Correspondence

Robust Coding Schemes for Indexing and Retrieval from Large Face Databases

Chengjun Liu and Harry Wechsler

***Abstract*—This paper introduces two new coding schemes, the probabilistic reasoning models (PRM) and the enhanced FLD (Fisher linear discrimimant) models (EFM), for indexing and retrieval from large image databases with applications to face recognition. The unifying theme of the new schemes is that of lowering the space dimension ("data compression") subject to increased fitness for the discrimination index.**

***Index Terms*—Enhanced FLD models (EFM), face recognition, Fisher linear discrimimant (FLD), indexing and retrieval, probabilistic reasoning models (PRM).**

I. Introduction

The contents of digital libraries consist of increasing amounts of digital media and, in particular, pictorial information characteristic of still images and video data. Computational challenges associated to the large volumes of data involved require low-dimensional image features for indexing and retrieval to support efficient, robust, and scalable image matching. The range of applications is wide open and it covers amongst others security (biometrics and forensics), journalism, telecommunication, electronic commerce, and human computer interaction. This paper is concerned in particular with indexing and retrieval from large facial image databases with applications to face recognition.

Learning to recognize visual objects, such as human faces, requires the ability to find meaningful patterns in spaces of very high dimensionality. 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" [4]. Low-dimensional representations are also important when one considers the intrinsic computational aspect. Principal component analysis (PCA) [6] is the method behind the eigenfaces coding scheme [17] whose primary goal is to project the similarity judgment for face recognition in a low-dimensional space. One should be aware, however, that PCA driven coding schemes are optimal and useful only with respect to data compression and decorrelation of low (second) order statistics. The recognition aspect is not considered and one should thus not expect optimal performance for tasks such as face recognition when using such PCA-like coding schemes. To address this obvious shortcoming, one reformulates the original problem as one where the search is still for low-dimensional patterns but is now also subject to a high discrimination index, characteristic of separable low-dimensional patterns. One solution proposed to solve the new problem is to use 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) [16] and the Fisherfaces [1].

As it is apparent that robust indexing and retrieval schemes require both low-dimensional representations and enhanced discrimination abilities, we address those complementary requirements using an approach similar to constrained optimization (regularization), where the task is that of lowering the space dimension subject to increased performance for the discrimination index. Toward that end we introduce two new coding schemes: the probabilistic reasoning models (PRM) and the enhanced FLD models (EFM). Lowering the space dimension for these coding schemes is done using PCA. The coding schemes assess and modulate the discrimination index using criteria such as the Bayes classifier and specific tradeoffs between PCA and FLD.

II. Probabilistic Reasoning Models (PRM)

We present in this section two PRM's that combine PCA and the Bayes classifier, and then show their feasibility on the face indexing and retrieval problem. PCA is applied first for dimensionality reduction with the goal of signal approximation. The PRM use the within-class scatter to estimate the covariance matrix for each class in order to approximate the conditional pdf, and then apply the MAP rule as the classification criterion. The MAP decision rule optimizes the class separability in the sense of the Bayes error and should improve on PCA and FLD based encoding schemes, which utilize criteria not related to the Bayes error [6].

A. Principal Component Analysis (PCA)

One popular technique for feature selection and dimensionality reduction is PCA [6]. Let $X \in \mathbb{R}^N$ be a random vector representing an image, where N is the dimensionality of the image space. The covariance matrix of X is defined as

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

where

$E(\cdot)$ 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:

$$\Sigma_X = \Phi \Lambda \Phi^t \text{ with } \Phi = [\phi_1, \phi_2, \cdots, \phi_N],$$
$$\Lambda = \text{diag}\{\lambda_1, \lambda_2, \cdots, \lambda_N\} \tag{2}$$

where

$\Phi \in \mathbb{R}^{N \times N}$ orthonormal eigenvector matrix

$\Lambda \in \mathbb{R}^{N \times N}$ diagonal eigenvalue matrix with diagonal elements in decreasing order.

An important property of PCA is its optimal signal reconstruction in the sense of minimum mean square error (MSE) when only a subset

Manuscript received October 9, 1998; revised July 27, 1999. This work was supported in part by the DoD Counterdrug Technology Development Program, with the U.S. Army Research Laboratory as Technical Agent, under Contract DAAL01-97-K-0118. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Thomas S. Huang.

C. Liu was with the Department of Computer Science, George Mason University, Fairfax, VA 22030 USA. He is now with the Department of Mathematics and Computer Science, University of Missouri, St. Louis, MO 63121 USA (e-mail: cliu@cs.umsl.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(00)00190-1.

of principal components, $P = [\phi_1, \phi_2, \cdots, \phi_m]$ where $m < N$, are used to represent the original signal. Following this property, an immediate application of PCA is the dimensionality reduction

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

PCA was first applied to reconstruct human faces by Kirby and Sirovich [8]. Turk and Pentland [17] further developed a well-known face recognition method, known as eigenfaces, where the eigenfaces correspond to the eigenvectors associated with the dominant eigenvalues of the face covariance matrix. The eigenfaces define a feature space, or "face space," which drastically reduces the dimensionality of the original space, and face detection and identification are carried out in the reduced space.

B. MAP Bayes Classification Rule

The Bayes classifier yields the minimum error when the underlying probability density functions (pdf's) are known. This error, called the Bayes error, is the optimal measure for feature effectiveness when classification is of concern, since it is a measure of class separability [6]. The *a posteriori* probability function of ω_i given X is defined as

$$P(\omega_i|X) = \frac{p(X|\omega_i)P(\omega_i)}{p(X)} \tag{4}$$

where

$p(X|\omega_i)$ conditional pdf of ω_i
$p(X)$ mixture density.

The MAP decision rule for the Bayes classifier is

$$p(X|\omega_i)P(\omega_i) = \max_j \{p(X|\omega_j)P(\omega_j)\} \quad X \in \omega_i. \tag{5}$$

The image X is classified to ω_i of whom the *a posteriori* probability given X is the largest among all the other classes.

The Bayes classifier involves the estimation of the conditional pdf for each class and requires a large number of training samples in order to give accurate results. To derive an accurate density estimation nonparametrically is extremely difficult, especially in high-dimensional spaces [6]. As there are usually not enough samples, one would then assume a particular form for the pdf and convert the general density estimation into a parametric one. The within-class densities are usually modeled as normal distributions [10]

$$p(X|\omega_i) = \frac{1}{(2\pi)^{N/2}|\Sigma_i|^{1/2}} \times \exp\left\{-\tfrac{1}{2}(X - M_i)^t \Sigma_i^{-1}(X - M_i)\right\} \tag{6}$$

where M_i and Σ_i are the mean and covariance matrix of class ω_i, respectively.

Under the parametric pdf (multivariate normal densities) assumption, Moghaddam and Pentland [11] used the Bayesian classifier to develop the probabilistic visual learning (PVL) method. In contrast to previously PCA inspired methods which use PCA for dimensionality reduction, PVL works by using the eigenspace decomposition as an integral part of estimating the conditional pdf in the original high-dimensional image space.

C. Unified Statistical Framework

The PRM to be detailed later in this section integrate PCA and the Bayes classifier. The first step is to use PCA for reducing the dimensionality of the original image space from N to m (3): $X \in \mathbb{R}^N \Rightarrow Y \in \mathbb{R}^m$ where $m < N$. The rationale behind applying PCA first for dimensionality reduction instead of exploiting the Bayes classifier and the MAP rule directly on the original data is two-fold. On the one hand, the high dimensionality of the original image space makes the parameter estimation very difficult, if not impossible, due to the fact that the high-dimensional space is mostly empty. This problem of sparsity limits the success of direct Bayesian analysis in the original space, since the amount of training data needed to get reasonably low variance estimators becomes "ridiculously high" [7]. On the other hand, it has been confirmed by many researchers that the PCA representation enjoys image constancy in the sense that it suppresses input noise [8], [12]. The second step of PRM is then to estimate the within-class density and under the normal probability distribution assumption this step is equivalent to estimate the within-class covariance matrices (6).

Estimating the covariance matrix Σ_i in (6) with respect to each class is still difficult due to the limited number of samples for each class. Note that while the mixture covariance matrix is diagonal following PCA, the within-class covariance matrices are not necessarily diagonal. Using different assumptions on the derivation of the within-class covariance matrices but within a unified statistical framework leads to alternative indexing and retrieval methods as follows.

1) Assume the within-class covariance matrices to be unit ones: $\Sigma_i = \Sigma_I = I_m$. Under this assumption, the conditional pdf (6) relaxes to $p(Y|\omega_i) = (1/(2\pi)^{m/2}) \exp\{-(1/2)(Y - M_i)^t(Y - M_i)\}$, where $M_i = E(Y|\omega_i)$. As a result, the MAP rule (5) leads to a distance classifier which corresponds to the eigenfaces method used by Turk and Pentland [17].
2) Following dimensionality reduction using PCA (3), one would use FLD to derive the new feature set $W = P_{\mathrm{FLD}}^t Y$. In the FLD space, assume all the within-class covariance matrices to be unit matrices: $\Sigma_i = \Sigma_I = I_m$. Under this assumption the conditional pdf (6) reduces to $p(W|\omega_i) = (1/(2\pi)^{m/2}) \exp\{-(1/2)(W - M_i)^t(W - M_i)\}$, where $M_i = E(W|\omega_i)$. Again the MAP rule (5) leads to a distance classifier which corresponds to the Fisherfaces method as described by Belhumeur*et al.* [1].
3) In the PCA space (3), assume all the within-class covariance matrices are identical, diagonal: $\Sigma_i = \Sigma_I = \mathrm{diag}\{\sigma_1^2, \sigma_2^2, \cdots, \sigma_{\mathrm{m}}^2\}$. Each diagonal element is estimated by the sample variance in the one dimensional PCA space. This case corresponds to our new method labeled PRM-1 (see Section II-D).
4) Compute the within-class covariance matrix based on the within-class scatter Σ_w in the reduced PCA space. Diagonalize Σ_w, and use the ordered diagonal elements as estimations of the within-class covariance matrices corresponding to those derived in 3). This case corresponds to our new method labeled PRM-2 (see Section II-C).

D. PRM-1 and PRM-2 Models

The two probabilistic reasoning models, PRM-1 and PRM-2, utilize the within-class scatters to derive the estimations of within-class covariance matrices. The PRM-1 model assumes the within-class covariance matrices are identical and diagonal under the assumption of 3) in Section II-D

$$\Sigma_i = \Sigma_I = \mathrm{diag}\{\sigma_1^2, \sigma_2^2, \cdots, \sigma_{\mathrm{m}}^2\}. \tag{7}$$

Each component σ_i^2 is estimated by sample variance in the one-dimensional (1–D) PCA space

$$\sigma_i^2 = \frac{1}{L}\sum_{k=1}^{L}\left\{\frac{1}{N_k - 1}\sum_{j=1}^{N_k}(y_{ji}^{(k)} - m_{ki})^2\right\} \tag{8}$$

where $y_{ji}^{(k)}$ is the ith element of the image $Y_j^{(k)}$ which belongs to class ω_k, m_{ki} the ith element of M_k, and $M_k = E(Y|\omega_k)$.

From (6) and (7), it follows

$$p(Y|\omega_i) = \frac{1}{(2\pi)^{m/2}\prod_{j=1}^{m}\sigma_j} \cdot \exp\left\{-\frac{1}{2}\sum_{j=1}^{m}\frac{(y_j - m_{ij})^2}{\sigma_j^2}\right\} \quad (9)$$

where $Y = (y_1, y_2, \cdots, y_m)^t$. Thus the MAP rule (5) specifies the following classifier (note that the *priors* are set to be equal)

$$\sum_{j=1}^{m}\frac{(y_j - m_{ij})^2}{\sigma_j^2} = \min_k\left\{\sum_{j=1}^{m}\frac{(y_j - m_{kj})^2}{\sigma_j^2}\right\} \Longrightarrow Y \in \omega_i. \quad (10)$$

The PRM-2 model estimates the within-class scatter matrix Σ_w in the reduced PCA space as

$$\Sigma_w = \frac{1}{L}\sum_{k=1}^{L}\left\{\frac{1}{N_k}\sum_{j=1}^{N_k}(Y_j^{(k)} - M_k)(Y_j^{(k)} - M_k)^t\right\}. \quad (11)$$

To avoid the explicit calculation of Σ_w and to improve numerical accuracy, we calculate the singular value decomposition (SVD) of matrix Z, where $\Sigma_w = ZZ^t$

$$Z = USV^t \quad (12)$$

where

U and V	orthonormal matrices,
S	diagonal matrix

$$S = \mathrm{diag}\{\mathrm{s}_1, \mathrm{s}_2, \cdots, \mathrm{s}_\mathrm{m}\} \quad (13)$$

with nonnegative singular values as diagonal elements. Order the squared diagonal elements as

$$(s_{(1)}^2, s_{(2)}^2, \cdots, s_{(m)}^2) = \text{ order } \{s_1^2, s_2^2, \cdots, s_m^2\}. \quad (14)$$

Finally, under the assumption of d) in Section II-C, the within-class covariance matrix is derived as

$$\Sigma_I = \mathrm{diag}\{\mathrm{s}_{(1)}^2, \mathrm{s}_{(2)}^2, \cdots, \mathrm{s}_{(\mathrm{m})}^2\}. \quad (15)$$

From (5), (6), and (15) the MAP rule specifies another classifier (note that again the *priors* are set to be equal)

$$\sum_{j=1}^{m}\frac{(y_j - m_{ij})^2}{s_{(j)}^2} = \min_k\left\{\sum_{j=1}^{m}\frac{(y_j - m_{kj})^2}{s_{(j)}^2}\right\} \Longrightarrow Y \in \omega_i. \quad (16)$$

E. Experimental Results

After estimating the within-class covariance matrices using (7) and (8) (for PRM-1) and (12) and (15) (for PRM-2), indexing and retrieval as appropriate for face recognition tasks can be carried out using the Bayes decision rule (5). The *a prior* probabilities are assigned values according to *a prior* knowledge, and are set to be equal in our experiments (10), (16).

The experimental data consists of 1107 facial images corresponding to 369 subjects and comes from the U.S. Army FERET database [14]. 600 out of the 1107 images correspond to 200 subjects with each subject having three images—two of them are the first and the second shot, and the third shot is taken under low illumination. For the remaining 169 subjects, there are also three images for each subject, but two out of the three images are duplicates taken at a different time. Two images of each subject are used for training with the remaining image for testing. The images are cropped to the size of 64 × 96, and the eye coordinates are manually detected.

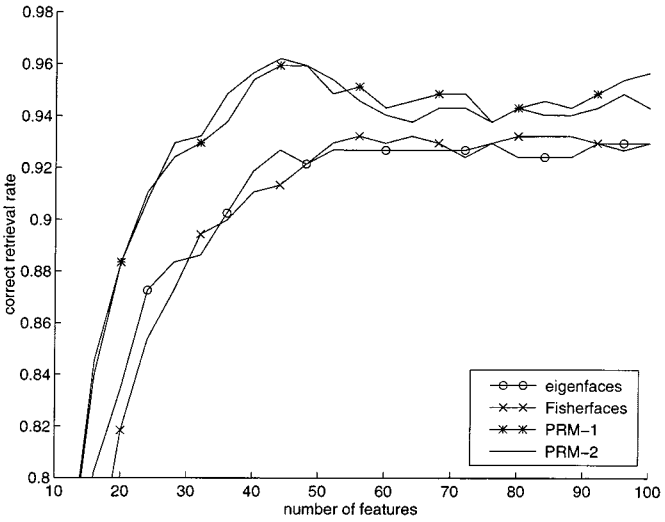

Fig. 1. Comparative testing performance for the eigenfaces, the Fisherfaces (using 200 principal components), and the PRM method (PRM-1, PRM-2).

Fig. 1 shows the comparative performance of eigenfaces, Fisherfaces, PRM-1 and PRM-2, and one can see that the PRM consistently better the eigenfaces and Fisherfaces methods. In particular, the PRM class (PRM-1 and PRM-2) increases the peak retrieval rate by 5% when compared to eigenfaces and Fisherfaces methods; the peak retrieval rate for both PRM is about 96% using 44 features.

III. Enhanced FLD Models (EFM)

We present in this section two EFM's in order to improve the generalization capability of FLD based classifiers. Similar to Fisherfaces, both EFM models (EFM-1 and EFM-2) apply first PCA for dimensionality reduction before proceeding with FLD analysis. EFM-1 implements the dimensionality reduction with the goal to balance between the need that the selected eigenvalues account for most of the spectral energy of the raw data and the requirement that the eigenvalues of the within-class scatter matrix in the reduced PCA space are not too small. EFM-2 implements the dimensionality reduction as the Fisherfaces method does. It proceeds then with the whitening of the within-class scatter matrix in the reduced PCA space and then chooses a small set of features (corresponding to the eigenvectors of the within-class scatter matrix) so that the smaller trailing eigenvalues are not included further in the computation of the between-class scatter matrix.

A. Fisher Linear Discriminant (FLD)

The FLD is a widely used technique for feature selection. Let $\omega_1, \omega_2, \cdots, \omega_L$ and $N_1, N_2, \cdots, N_L$ denote the classes and the number of images within each class, respectively. L is the number of the classes. Let $M_1, M_2, \cdots, M_L$ and M be the means of the classes and the grand mean. Let Σ_w and Σ_b be the within- and between-class scatter matrices. The standard FLD procedure derives a projection matrix Ψ that maximizes the ratio $|\Psi^t\Sigma_b\Psi|/|\Psi^t\Sigma_w\Psi|$. This ratio is maximized when Ψ consists of the eigenvectors of the matrix $\Sigma_w^{-1}\Sigma_b$ [6]

$$\Sigma_w^{-1}\Sigma_b\Psi = \Psi\Delta \quad (17)$$

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

The FLD overcomes one of PCA's drawbacks since it can distinguish within- and between-class scatters. Furthermore, FLD induces nonorthogonal projection axes, a characteristic known to have great functional significance in biological sensory systems [3]. The drawback of FLD is that it requires large sample size for good generalization. As this requirement is rarely met, FLD overfits and thus generalizes poorly when compared against PCA schemes [9]. One possible remedy for this drawback is to artificially generate additional data and thus increase the sample size [5].

FLD is behind several face recognition methods [1], [5], [16]. 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 developed by Belhumeur *et al.* [1]. Using similar arguments, Swets and Weng [16] point out that the eigenfaces derive only the most expressive features (MEF's). As explained earlier, such PCA inspired features do not necessarily provide for good discrimination. As a consequence, subsequent FLD projections are used to build the most discriminant 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 [16].

B. Standard FLD-Based Methods and Overfitting

The standard FLD based methods apply first PCA for dimensionality reduction and then FLD for discriminant analysis [1], [16]. Relevant questions concerning PCA representation are usually related to the range of principal components used and how it affects performance. Regarding the FLD discriminant analysis one has to understand the reasons for overfitting and how to avoid it. For image representation, the more principal components are used the better the quality of reconstruction. The same reasoning does not apply, however, for class discrimination. One can actually show that using more principal components may lead to decreased classification performance [9]. The explanation for such behavior is that the trailing eigenvalues (resulting from the more principal components used) correspond to high-frequency components and usually encode noise. As a result, when these trailing but small valued eigenvalues are used to define the reduced PCA space, the FLD procedure has to fit for noise as well and as a consequence overfitting takes place.

The FLD procedure involves the simultaneous diagonalization of the two within- and between-class scatter matrices and it is stepwise equivalent to two operations as pointed out by Fukunaga [6]: first whitening the within-class scatter matrix, and second applying PCA on the between-class scatter matrix using the transformed data. The purpose of the whitening step is to normalize (to unity) the within-class scatter matrix for uniform gain control. The second operation maximizes then the between-class scatter to separate different classes as much as possible. As during whitening the eigenvalues of the within-class scatter matrix appear in the denominator, the small (trailing) eigenvalues cause the whitening step to fit for misleading variations and thus generalize poorly when exposed to new data. The EFM, detailed in the following subsections, successfully address the overfitting problem and as a result would display increased generalization capabilities.

C. EFM-1

For EFM-1, one wants to choose m, the number of principal components in (3), such that a proper balance is maintained between the need

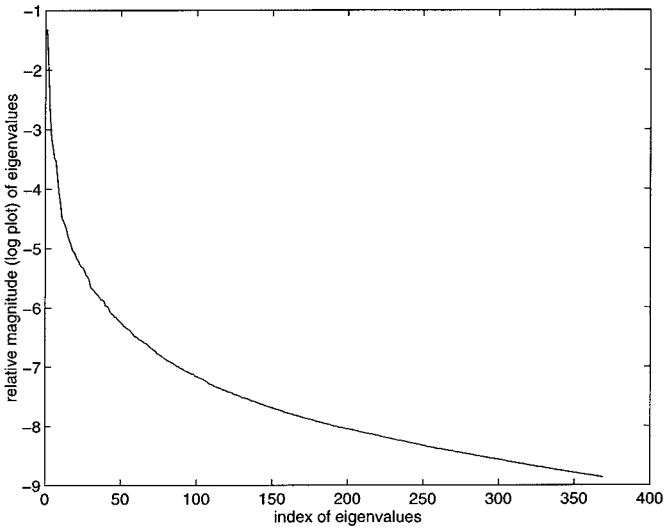

Fig. 2. Relative magnitude $(\lambda_i / \sum_{k=1}^{m} \lambda_k)$ of the eigenvalues of the covariance matrix.

that the selected eigenvalues account for most of the spectral energy of the raw data and the requirement that the eigenvalues of the within-class scatter matrix (in the reduced PCA space) are not too small. The eigenvalue spectrum of PCA is derived by (2), and the relative magnitude of the eigenvalues for the face data used in our experiments (see Section II-E) is shown in Fig. 2. The index for the eigenvalues ranges from 1 to $m = L$, where L stands for the number of (face) classes considered in our experiments. One can see from Fig. 2 that the first 50 eigenvalues capture most of the energy and that the eigenvalues whose indices are greater than 100 are fairly small and most likely capture noise.

To calculate the eigenvalue spectrum of the within-class scatter matrix in the reduced PCA space one needs to decompose the FLD procedure into two operations: whitening and diagonalization [6]. In particular, let Σ'_w, $\Sigma'_b \in \mathbb{R}^{m \times m}$ represent the within- and between-class scatter matrices in the reduced PCA space, respectively. The stepwise FLD procedure derives the eigenvalues and eigenvectors of $\Sigma'^{-1}_w \Sigma'_b$ as the result of the simultaneous diagonalization of Σ'_w and Σ'_b. First whiten the within-class scatter matrix

$$\Upsilon^{-1/2} \Xi^t \Sigma'_w \Xi \Upsilon^{-1/2} = I \tag{18}$$

where Υ, $\Xi \in \mathbb{R}^{m \times m}$ are the eigenvalue and eigenvector matrices of Σ'_w

$$\Sigma'_w \Xi = \Xi \Upsilon \quad \text{and} \quad \Xi^t \Xi = I. \tag{19}$$

The eigenvalue spectrum of the within-class scatter matrix in the reduced PCA space can be derived by (19), and different spectra are obtained corresponding to different number of principal components that are utilized as shown in Fig. 3. Now one has to simultaneously optimize the behavior of the trailing eigenvalues in the reduced PCA space (Fig. 3) with the energy criteria for the original image space (Fig. 2). Note that different choices on the cutoff principal component index for the original image space yield different within-class spectra as shown in Fig. 3. As one can see from Figs. 2 and 3, a suitable choice is to set $m = 50$, since this choice not only accounts for most of the spectral energy of the raw data (Fig. 2) but also meets the requirement that the eigenvalues of the within-class scatter matrix (in the reduced 50 dimensional PCA space) are not too small (Fig. 3).

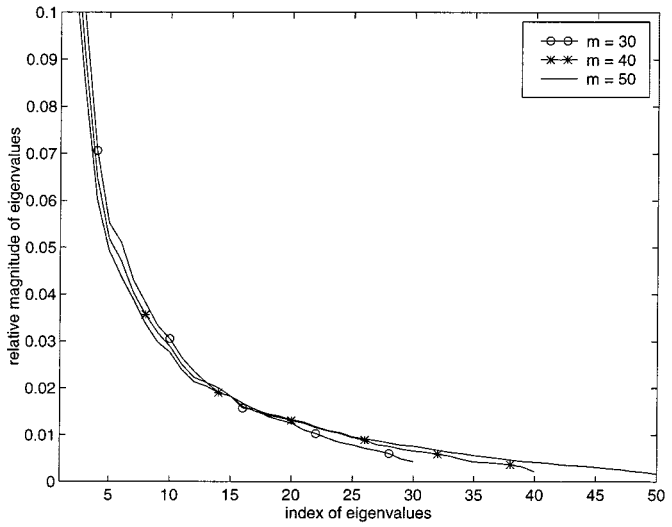

Fig. 3. Relative magnitude of the eigenvalues spectra derived using the FLD whitening step with different number of principal components ($m = 30$, $m = 40$, and $m = 50$) for dimensionality reduction.

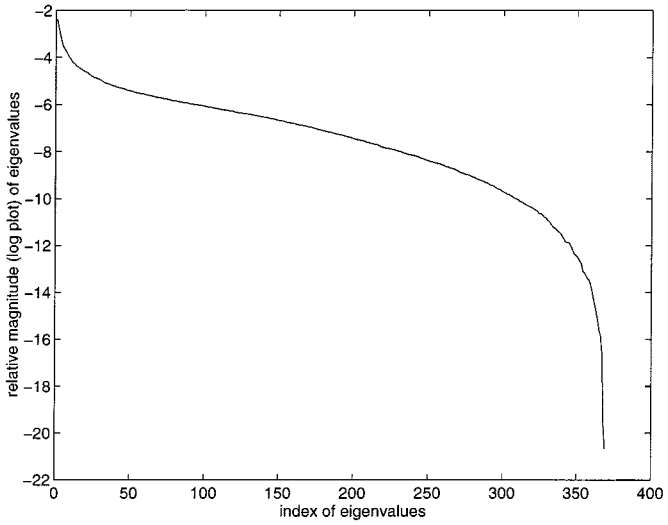

Fig. 4. Relative magnitude ($\lambda_i / \sum_{k=1}^{m} \lambda_k$) of the eigenvalues derived from the FLD whitening step in the $m = 369$ dimensional PCA space.

After the number of principal components is set, EFM-1 computes the between-class scatter matrix K_b as

$$K_b = \Upsilon^{-1/2} \Xi^t \Sigma_b' \Xi \Upsilon^{-1/2}. \tag{20}$$

Diagonalize the new between-class scatter matrix K_b

$$K_b \Theta = \Theta \Delta \quad \text{and} \quad \Theta^t \Theta = I. \tag{21}$$

The overall transformation matrix (in the reduced PCA space) for EFM-1 is now defined as

$$T = \Xi \Upsilon^{-1/2} \Theta. \tag{22}$$

D. EFM-2

The EFM-2 model uses the same number m of principal components as utilized by Fisherfaces [1] and MDF method [16] for dimensionality reduction, namely, $L \leq m \leq n - L$, where n is the number of training samples and L the number of classes. Note that in our experiments (see Section II-E) the number of classes $L = 369$ and the number of training images $n = 738$, therefore $m = 369$. The relative magnitude of the eigenvalues derived from the FLD whitening step in the $m = 369$ dimensional PCA space is shown in Fig. 4. The EFM-2 proceeds then by choosing a small set of features [corresponding to the eigenvectors of the within-class scatter matrix—see (19)] after the whitening procedure so that the smaller trailing eigenvalues are not included.

Let $s (s < m)$ be the number of features chosen in the whitening space according to the eigenvalue spectrum of Σ_w' (see Fig. 4, a suitable choice is to set $s = 100$), and a new projection matrix $\Xi^* \in \mathbb{R}^{m \times s}$ is derived

$$\Xi^* = [\xi_1, \xi_2, \cdots, \xi_s] \tag{23}$$

where $\xi_1, \xi_2, \cdots, \xi_s$ are the eigenvectors of Σ_w' corresponding to the leading eigenvalues $\gamma_1, \gamma_2, \cdots, \gamma_s$. The new diagonal matrix is

$$\Gamma^* = \operatorname{diag}\{\gamma_1, \gamma_2, \cdots, \gamma_s\}. \tag{24}$$

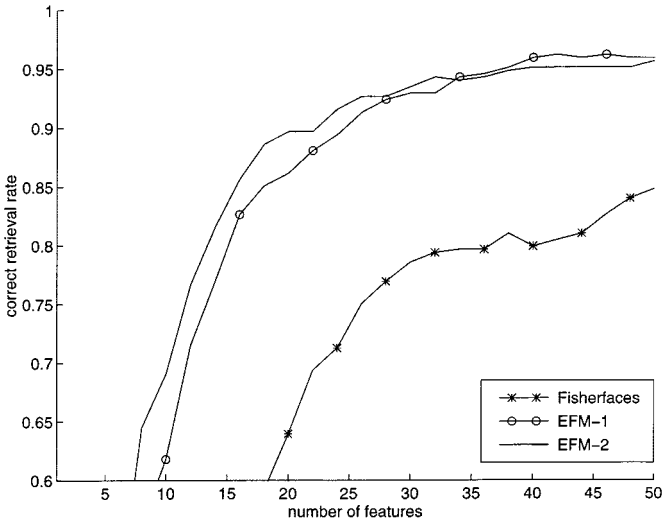

Fig. 5. Comparative performance for Fisherfaces (using 369 principal components), EFM-1 and EFM-2 models.

The new whitening transformation matrix in the reduced s dimensional whitened space is

$$Q = (\Xi^*)(\Gamma^*)^{-1/2}. \tag{25}$$

The between-class scatter matrix Σ_b in this space is transformed to K_b^*

$$K_b^* = Q^t \Sigma_b Q. \tag{26}$$

Note that the new between-class scatter matrix K_b^* is $s \times s$ now instead of $m \times m$. Diagonalize K_b^*

$$K_b^* \Theta^* = \Theta^* \Delta^* \quad \text{and} \quad \Theta^{*t} \Theta^* = I. \tag{27}$$

The overall transformation matrix (after PCA) for EFM-2 is

$$T^* = \Xi^* \Gamma^{*-1/2} \Theta^*. \tag{28}$$

E. Experimental Results

The experimental data used is as described in Section II-E. Fig. 5 shows the comparative performance for Fisherfaces and our new EFM-1 and EFM-2 models. For the Fisherfaces, the principal components is optimally set to L, because $L \leq m \leq n - L$ and $n = 2L$ in our experiments (the number of classes $L = 369$, the number of training images $n = 738$). The EFM methods increase the correct retrieval rate by 10–15% compared to Fisherfaces.

IV. Conclusions

We introduced in this paper two new coding schemes, the probabilistic reasoning models (PRM) and the enhanced FLD models (EFM), for indexing and retrieval from large image databases with applications to face recognition. Experimental results using in excess of one thousand face images from the FERET facial image database show the feasibility and the robustness of our approach. Robustness is shown in terms of both absolute performance indices and comparative performance against traditional face recognition methods such as eigenfaces and Fisherfaces. Even though we have used the FERET digital face library to demonstrate the robustness of our encoding schemes, our general approach is not restricted to face images and can be applied to other digital image libraries as well. The methods provide robust coding schemes for image content-based indexing and retrieval based on image features derived in very low dimensional space. Because of the low dimensionality of the feature space, these methods are suitable for real time indexing and retrieval of large databases.

References

[1] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, pp. 711–720, 1997.

[2] R. Chellappa, C. Wilson, and S. Sirohey, "Human and machine recognition of faces: A survey," *Proc. IEEE*, vol. 83, pp. 705–740, 1995.

[3] J. Daugman, "An information-theoretic view of analog representation in striate cortex," in *Computational Neuroscience*, E. Schwartz, Ed. Cambridge, MA: MIT Press, 1990, pp. 403–424.

[4] 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, New York: Academic , 1990.

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

[6] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed, New York: Academic , 1991.

[7] N. Intrator and L. Cooper, "Objective function formulation of the BCM theory of visual cortical plasticity: Statistical connections, stability conditions," *Neural Networks*, vol. 5, pp. 3–17, 1992.

[8] M. Kirby and L. Sirovich, "Application of the Karlumen–Loeve procedure for the characterization of human faces," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, pp. 103–108, 1990.

[9] C. Liu and H. Wechsler, "Enhanced Fisher linear discriminant models for face recognition," in *Proc. 14th Int. Conf. Pattern Recognition*, Queensland, Australia, Aug. 17–20, 1998.

[10] ——, "Probabilistic reasoning models for face recognition," in *Proc. IEEE Computer Soc. Conf. Computer Vision and Pattern Recognition*, Santa Barbara, CA, June 23–25, 1998.

[11] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, pp. 696–710, 1997.

[12] P. Penev and J. Atick, "Local feature analysis: A general statistical theory for object representation," *Network: Comput. Neural Syst.*, vol. 7, pp. 477–500, 1996.

[13] P. Phillips, "Matching pursuit filters applied to face identification," *IEEE Trans. Image Processing*, vol. 7, pp. 1150–1164, 1998.

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

[15] A. Samal and P. Iyengar, "Automatic recognition and analysis of human faces and facial expression: A survey," *Pattern Recognit.*, vol. 25, pp. 65–77, 1992.

[16] D. L. Swets and J. Weng, "Using discriminant eigenfeatures for image retrieval," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, pp. 831–836, 1996.

[17] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cogn. Neurosci.*, vol. 13, pp. 71–86, 1991.

Detection of Moving Objects in Video Using a Robust Motion Similarity Measure

Hieu T. Nguyen, Marcel Worring, and Anuj Dev

***Abstract*—This correspondence deals with the segmentation of a video clip into independently moving visual objects. This is an important step in structuring video data for storage in digital libraries. The method follows a bottom-up approach. The major contribution is a new well-founded measure for motion similarity leading to a robust method for merging regions. The improvements with respect to existing methods have been confirmed by experimental results.**

***Index Terms*—Content-based indexing, motion segmentation, motion similarity, video analysis.**

I. Introduction

Video is becoming an important datatype in digital libraries. Besides the traditional verbal user queries, the library should support queries based on object shape and motion. Objects and their characteristics therefore form basic units for content-based retrieval and presentation of video. They are further useful for video compression and the creation of hypervideo documents.

Our correspondence deals with motion segmentation, i.e., the decomposition of a video scene into independently moving visual objects. Starting from an oversegmentation of the scene, it merges regions based on motion similarity.

This work is structured as follows. In Section II we formulate the problem and provide a short review of existing literature. Our new region merging method is described in Section III. Section IV shows the results of applying the method on some standard test image sequences.

Manuscript received December 1, 1998; revised July 27, 1999. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. A. Murat Tekalp.

The authors are with the Intelligent Sensory Information Systems Group, University of Amsterdam, Faculty of WINS, NL-1098 SJ, Amsterdam, The Netherlands (e-mail: tat@wins.uva.nl; worring@wins.uva.nl; anuj@wins.uva.nl).

Publisher Item Identifier S 1057-7149(00)00188-3.