followed in order by the L_2 distance measure and the cosine similarity measure as shown in Fig. 6. The Mahalanobis distance measure performs better than the other similarity measures because the Mahalanobis distance measure counteracts the fact that L_1 and L_2 distance measures in the PCA space weight preferentially for low frequencies, and this is consistent with the results reported by Moghaddam and Pentland [28] and Sung and Poggio [20]. As the L_2 measure weights more the low frequencies than L_1

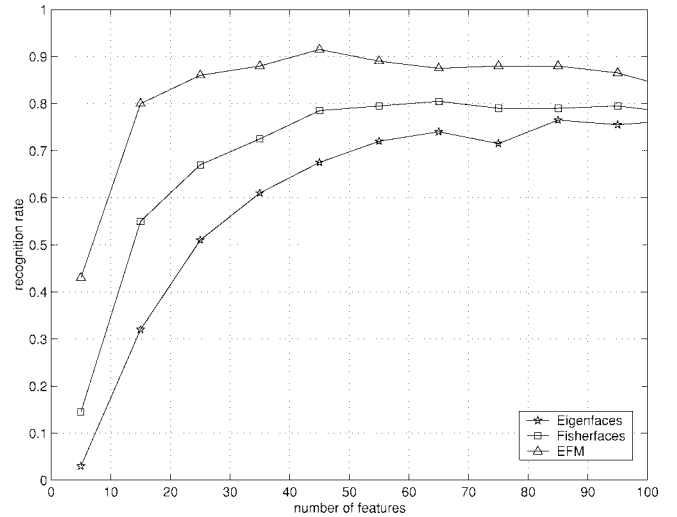


Fig. 5. Comparative face recognition performance of the eigenfaces method, the Fisherfaces method, and the EFM method using the original images.

does, the L_1 distance measure should perform better than the L_2 distance measure, a conjecture validated by our experiments as shown in Fig. 6. The cosine similarity measure does not compensate the low frequency preference, and it performs the worst among all the measures. Actually, the superiority of the cosine similarity measure to the others can be revealed only when the discriminating features (derived by the GFC method) rather than the expressive features (derived by the PCA) are used for classification [24].

The next series of experiments exploits the Gabor wavelet representation, $\mathcal{S} = \{O_{\mu,\nu}(z) : \mu \in \{0, \dots, 7\}, \nu \in \{0, \dots, 4\}\}$, derived in Section II-B, using the L_1 , L_2 and cosine similarity measures, respectively. (The Mahalanobis metric is not used here because it involves transformed data and covariance matrix suitable for PCA-like schemes. The L_1 , L_2 and cosine metrics are here compared at different downsampling rates without further data transformations.) For the first set of

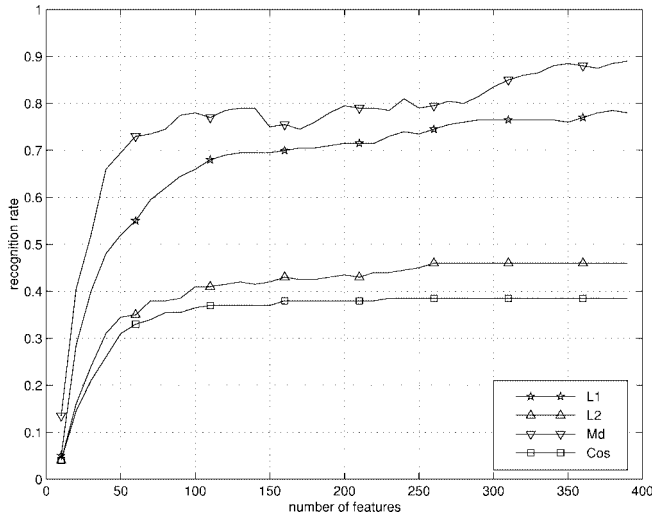


Fig. 6. Face recognition performance of the eigenfaces method using the original face images and the four different similarity measures: L1 (L_1 distance measure), L2 (L_2 distance measure), Md (Mahalanobis distance measure), and Cos (cosine similarity measure).

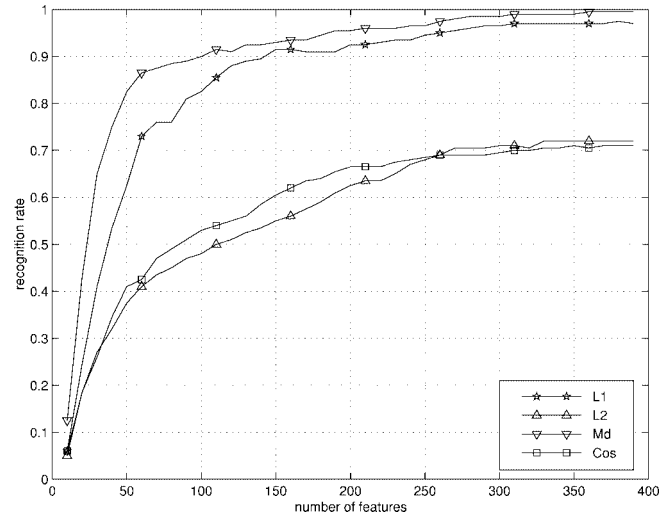


Fig. 7. Face recognition performance of PCA on the augmented Gabor feature vector $\mathcal{X}^{(\rho)}$, $\rho = 64$, using four different similarity measures: L1 (L_1 distance measure), L2 (L_2 distance measure), Md (Mahalanobis distance measure), and Cos (cosine similarity measure).

TABLE I

FACE RECOGNITION PERFORMANCE USING THE GABOR WAVELET REPRESENTATION WITH THE THREE DIFFERENT SIMILARITY MEASURES, RESPECTIVELY: L1 (L_1 DISTANCE MEASURE), L2 (L_2 DISTANCE MEASURE), AND COS (COSINE SIMILARITY MEASURE). $\mathcal{S}^{(16)}$ REPRESENTS THE GABOR REPRESENTATION DOWNSAMPLED BY A FACTOR OF 16 AND NORMALIZED TO UNIT LENGTH OF EACH $O_{\mu,\nu}(z)$, AS SUGGESTED BY [10]. $\mathcal{X}^{(4)}$, $\mathcal{X}^{(16)}$, AND $\mathcal{X}^{(64)}$ REPRESENT THE AUGMENTED GABOR FEATURE VECTOR $\mathcal{X}^{(\rho)}$ AS DEFINED BY (6) USING THREE DIFFERENT DOWNSAMPLING FACTORS, RESPECTIVELY: $\rho = 4, 16$, OR 64 .

measure \ representation	$\mathcal{S}^{(16)}$	$\mathcal{X}^{(4)}$	$\mathcal{X}^{(16)}$	$\mathcal{X}^{(64)}$
L1	76%	76.5%	76.5%	76.5%
L2	73.5%	72%	72%	72%
Cos	72%	70.5%	70.5%	70%

experiments, we downsampled the Gabor wavelet representation set, $\mathcal{S} = \{O_{\mu,\nu}(z) : \mu \in \{0, \dots, 7\}, \nu \in \{0, \dots, 4\}\}$, by a factor of 16 to reduce the dimensionality and normalized each $O_{\mu,\nu}(z)$ to unit length, as suggested by Donato *et al.* [10]. The face recognition performance using such a Gabor representation corresponding to different similarity measures is tabulated in Table I, which shows that the best performance is achieved using the L_1 similarity measure. Comparing Table I with Fig. 6, we found that 1) under the L_2 and cosine similarity measures, the Gabor features carry more discriminating information than the PCA features do, a finding consistent with that reported by Donato *et al.* [10] on facial action classification and 2) the performance with the three similarity measures, L_1 , L_2 and cosine, varies less drastically than that shown in Fig. 6. The second finding indicates Gabor representation is less likely affected by preferential low frequency weighting, which qualifies the Gabor representation as a discriminating representation method. We have also experimented on the augmented Gabor feature vector $\mathcal{X}^{(\rho)}$ as defined by (6) with three different downsampling factors, respectively: $\rho = 4$,

16, or 64. From the face recognition performance shown in Table I, we found that 1) the augmented Gabor feature vector $\mathcal{X}^{(\rho)}$ carries quite similar discriminating information to the one used by Donato *et al.* [10] and 2) the performance differences among using the three different downsampling factors are not significant. As a result, we choose the downsampling factor of 64 for the next series of experiments, since it reduces to a larger extent the dimensionality of the vector space than the other two factors do. (We experimented with other downsampling factors as well. When the downsampling factors are 256 and 1024, the performance is marginally less effective; when the downsampling factor is 4096, the recognition rate drops drastically.)

Even though the performance shown in Table I indicates that Gabor feature representation carries discriminating information, it is still not convenient to use such representation directly for classification, since the dimensionality of the augmented Gabor feature vector space is very high. To reduce the dimensionality of the vector space, we applied PCA on the augmented Gabor feature vector $\mathcal{X}^{(\rho)}$, where the downsampling factor ρ is set to be 64. Fig. 7 shows the face recognition performance of PCA using the augmented Gabor feature vector $\mathcal{X}^{(\rho)}$. Our results indicate that (i) the recognition performance improves by a large margin for all the similarity measures as compared with Fig. 6; and (ii) Mahalanobis and L_1 distance measures perform better than the other two similarity measures, which shows again that PCA derives features that preferentially weight low frequencies. Our last series of experiments, performed using the novel Gabor–Fisher Classifier (GFC) method described in this paper, show that the GFC derives discriminating Gabor features with low dimensionality and enhanced discrimination power. Fig. 8 shows comparative face recognition performance of the combination of Gabor and the Eigenfaces method, and the combination of Gabor and the Fisherfaces method, and the GFC method, using the augmented Gabor feature vector $\mathcal{X}^{(\rho)}$ downsampled by a factor of 64, i.e., $\rho = 64$. The GFC method performs better than both of the other two methods. In particular, GFC method achieves 100% correct recognition accuracy when using only 62

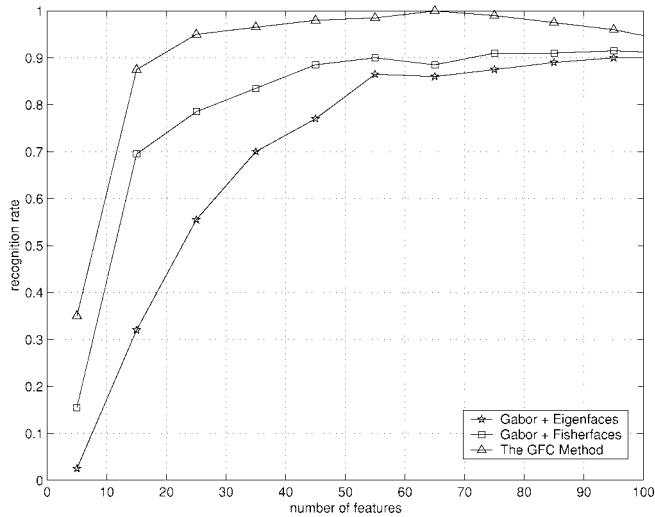


Fig. 8. Comparative face recognition performance of the combination of Gabor and the eigenfaces method, the combination of Gabor and the Fisherfaces method, and the GFC method, using the augmented Gabor feature vector $\mathcal{X}^{(\rho)}$ downsampled by a factor of 64, i.e., $\rho = 64$.

features (note that the curves in Fig. 8 were drawn with an interval resolution of 10 features, and it shows that 100% correct recognition rate happens when 65 features are used).

V. CONCLUSIONS

We have introduced in this paper a novel Gabor–Fisher classification method for face recognition. The GFC method, which is robust to variations in illumination and facial expression, applies the EFM method to an augmented Gabor feature vector derived from the Gabor wavelet representation of face images. The Gabor transformed face images yield features that display scale, locality, and orientation selectivity. The feasibility of the new GFC method has been successfully tested on face recognition using a data set from the FERET database, which is a standard testbed for face recognition technologies. Specifically we used 600 FERET frontal face images corresponding to 200 subjects, which were acquired under variable illumination and facial expressions. The effectiveness of the GFC method is shown in terms of both absolute performance indices and comparative performance against some popular face recognition schemes such as the Gabor wavelet method, the Eigenfaces method, the Fisherfaces method, the EFM method, the combination of Gabor and the Eigenfaces method, and the combination of Gabor and the Fisherfaces method. In particular, the novel GFC method achieves 100% recognition accuracy using only 62 features.

The excellent performance shown by the GFC method is the direct result of coupling an augmented Gabor feature vector with the EFM method. The benefits resulting from using Gabor wavelets come from them being the result of evolutionary pressure on the mammalian visual system to develop an optimal sensory architecture tuned to an environment where people live and operate on a regularly basis [14], [15]. While the effectiveness of Gabor wavelets has been shown so far to match and capture only the statistics of natural scenes [14], it is also quite possible that the same Gabor wavelets are also tuned for face processing

tasks [32], another task highly relevant for people. In particular, Field [14] has shown for natural scenes that “they are approximately scale invariant with regards to both their power spectra and their phase spectra. Principally because of the phase spectra, self-similar wavelet-like codes are capable of producing a sparse but informative representation of these images.” What we see and the range of images we are likely to see are limited. Therefore, the human visual system might be tuned for both natural scenery and human faces, and it can take advantage of “the degree of predictability or redundancy in our environment”. Note that as the Gabor wavelets are scale invariant, the statistics of the image must remain constant as one magnifies any local region of the image. The two statistics Field refers to are 1) invariance in contrast across scale (reflected in the power spectrum) and 2) invariance in the local structure (reflected in the phase spectrum). Invariance in local structure means that there exists a number of structures which extend across different frequency bands. Therefore, for both consistency and robustness one can check if such structures are confirmed across several (bandwidth) channels, while moving from coarse to fine resolution, i.e., low to high frequency, even that some drifting may occur. The location of features at one scale can provide a guide for the search for features at other scales. In summary Gabor wavelets yield sparse codes. This does not mean dimensionality reduction, but rather something approaching a factorial code. The nonaccidental occurrence of coincidences in the self-similar wavelet code should facilitate enhanced associations and face recognition [15].

Our next goal is to further search for an optimal and sparse code resulting from the Gabor wavelet representation of face images, before forming the augmented Gabor feature vector and applying the GFC method for classification. The sparse code should represent the sparse structures as displayed by the features of the Gabor transformed face images in terms of spatial locality, scale and orientation selectivity, along the lines suggested by Olshausen and Field [29] for natural image analysis. Another possibility is to search, using evolutionary pursuit (EP) method [22], for the sparse features directly with the twin goals of reducing the amount of data used for classification and simultaneously providing enhanced discriminatory power. The search for such features would be driven by the need to increase the generalization ability of the learning classification machine as a result of leveraging the trade-off between minimizing the empirical risk encountered during training and narrowing the confidence interval for reducing the guaranteed risk while testing on unseen data [37].

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their critical and constructive comments and suggestions.

REFERENCES

- [1] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, pp. 711–720, July 1997.
- [2] D. M. Blackburn, M. Bone, and P. J. Phillips, “Facial recognition vendor test 2000,” <http://www.dodcounterdrug.com/facialrecognition/FRVT2000/documents.htm>, 2001.

- [3] D. Burr, M. Morrone, and D. Spinelli, "Evidence for edge and bar detectors in human vision," *Vis. Res.*, vol. 29, no. 4, pp. 419–431, 1989.
- [4] R. Chellappa, C. L. Wilson, and S. Sirohey, "Human and machine recognition of faces: A survey," *Proc. IEEE*, vol. 83, pp. 705–740, May 1995.
- [5] J. G. Daugman, "Two-dimensional spectral analysis of cortical receptive field profiles," *Vis. Res.*, vol. 20, pp. 847–856, 1980.
- [6] —, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional cortical filters," *J. Opt. Soc. Amer.*, vol. 2, no. 7, pp. 1160–1169, 1985.
- [7] —, "Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 36, pp. 1169–1179, July 1988.
- [8] J. Daugman, "Face and gesture recognition: Overview," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 7, pp. 675–676, 1997.
- [9] K. I. Diamantaras and S. Y. Kung, *Principal Component Neural Networks: Theory and Applications*. New York: Wiley, 1996.
- [10] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, "Classifying facial actions," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 21, pp. 974–989, Oct. 1999.
- [11] S. Edelman, *Representation and Recognition in Vision*. Cambridge, MA: MIT Press, 1999.
- [12] K. Etemad and R. Chellappa, "Discriminant analysis for recognition of human face images," *J. Opt. Soc. Amer. A*, vol. 14, pp. 1724–1733, 1997.
- [13] D. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Amer. A*, vol. 4, no. 12, pp. 2379–2394, 1987.
- [14] —, "Scale-invariance and self-similar 'wavelet' transforms: An analysis of natural scenes and mammalian visual systems," in *Wavelets, Fractals and Fourier Transforms: New Developments and New Applications*, M. Farge, J. Hunt, and J.C. Vassilicos, Eds. Oxford, U.K.: Oxford Univ. Press, 1993, pp. 151–193.
- [15] —, "What is the goal of sensory coding," *Neural Comput.*, vol. 6, pp. 559–601, 1994.
- [16] R. A. Fisher, "The use of multiple measures in taxonomic problems," *Ann. Eugenics*, vol. 7, pp. 179–188, 1936.
- [17] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. New York: Academic, 1991.
- [18] J. Jones and L. Palmer, "An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex," *J. Neurophys.*, pp. 1233–1258, 1987.
- [19] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, pp. 103–108, Jan. 1990.
- [20] K. K. Sung and T. Poggio, "Example-based learning for view-based human face detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, no. 1, pp. 39–51, 1998.
- [21] M. Lades, J. C. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R. P. Wurtz, and W. Konen, "Distortion invariant object recognition in the dynamic link architecture," *IEEE Trans. Comput.*, vol. 42, pp. 300–311, 1993.
- [22] C. Liu and H. Wechsler, "Evolutionary pursuit and its application to face recognition," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, pp. 570–582, June 2000.
- [23] —, "Robust coding schemes for indexing and retrieval from large face databases," *IEEE Trans. Image Processing*, vol. 9, pp. 132–137, Jan. 2000.
- [24] —, "A Gabor feature classifier for face recognition," in *Proc. 8th IEEE Int. Conf. Computer Vision*, Vancouver, BC, Canada, July 9–12, 2001.
- [25] M. J. Lyons, J. Budynek, A. Plante, and S. Akamatsu, "Classifying facial attributes using a 2-D Gabor wavelet representation and discriminant analysis," in *Proc. 4th IEEE Int. Conf. Automatic Face and Gesture Recognition*, 2000.
- [26] M. J. Lyons, J. Budynek, and S. Akamatsu, "Automatic classification of single facial images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 21, pp. 1357–1362, Dec. 1999.
- [27] S. Marcelja, "Mathematical description of the responses of simple cortical cells," *J. Opt. Soc. Amer.*, vol. 70, pp. 1297–1300, 1980.
- [28] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, pp. 696–710, July 1997.
- [29] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 13, pp. 607–609, 1996.
- [30] A. Pentland, "Looking at people: Sensing for ubiquitous and wearable computing," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, pp. 107–119, Jan. 2000.
- [31] P. J. Phillips, H. Wechsler, J. Huang, and P. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms," *Imag. Vis. Comput.*, vol. 16, pp. 295–306, 1998.
- [32] R. Rao and D. Ballard, "An active vision architecture based on iconic representations," *Artif. Intell.*, vol. 78, pp. 461–505, 1995.
- [33] A. Samal and P. A. Iyengar, "Automatic recognition and analysis of human faces and facial expression: A survey," *Pattern Recognit.*, vol. 25, no. 1, pp. 65–77, 1992.
- [34] B. Schiele and J. L. Crowley, "Recognition without correspondence using multidimensional receptive field histograms," *Int. J. Comput. Vis.*, vol. 36, no. 1, pp. 31–52, 2000.
- [35] D. L. Swets and J. Weng, "Using discriminant eigenfeatures for image retrieval," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, pp. 831–836, Aug. 1996.
- [36] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cogn. Neurosci.*, vol. 13, no. 1, pp. 71–86, 1991.
- [37] Y. N. Vapnik, *The Nature of Statistical Learning Theory*. Berlin, Germany: Springer-Verlag, 1995.
- [38] L. Wiskott, J. M. Fellous, N. Kruger, and C. von der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, pp. 775–779, July 1997.



Chengjun Liu (S'98–M'99) received the Ph.D. degree from George Mason University, Fairfax, VA, in 1999.

He is an Assistant Professor of computer science at the New Jersey Institute of Technology, Newark. 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 methods for image/video retrieval and object detection, tracking and recognition based upon

statistical and machine learning concepts. The class of new methods includes the probabilistic reasoning models (PRM), the enhanced Fisher models (EFM), the enhanced independent component analysis (EICA), the Gabor–Fisher Classifier (GFC), and the independent Gabor features (IGF) method. 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 (SM'86–F'91) received the Ph.D. degree in computer science from the University of California, Irvine, in 1975.

He is Professor of computer science at George Mason University, Fairfax, VA. 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, U.K., 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 and the book *Computational Vision* (New York: Academic, 1990). He was the editor of *Neural Networks for Perception Vols. 1 and 2* (New York: Academic, 1991).

Dr. Wechsler was elected an International Association of Pattern Recognition (IAPR) Fellow in 1998.