CAR-TR-914
CS-TR-4009

N00014-95-1-0521
April 1999

Subspace Linear Discriminant Analysis for Face Recognition

W. Zhao, R. Chellappa
Center for Automation Research
University of Maryland
College Park, MD 20742-3275
{wyzhao, rama}@cfar.umd.edu
Tel/Fax: (301) 405-1235, 314-9115

P.J. Phillips
National Institute of Standards
and Technology
Gaithersburg, MD 20899
jonathon@nist.gov

EDICS: Image Analysis (IP 1.9)

Abstract

In this paper we describe a holistic face recognition method based on subspace Linear Discriminant Analysis (LDA). The method consists of two steps: first we project the face image from the original vector space to a *face subspace* via Principal Component Analysis where the subspace dimension is carefully chosen, and then we use LDA to obtain a linear classifier in the subspace. The criterion we use to choose the subspace dimension enables us to generate class-separable features via LDA from the *full* subspace representation. Hence we are able to solve the *generalization/overfitting* problem when we perform face recognition on a large face dataset but with *very few* training face images available per testing person. In addition, we employ a weighted distance metric guided by the LDA eigenvalues to improve the performance of the subspace LDA method. Finally, the improved performance of the subspace LDA approach is demonstrated through experiments using the FERET dataset for face recognition/verification, a large mugshot dataset for person verification, and the MPEG-7 dataset. We believe that this approach provides a useful framework for other image recognition tasks as well.

Partially supported by the Office of Naval Research under Grant N00014-95-1-0521.

1 Introduction

The problem of automatic face recognition involves detection and location of faces in a cluttered background, normalization, identification and/or verification. Depending on the nature of the application, e.g. the sizes of the training and testing databases, clutter and variability of the background, noise, occlusion, and finally speed requirements, some of the subtasks can be very challenging. In this paper, we focus on the subtasks of identification and verification and demonstrate performance on a database of about 3800 images collected as part of the FERET test, a mugshot database of about 2200 persons, and a MPEG-7 content set of face images provided by the Heinrich Hertz Institute of Germany. In identification problems, the input to the system is an unknown face, and the system reports back the decided identity from a database of known individuals, whereas in verification problems, the system needs to confirm or reject the claimed identity of the input face.

Many methods have been proposed for face recognition [2, 1]. Basically they can be divided into holistic template matching based systems [6, 8, 7, 10, 34, 23, 9] and geometrical local-feature-based schemes [25, 11]. Even though both types of systems have been successfully applied to the task of face recognition, they do have certain advantages and disadvantages. Thus appropriate schemes should be chosen based on the specific requirements of a given task.

In the feature-based approach, many systems have been developed in which wavelets play a key role in facial image representation, including Gabor 'jets' [26] and matching pursuit filters [35], for example. The rationale behind this is due to the relative robustness of wavelet coefficients to illumination change, etc. While there are many wavelets available, Gabor wavelets have been the most popular ones used in face recognition. Several methods of obtaining the local feature representation have also been considered: one is a set of wavelet coefficients for different scales and rotations based on fixed wavelet bases (called 'jets' in [26]), and another is coefficients based on adaptive wavelet bases (called matching pursuit filters in [35]). After obtaining the local features, there are several ways to group the local representations in order to match the input image with a database of stored images. For example, Elastic Graph Matching (EGM) is proposed in [25] as an efficient way to perform matching based on the 'jets' representation.

Proposed as an effective dimensionality reduction method [3, 12, 13], the PCA plays a fundamental role in the holistic methods. These systems typically involve first projecting the original face image into the eigen-subspace (eigenface), and then applying appropriate classifiers. Different pattern classifiers have been used for face recognition, including the Nearest-Neighbor rule [6], Bayesian [7], and LDA/FLD (Fisher Linear Discriminant) classifiers [34, 23, 10], to name a few.

Of those classifiers, LDA is an attractive choice for face recognition/verification tasks. This is because of the following reasons: (1) Unlike PCA which codes information relevant to compression, LDA encodes discriminatory information; (2) LDA is a good classifier when the input features are linearly separable; and finally (3) LDA is simple to implement and hence is a good choice when computational complexity is of great concern. For large databases such as the FERET database, there exist questions about how to *generalize* the chosen classifier based on available training samples to new testing samples from trained classes or new classes [23]. In this paper, we treat the PCA coefficients as *generalizable* features and thus the choice of dimensionality of the subspace is based mainly on the characteristics of the eigenvectors instead of the eigenvalues. Combining PCA and LDA we propose a subspace LDA based face recognition method which solves the generalization/overfitting problem. In addition, in order to obtain better performance, we propose using a weighted distance metric in the composite PCA, LDA projection space. The efficacy of this approach is supported by extensive experiments using the

FERET database, a large mugshot database and an MPEG-7 face dataset.

This paper is organized as follows: The following section describes the subspace LDA approach to face recognition. Section 3 presents experimental results using various face datasets, and a performance comparison between subspace LDA and other algorithms. The last section contains discussions, conclusions and comments on future research directions in this area.

2 Subspace LDA for face recognition

2.1 LDA as a pattern classifier

In typical applications, two or four training samples per person are available (as in the FERET development set) and the original feature dimension is 2016 (with face image size 48×42) or higher. Even with face subspace projection, the feature dimension is still around 300 [1], for example. With a dimension/sample ratio of 150 or 75, this problem leads to a typical *dimensionality curse* phenomenon. However, LDA can still be applied as pattern classifier in such bad situations.

LDA training is carried out via scatter matrix analysis [3]. For an M-class problem, the within- and between-class scatter matrices S_w, S_b are computed as follows:

$$S_w = \sum_{i=1}^{M} Pr(C_i)\Sigma_i, \tag{1}$$

$$S_b = \sum_{i=1}^{M} Pr(C_i)(\mathbf{m}_i - \mathbf{m}_0)(\mathbf{m}_i - \mathbf{m}_0)^T \tag{2}$$

where $Pr(C_i)$ is the *prior* class probability and usually is replaced by $1/M$ in practice with the assumption of equal priors. Here S_w is the *Within-class Scatter Matrix* showing the average scatter Σ_i of the sample vectors $\mathbf{x}$ of different classes C_i around their respective means $\mathbf{m}_i$:

$$\Sigma_i = E[(\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T | C = C_i] \tag{3}$$

Similarly S_b is the *Between-class Scatter Matrix*, representing the scatter of the conditional mean vectors $\mathbf{m}_i$ around the overall mean vector $\mathbf{m}_0$.

Various measures are available for quantifying the discriminatory power [2], a commonly used one being the ratio of the determinant of the between-class scatter matrix of the projected samples to the within-class scatter matrix of the projected samples:

$$\mathcal{J}(T) = \frac{|T^T S_b T|}{|T^T S_w T|}. \tag{4}$$

Let us denote the optimal projection/matrix which maximizes $\mathcal{J}(T)$ by W; then W can be obtained via solving the generalized eigenvalue problem [17]:

$$S_b W = S_w W \Lambda_W \tag{5}$$

Finally the linear discriminant functions for LDA are

$$d_i(\mathbf{x}) = W^T(\mathbf{x} - \mathbf{m}_i) \tag{6}$$

[1] The subspace dimension is carefully chosen and should not be made small, as we show later.

[2] A lot of these criteria are equivalent.

2.1.1 Compute optimal LDA projection

There are several ways to solve the *generalized eigen-problem* of (5). One is to directly compute the inverse of S_w and solve a non-symmetric (in general) eigen-problem for matrix $S_w^{-1}S_b$. But this approach is numerically unstable since it involves the direct inversion of a potentially very large matrix which is probably close to being singular. Another way to solve this equation is to recover a symmetric eigen-problem by using Cholesky decomposition (refer to [28]) or solving the eigen-problem for S_w first [3, 10]. Since S_w is a real symmetric matrix, there exist orthonormal Q and diagonal Λ_Q such that $S_w = Q\Lambda_Q Q^T$. Then the original problem (5) becomes

$$(\Lambda_Q^{-\frac{1}{2}}Q^T)S_bW = (\Lambda_Q^{\frac{1}{2}}Q^T)W\Lambda_W, \tag{7}$$

and it can be further simplified as

$$(RS_bR^T)W_R = W_R\Lambda_W, \tag{8}$$

with $R = \Lambda_Q^{-\frac{1}{2}}Q^T$ and $W_R = \Lambda_Q^{\frac{1}{2}}Q^TW$. So we can first solve a symmetric eigen-problem (8) to obtain Λ_W and W_R, and then compute W as $Q\Lambda_Q^{-\frac{1}{2}}W_R$.

For face applications, the matrix size is usually very large and it is likely that matrix S_w is close to being singular due to round-off error. In order to solve this numerical problem, we slightly modify matrix S_w to $S_w + \Delta I$, where Δ is a (relatively) very small positive number such that $S_w + \Delta I$ is strictly positive definite. By adding such a small diagonal positive number to matrix S_w, we effectively carry out a *regularization* procedure for the *ill-posed problem*. This approach is quite different from the traditional one which usually performs *pure* dimension reduction by applying PCA first in order to obtain a full-rank S_w. Both our approach and the traditional dimension reduction approach provide solutions to the following problem: how to perform LDA when the number of training samples is less than the image/vector size but large enough to guarantee the full rank of S_w (in reduced dimension).

However, with the traditional dimension reduction approach one cannot solve the following practical problem: How to apply LDA when only one sample per class is available for training. In such a case, S_w is augmented to be an identity matrix (with a small constant factor); hence the problem reduces to a standard eigen-problem with S_b being the "covariance" matrix.

2.1.2 Drawback of the LDA face recognition system

We implemented an LDA-based face recognition system and took part in the FERET test in September 1996, and the results were encouraging but not satisfactory on a large dataset of persons not originally present in the training set. Although the pure LDA algorithm does not have any problem in discriminating the trained samples, we have observed that it does not perform very well in the following three cases:

1. when the testing samples are from persons not in the training set
2. when markedly different samples of the trained classes are presented
3. when samples with different backgrounds are presented

Basically this is a *generalization/over-fitting problem* since the pure LDA based system is very much tuned to the specific training set, which has the same number of classes as persons, with two or four samples per class.

2.2 Observation: *face subspace*

As a pattern classification task, face recognition has its own special characteristics. The basic purpose of face recognition is to distinguish persons via their face images; thus different faces of the same person are treated as belonging to one class. On the other hand, face images of all persons are a well defined super-class. Considering a 256×256 image space, all kinds of 2D images of objects are located in this image space: a bird, a car, a tree, a human face. Thus classifying an input image of that size as a face, bird, or car is a well-defined problem. This implies that if we directly apply *any* classifier in the image space, for example, a linear discriminant classifier, the classifier will extract the features which have the most discriminating power from all possible sources in this image space, even including noise or background. So in order to distinguish among face images it makes sense to compare them only in the face subspace. The concept of the face subspace was originally suggested by Sirovich and Kirby [5], and later by Turk and Pentland [6] and other researchers.

To demonstrate that using an arbitrarily small number of subspace components to classify classes is not a good idea, we randomly pick one image from the FERET development dataset and plot (Figure 1) the original image along with its reconstructed images with different numbers of eigenvectors. As can be seen, when only a few leading components are used, two originally different face images may be altered/reconstructed into very similar ones, resulting in bad performance for a large test set.

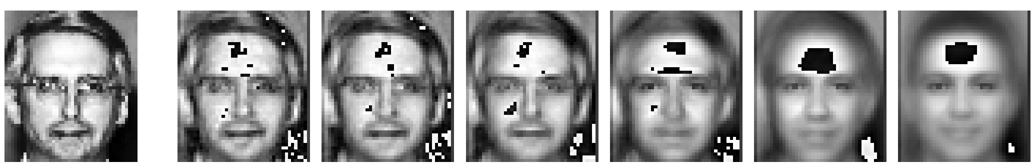


Figure 1: Original image and the reconstructed images using 300, 200, 100, 50, 20, 10 leading components respectively.

Unlike existing approaches, we argue that the dimension of the face subspace is fixed (for a given training set) regardless of the image size [27]. The property of relative invariance of the subspace dimension enables us to work with smaller face images without sacrificing performance. This claim is supported by our experiments using normalized face images of different sizes to obtain different face subspaces. The choice of such a fixed subspace dimension is mainly based on the characteristics of the eigenvectors instead of the eigenvalues [23]. For instance, in [10] the number of leading PCA eigenvalues is selected such that the retained eigenvalues contain 95% of the total energy, while in [16] the dimension is chosen to be 40% of the total number of eigenvectors based on empirical performance evaluation of the *PCA classifier* on large databases. Based on our training set of 1078 original FERET images and their mirror-reflected/noisy images, we have 4312 images. Thus the dimension of the face subspace is around 200 according to [10], and around 1600 according to [16]. We have noticed that the choice of 1600 agrees with our experimental result of applying PCA as the classifier. (In our experiment, we found that PCA with dimension 1000 outperforms PCAs with dimensions 15 and 300.) However, our performance evaluation of *subspace LDA* using different subspace dimensions (200, 300, 400, 500, for example) confirms our earlier choice of 300 as the fixed face subspace dimension.

2.3 Comparison of PCA, LDA and subspace LDA

To visualize the comparison of classifiers PCA, LDA and subspace LDA, we plot three different types of bases (PCA, LDA and subspace LDA) here. First of all, the three types of decompositions are listed here:

$$\mathbf{x} = \sum_{i=1}^{n} a_i \Phi_i \approx \sum_{i=1}^{m} a_i \Phi_i \qquad \text{PCA} \tag{9}$$

$$\mathbf{x} = \sum_{i=1}^{K} b_i W_i + \sum_{i=K+1}^{n} b_i W_i^C \qquad \text{LDA} \tag{10}$$

$$\mathbf{x} \approx \sum_{i=1}^{m} a_i \Phi_i = \sum_{i=1}^{k} c_i W_{x,i} + \sum_{i=k+1}^{m} c_i W_{x,i}^C \qquad \text{subspace LDA} \tag{11}$$

where n is the dimension of the original signal, m is the subspace signal dimension, $K = \min(n, M-1)$ and $k = \min(m, M-1)$, and W^C and W_x^C are the complementary subspaces of LDA's W and subspace LDA's W_x respectively. The a_i, b_i, and c_i's are the corresponding projection coefficients.

Since a linear projection can be viewed as a projection onto a set of bases (PCA bases are orthogonal, LDA and subspace LDA bases are not orthogonal in general), we can visualize these bases. Three different sets of bases from three different linear projections are shown here (Figure 2): (1) pure LDA projection W, (2) pure PCA projection Φ, and (3) subspace LDA projection W_x. All of these bases are computed using the FERET *training* set [14], the subspace LDA bases being based on the first 300 PCA eigenvectors. Through this visualization, we can clearly see that subspace LDA provides a constrained version of pure LDA, and thus has better *generalization* for both *trained classes* and *non-trained classes*.

For a classification task with a large number of classes, the subspace (PCA) projection *may* offer an additional advantage over the LDA classifier. Notice that the decomposition of the original signal $\mathbf{x}$ into the face subspace Φ can be expressed as in (9). Equivalently, the projection coefficients a_i of $\mathbf{x}$ can be expressed as

$$a_i = [\Phi^T \mathbf{x}]^i \tag{12}$$

where $[\cdot]^i$ denotes the ith element of the vector. Since the original signal $\mathbf{x}$ (mean-subtracted face image) is a well-defined class **face**, it is reasonable to assume that all sample images are i.i.d. realizations of the face images class. Hence by the central limit theorem, the projection coefficient vectors can be viewed as samples generated from a *white Gaussian random vector* [3]. Figure 3 shows the histogram of the signal components of the original face image and the PCA projections. Noticing that the components have similar histograms (Gaussian), we only plot the histogram for one component.

2.4 Implementation of subspace LDA

By projecting the original images onto the subspace with the dimension carefully chosen, we can solve the generalization/overfitting problem for face recognition. We have also noticed that the PCA is a conventional method to obtain such a subspace [4]. Thus our natural choice becomes

[3] The signal has been whitened through PCA projection.

[4] The feature subspace may also be obtained through nonlinear transformation such as wavelet projection [29].

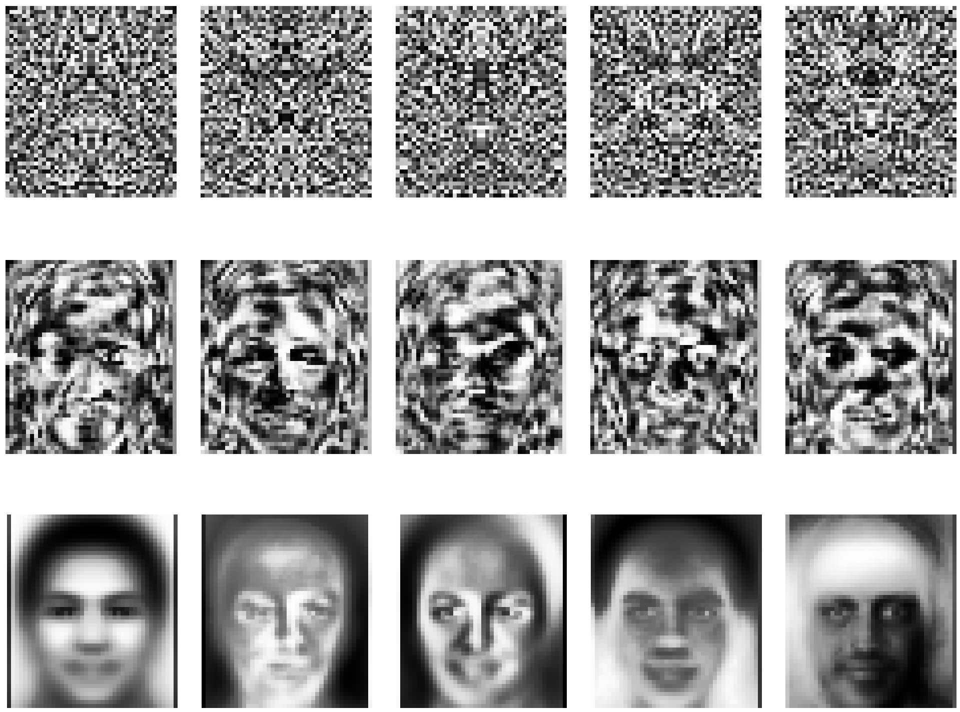


Figure 2: Different bases: The first row shows the first five pure LDA bases, the second row represents the first five subspace LDA bases, and finally the average face and first four eigenfaces are shown on the third row.

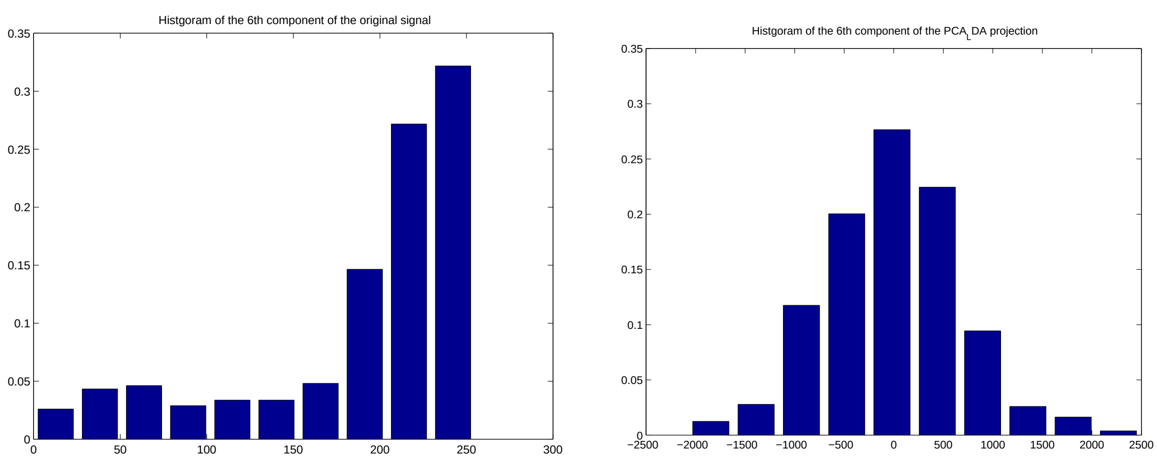


Figure 3: Histogram of the 6th component across all 1038 FERET development samples: the left plot is for the original face images while the right plot is for the PCA projection coefficient vector.

combining PCA and LDA as previously explored by Swet and Weng [10] and by Belhumeur *et al.* [34]. Combining PCA and LDA, we obtain a linear projection which maps the input image vector $\mathbf{x}$ (expanded from an image matrix and mean-subtracted) first into the face-subspace $\mathbf{y}$, and then into the classification space $\mathbf{z}$:

$$\mathbf{y} = \Phi^T \mathbf{x} \tag{13}$$
$$\mathbf{z} = W_y^T \mathbf{y} \tag{14}$$
$$\mathbf{z} = W_x^T \mathbf{x} \tag{15}$$

where Φ is the PCA transform, W_y is the best linear discriminating transform in the PCA feature space, and $W_x = \Phi W_y$. After this composite linear projection, classification is performed in the classification space based on the weighted distance measure guided by the LDA eigenvalues. Finally, recognition is performed by comparing the input face images with mean images of known persons. When the class labels are not known for the stored images, the system performs face image matching with the nearest neighbor rule, which is the case for the FERET test [14].

In the implementation of PCA, there are two ways to compute the eigenvalues and eigenvectors: SVD decomposition and regular eigen-computation. For efficient ways to compute or update the SVD, please refer to [21, 20]. In many cases, even though the matrix is a full-rank matrix, the large condition number will create a numerical problem. One way around this is to compute the eigenvalues and eigenvectors for $\Sigma + \kappa I$ instead of C, where κ is a positive number such that $\Sigma + \kappa I$ is strictly positive [23].

We argued earlier that the face subspace dimension is fixed for a given training set, hence there should exist a universal face subspace with a fixed dimension if the training set has an infinite number of samples. However, for a specific face recognition task with a varying number of training samples, the dimension of the obtained subspace may vary. Even so, for a specific face recognition task with a small number of classes, say, five persons, it may still be better to use the approximate universal subspace obtained with large number of samples. But to obtain the LDA projection, we may wish to use only the face images of the five persons.

After we obtain the subspace from the training samples, each input can be reconstructed through linear combinations of the eigenvectors with the coefficients being the projection coefficients. Then the reconstructed inputs can be fed into the LDA classifier. But this procedure can be simplified to applying the classifier to the projection coefficients, as people usually do in practice. The correctness of this simplified procedure is guaranteed by the following lemma.

Lemma 1 *Let the Φ_i's be the column vectors of Φ with dimension n. Suppose a signal can be expressed as $\sum_{i=1}^{m} a_i \Phi_i$ ($m \leq n$), where the a_i's are the coefficients of the original signal projected into the subspace; then applying the classifier to the a_i's is equivalent to $\sum_{i=1}^{m} a_i \Phi_i$.*

Finally, the subspace LDA face recognition system is illustrated in Figure 4 where we pre-compute and pre-store the composite projection coefficients for the representative images of each class (called *gallery images*), and compute the coefficients for each testing image (also called *the probe image*) and compare them to the pre-stored coefficients to decide on the class label (person ID).

2.4.1 Distance Measure for Subspace LDA: Weighted or Unweighted?

For the pure LDA distance measure, it has been suggested that weighted Euclidean distance will give better classification than simple Euclidean distance in the context of face recognition [9], where the weights are the normalized eigenvalues defined in (5). But it turns out that this

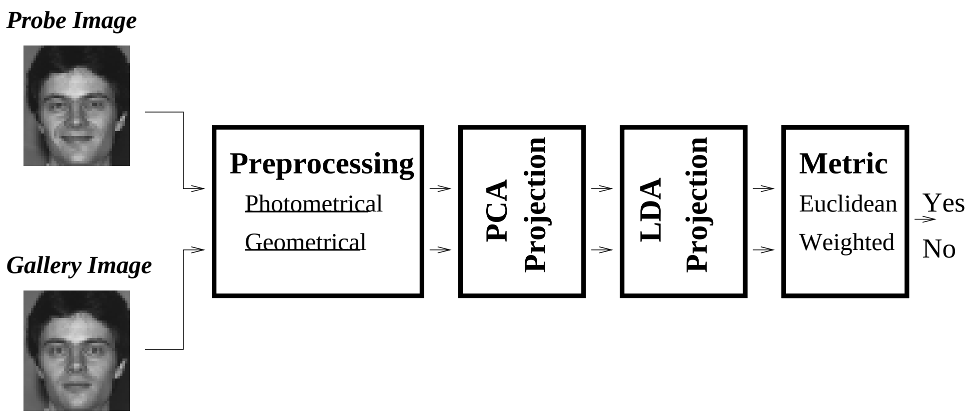


Figure 4: The subspace LDA face recognition system

weighted measure is sensitive to whether the corresponding persons have been seen before during the training stage and need some additional treatment [23]. Thus the preferred distance measure is the *adaptive* Euclidean distance, which is an eigenvalue-weighted distance when the testing samples are from the trained class and unweighted otherwise. In order to determine whether the input face images are from the trained class, we use a pre-processing step in which we compute the distance between the input image and its mirror-reflected image in the LDA projected space. If the distance is within a preset threshold the input image is declared to be within the trained classes; otherwise, it is declared as belonging to the untrained classes.

However, for the subspace LDA approach, the experiments suggest that the weighted Euclidean distance measure leads to bad performance, even for the testing samples from the trained class. Meanwhile, we believe that it is helpful to utilize the discrimination information revealed by LDA. Thus we apply a modified weighted distance measure where the weights are not directly the LDA eigenvalues, but rather regularized ones. This modified distance measure leads to a much better performance than both the weighted and unweighted ones. Figure 5 compares the normalized weights (LDA eigenvalues) with the modified weights we actually use in our subspace LDA approach. In Table 1 we also give a performance comparison of applying the subspace LDA approach to a testing set of 115 images with three different distance measures: unweighted, strictly weighted, and softly weighted. To represent all these distance measures in a unified manner, we use the following formula:

$$d(\mathbf{z}_1, \mathbf{z}_2) = \sqrt{\sum_{i=1}^{k} \tilde{\lambda}_i (z_{1,i} - z_{2,i})^2}, \qquad (16)$$

where $\mathbf{z}_i$ are the feature vectors obtained through (15) from the input images $\mathbf{x}_i$. Here $\tilde{\lambda}_i$ is the weight: 1 for the unweighted distance measure, eigenvalue λ_i for the strict distance measure, and the modified weight for the soft measure. Currently the modified weights are derived from the Numerical Recipe eigen-problem routines. We are planning to address the issue of constructing optimal modified weights in the near future.

3 Experiments

We present the details of a face recognition system based on the LDA of subspace features, tested using the FERET data consisting of about 3800 images and a mugshot dataset which

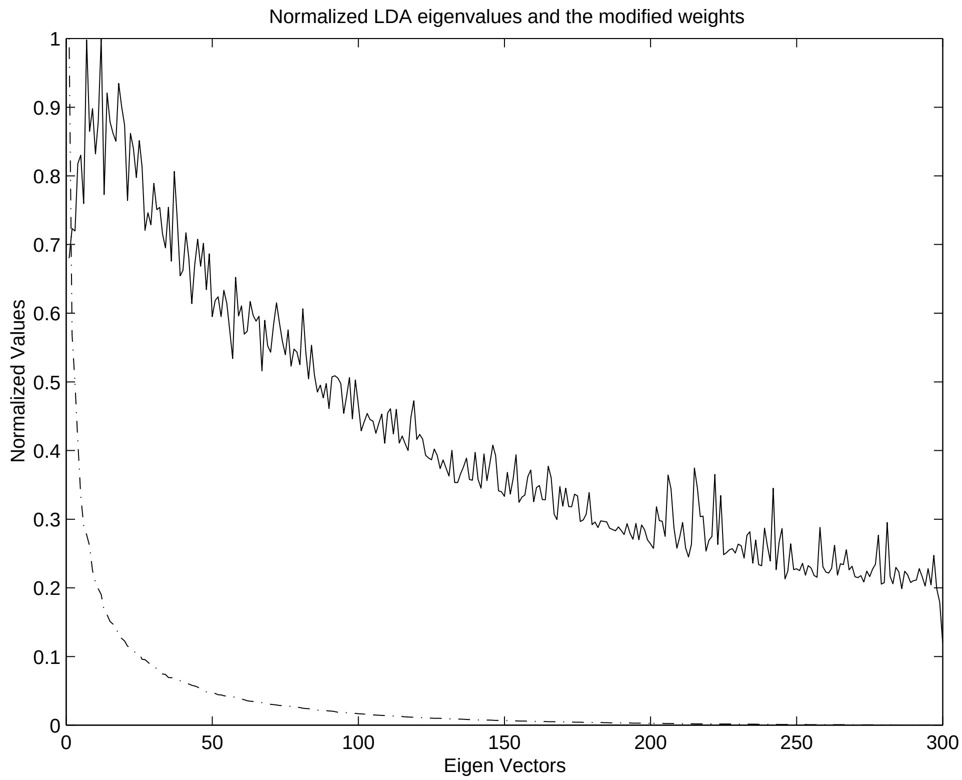


Figure 5: Comparison between the normalized LDA eigenvalues and the modified weights used in our subspace LDA approach. The dashed line denotes the eigenvalues associated with the eigenvectors, while the solid line denotes the modified weights.

Category	All (115)	Front-view (81)	Trained (61)
Strict	31.3	22.2	32.8
Unweighted	66.9	81.4	75.4
Soft	85.2	95.1	100

Table 1: Performance comparison (correct rate) using different distance metrics: the top row shows the performance using the strict (eigenvalue-based) weighted distance metric, the second row represents performance using the unweighted distance metric, while the last row is for results using the soft metric.

has about 2200 persons. The robust version [30] of this system has also been tested on an MPEG-7 content set [31]. For comprehensive performance comparison of this algorithm and other algorithms [9, 10, 7, 6, 25] on the FERET test, please refer to [14, 15].

For purposes of comparison, we have tested different algorithms on different face databases. The algorithms we tested are PCA [5], LDA (both weighted and unweighted), and subspace LDA. The databases we have used are the FERET development dataset [40], USC dataset [11], Olivetti dataset [6], and Stirling dataset [7]. Figure 6 shows some face images from the four different datasets.

Figure 6: Samples from the different face datasets we used: (1) FERET development dataset with more than 1000 images of size 384×256 (first two), (2) USC dataset with 36 images of size 128×128 (next two), (3) Olivetti dataset with 400 images of size 112×92, and (4) Stirling dataset with approximately 148 images of size 256×196 (last two).

3.1 Our Experiments

To process the face images, we manually locate the eyes and then perform geometric and intensity normalization. Normalization in intensity is done using either histogram equalization or the zero-mean-unit-variance operation. Geometric normalization consists of rotating, translating, and scaling the images so that the eyes are in fixed positions and all the normalized images are of the same size.

PCA vs. LDA

We first conducted experiments on the Olivetti face database which has 40 persons with 10 images per person. The normalized image size was chosen to be 37×30 and the image intensity normalization was done using histogram equalization. The results are given in Table 2. A few conclusions can be drawn from this experiment conducted using this relatively small face database:

1. Adding mirror/noisy images to the training set is good for an LDA classifier, but is a bad idea for the PCA classifier.

2. Both weighted and regular LDA perform better than PCA, while weighted LDA is better than regular LDA. (Notice here that the issue of testing samples from non-trained classes does not exist.)

Obtaining the Subspace

We then mainly worked on the FERET development dataset for training and used different face datasets for testing. The normalized image size is chosen to be 48×42. This is because the

[5] Our experiment suggests that weighted PCA is not a good choice; thus we only apply unweighted PCA. For a comprehensive analysis of PCA algorithms with different distance measures, please refer to [16].

[6] Internet Address: http://www.cam-orl.co.uk/facedatabase.html

[7] Internet address: http://pics.psych.stir.ac.uk/

M	N_T	Intensity Preprocessing	Set-Augmentation	PCA (%)	LDA (%)	W. LDA (%)
5	3	None	No	52	64	82
5	3	None	Yes	48	82	94
5	3	Histogram	No	60	66	90
5	5	None	No	66	94	96
5	5	None	Yes	28	98	100
5	5	Histogram	No	72	96	98
5	10	None	No	64	100	100
5	10	None	Yes	30	80	100
5	10	Histogram	No	70	100	100
10	5	None	No	51	75	97
10	5	None	Yes	21	85	99
20	5	None	No	41.5	75	98.5
20	5	None	Yes	31	63.5	96
20	10	None	No	36	90	100
20	10	None	Yes	42	90	100
40	5	None	No	35.25	62.75	92.75
40	5	None	Yes	31.5	56	95
40	5	Histogram	Yes	21.5	45.75	96.25

Table 2: Performance comparison of different classifiers based on the Olivetti dataset: PCA, unweighted LDA and weighted LDA. M represents how many persons (classes) out of the total of 40 persons used in the test, while N_T represents the number of samples per person out of the total of 10 used in the training. Set-augmentation means adding mirror/noisy images for training. The recognition rates for all classifiers are obtained from classifying all 10 samples in each trained class.

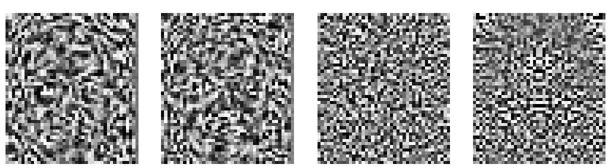
dimension of the subspace is around 300 which is far less than $48 \times 42 = 2016$. Basically we can work with even smaller images such as 24×21 because the subspace dimensionality appears not to be sensitive to the image size [27]. Further, similar performance was observed when image sizes ranging from 96×84, 48×42, 24×21, to 19×17 were used in our experiments.

To obtain the face subspace, we used 1038 original FERET images from 444 classes and increased the size of the training set to 4532 by adding mirror, noisy, and mirror plus noisy images. Then we retained eigenvectors corresponding to the top 300 eigenvalues, based on the observation that the higher-order eigenvectors do not look like a face (Figures 7, 8, and see Figure 9) [8]. *A wrong choice of this number will result in bad performance.* We have tested the algorithm that performs subspace LDA using the first 15 eigenvectors and 1000 eigenvectors on both the USC and Stirling datasets. Both choices produced lower scores while the latter choice did better than the pure LDA algorithm. To visualize the statistical effects of the LDA classifier, we have computed the between-class Euclidean distances among all 444 classes; the histograms of these distances are plotted in Figure 10. It can be seen that after the LDA projection, the distances are more meaningful for classification purposes (with a sharper histogram).

The average face and first four eigenfaces

Eigenfaces 15, 100, 200, 250, 300

Figure 7: Useful eigenfaces



Eigenfaces 400, 450, 1000, 2000

Figure 8: Suspicious eigenfaces: statistically insignificant.

Subspace LDA

We also conducted experiments using pure PCA, pure LDA (with adaptive Euclidean distance) and subspace LDA based on the FERET development dataset for training and testing, and other datasets for testing as well. All the experiments conducted here are similar to the FERET test: we have a gallery set and a probe set, and for each image in the probe set a rank ordering of all the images in the gallery set is produced. The cumulative match scores in

[8] This criterion is quite different from the energy criterion (eigenvalue) which is commonly used by many people.

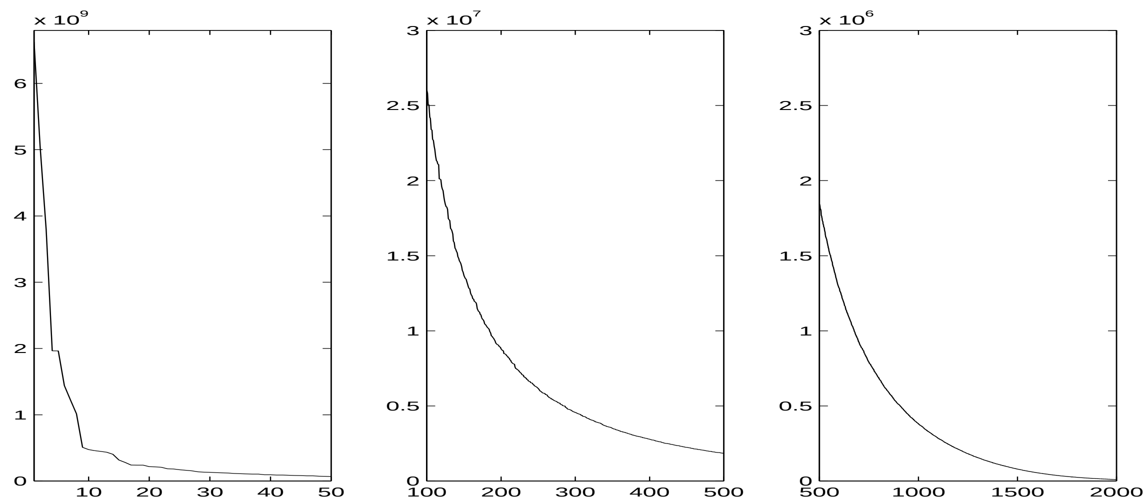


Figure 9: Eigenvalues in three segmentations: Since the first-last eigenvalue ratio is $8.39e^5$ it is hard to visualize all parts of the plot if it is plotted as a whole. As can be seen, the value drops to 0.02 from the first eigenvalue to the 30th eigenvalue. From the Principal Component point of view, the first 30 are enough, while the ratio of the 300th eigenvalue to the first eigenvalue is $6.88e^{-4}$.

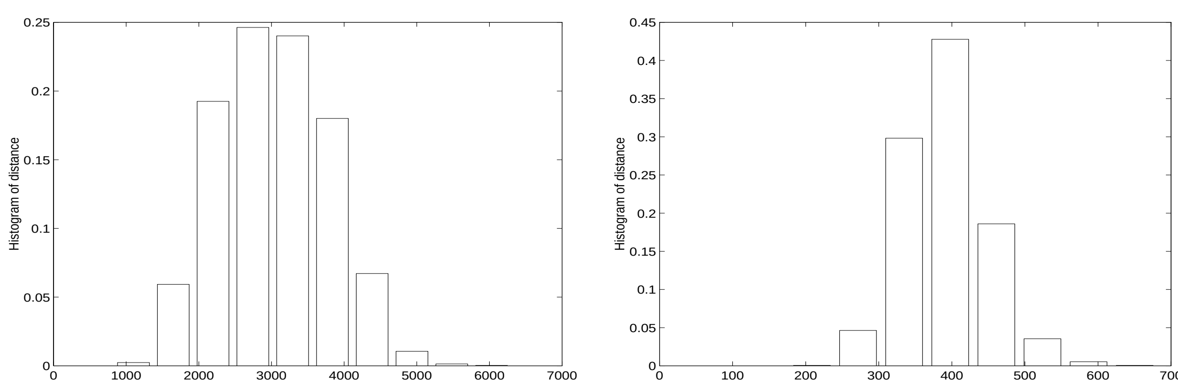


Figure 10: The Euclidean distance histograms for the FERET dataset of 444 classes: the left plot shows the widespread distance histogram computed from the original face images, the right plot shows the much concentrated distance histogram computed from the LDA projection coefficients.

Figures 11,12 are computed the same way as in the FERET test [14]. In our experiments, the gallery set contains 738 images, with 721 from the FERET training set and 17 from the USC dataset. The probe set has 115 images with 78 images in the training set (24 persons), 18 images from the FERET data set but not trained on (9 persons, also rotated out of the image plane), and 19 images (13 persons) from the USC dataset. These 115 testing face images correspond to a total of 46 persons. Except for LDA, we conducted experiments using 15, 300 and 1000 eigenvectors for PCA and subspace LDA and we found that the performance ordering is $subLDA300 > subLDA1000 > LDA > PCA1000 \geq PCA300 > subLDA15 > PCA15$ (Figure 11).

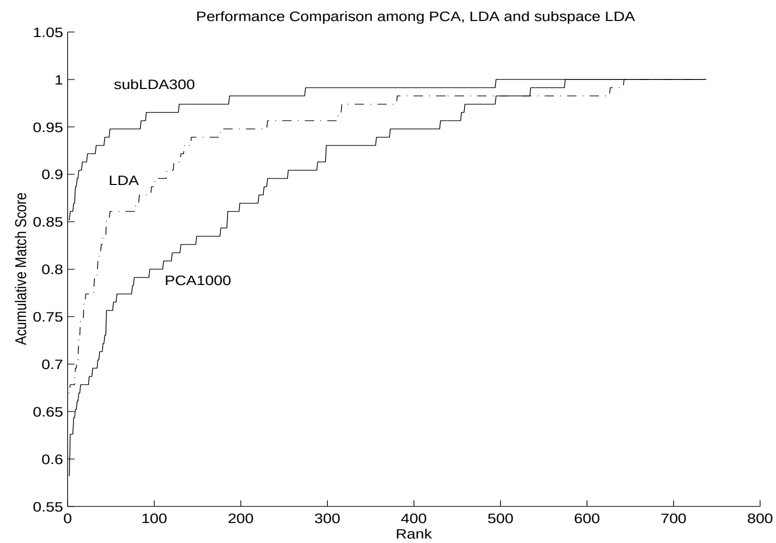
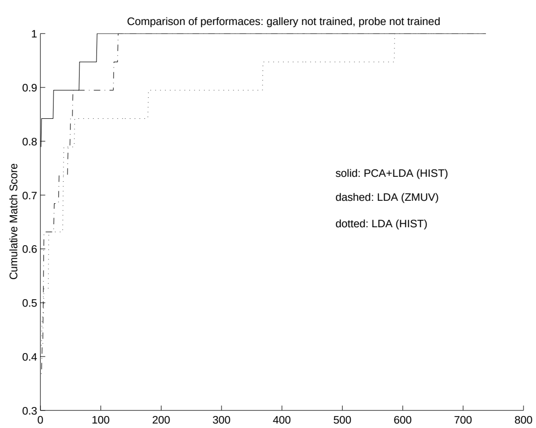
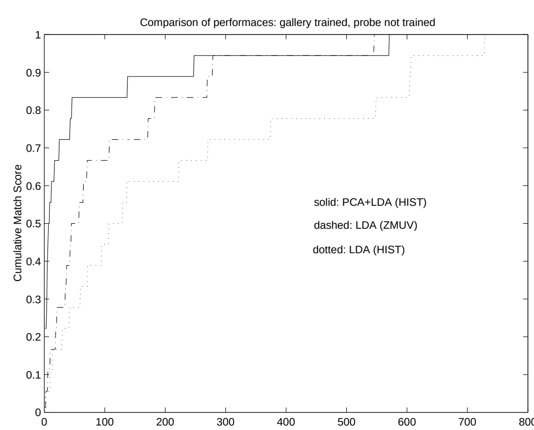


Figure 11: Performance comparison among PCA, LDA and subspace LDA. Three performance curves representing the best of three algorithms are plotted here: $subLDA300$, LDA, and $PCA1000$.



(a)



(b)

Figure 12: (a) Performance comparison on the 19 images from the USC dataset, (b) Performance comparison on the 18 images from the FERET dataset not included in the training set.

For these 78 trained images, both systems (LDA and subspace LDA) perform perfectly even though most of these images do not appear in the gallery set. But for the other 18 and 19 images from the FERET and USC datasets, the performances of these two methods are quite different. Figure 12 shows the performance comparison between pure LDA with different preprocessing techniques and subspace LDA with histogram equalization preprocessing.

In addition to the above experiments, we also conducted a sensitivity test of our system. We took one original face image, and then electronically modified the image by creating occlusions, applying Gaussian blur, randomizing the pixel location, and adding an artificial background. Figure 13 shows some electronically modified face images which were correctly identified. The tolerance to image noise/image distortion is a consequence of PCA's reconstruction power. We reconstructed the modified images using 300 PCA projection coefficients of the images; the results are shown in Figure 14.

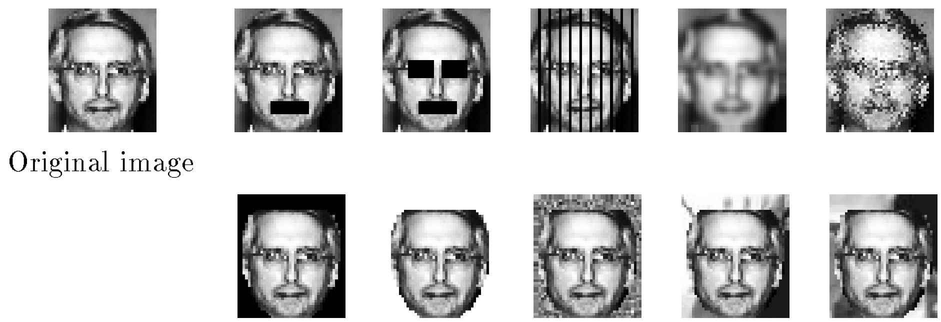


Figure 13: Electronically modified images which have been correctly identified.

Figure 14: Reconstructed images using 300 PCA projection coefficients for these electronically modified images.

3.2 FERET Test

Although we are not one of the participants in the FERET program, we agreed to take the FERET test in September 1996 to test the efficacy of the pure LDA approach. The gallery and probe datasets had 3323 and 3816 images respectively. Thus for each image in the probe set we produced a set of 3323 ordered images from the gallery set. A detailed description of the FERET test can be found in [14]. In March 1997, we re-took the FERET test to test the effect of different intensity preprocessing for LDA and mainly to test the improvement due to subspace LDA. Figure 15 shows significant improvement of the subspace LDA approach over

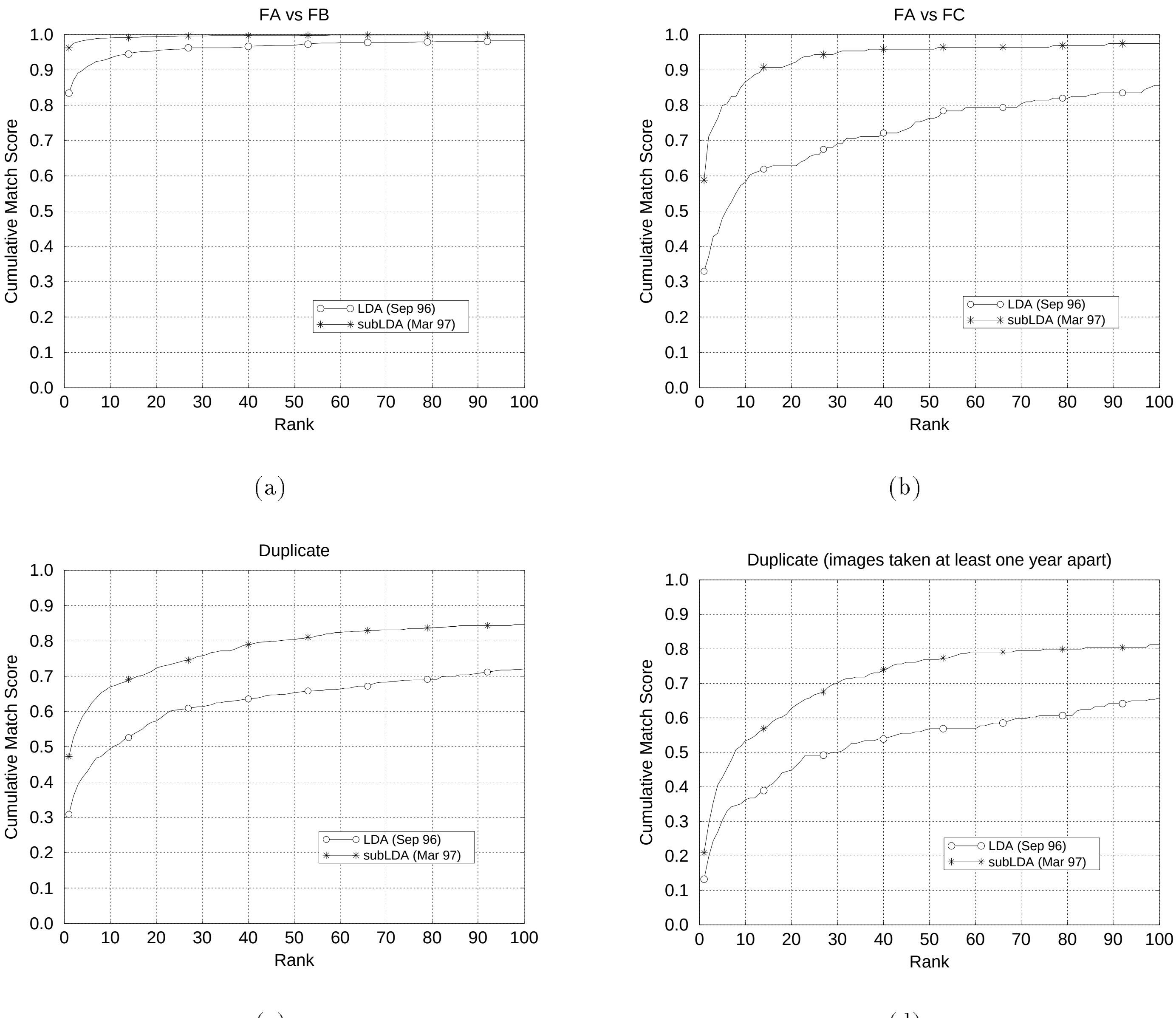


Figure 15: FERET test results from September 96 and March 97: (a)FA vs FB, (b)FA vs FC, (c)Duplicate, (d)Duplicate (images taken at least one year apart) (*Courtesy of Army Research Laboratory*)

LDA in every category [9]. As used in the FERET test, the x-axes in these sub-figures represent the *rank* of the ordering of the gallery from a match with a probe, while the y-axes represent the fraction of the probes correctly identified in terms of *top N* match. More recently, some preliminary results show that our system's performance on the task of *person verification* is very competitive [15].

3.3 Mugshot Dataset

One potential application for face identification algorithms is the electronic mugbook. In an electronic mugbook the gallery consists of digital mugshots of known people and the probe is a digital mugshot of an individual to be identified. In this experiment we give performance results on mugshot data provided by a law enforcement agency.

In this experiment, the gallery consisted of digital mugshots of 2175 persons with one frontal image per person (side mugshots were not collected). The probe set consisted of 2126 individuals with one frontal image per person. The two images of each person were taken within a few minutes of each other; however, the subjects were not necessarily cooperative.

We ran both PCA and subspace LDA on this dataset. In both implementations, the images were placed in a standard position with the eye locations fixed, and the pixel values were processed by a histogram equalization algorithm. The eigenfaces were trained on a subset of 500 images, and the faces were represented by the first 200 eigenvectors. In addition, the face images have been masked for the PCA classifier.

The results are presented in Figure 16 on a cumulative match plot as in the FERET test. Though subspace LDA produces a much better result than PCA, the result is not very im-

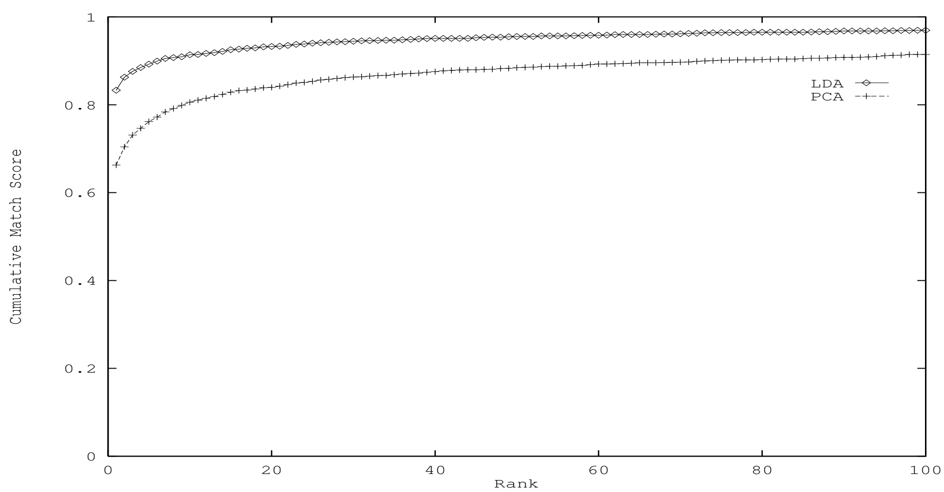


Figure 16: Performance comparison of PCA and subspace LDA on the mugshot dataset.

pressive. One possible reason is that the subspace is obtained by training on only 500 face images, hence is hardly representative and the choice of subspace dimension is made mainly to accommodate the PCA classifier (retaining the top 40% of the eigenvectors). Hence it does not satisfy the requirements for applying subspace LDA. Also in applying subspace LDA, we did

[9] Even though the zero-mean-unit-variance preprocessing showed better results for the pure LDA approach than histogram equalization for the experiment reported in Figure 12, the FERET test showed inferior performance. The plots here are only for the histogram equalization preprocessing case.

not apply any face masking step to improve the performance as in all the other experiments (except the MPEG-7 query test).

3.4 MPEG-7 Query Test

In this section, we briefly discuss the application of our face recognition algorithm to the development of MPEG-7 standards [33]. We have recently submitted a proposal titled *Descriptor for Human Face Image Objects in Multimedia Databases* to MPEG-7 using the subspace LDA method [31].

The performance of this proposed descriptor for retrieval of face image objects in a database was evaluated using the MPEG-7 Test Content Set S4 [32]. This set contains a total of 178 face images obtained from 14 different persons (classes) [10]. Of the 178 images, 140 views are frontal views, and the remaining 38 are non-frontal views (rotated out of the image plane) of faces. After being given a query input/image, the querying procedure usually consists of two key components: (1) processing of the input/image to obtain a representation of the input/image, and (2) a similarity measure for the item retrieved and ranking of the retrieved items.

3.4.1 Representation of the Face Image Object

Many databases contain images and video of humans, and especially photographs containing human faces. In many applications, it is important to be able to use a query face image to retrieve from these databases those multimedia data that contain faces similar to the query. This capability is facilitated (more efficiently and accurately) by storing/associating a standard representation (called a descriptor) with each instance of a human face in the database. We propose using the subspace LDA projection coefficients as the descriptor of a face image object. Recall from (15) that the descriptor is basically the vector $\mathbf{z}$ if the image is $\mathbf{x}$. So the full representation for a database is the PCA projection matrix Φ, the LDA projection matrix W, and the vector $\mathbf{z}$ for each face image. The PCA projection Φ (with dimension 2016×300) was computed from the 1038 FERET images. However, the LDA projection W was calculated from the available MPEG-7 images. The MPEG-7 S4 database contains only 14 persons/classes; thus the dimension of W is 300×13. Also, we applied a face mask and illumination compensation to compute $\mathbf{z}$ in addition to geometrical normalization [11] and histogram equalization [30].

3.4.2 Similarity Measure

After we compute the face descriptor, we can compare the descriptor value with the query image against descriptor values associated with each database image based on some similarity measure. The similarity measure we used is the weighted distance measure guided by LDA eigenvalues described in Section II-D.1. More specifically, assuming $\mathbf{q}$ is the query image and $\mathbf{x}$ is an image in the database, and $\mathbf{z}_q$, $\mathbf{z}$ are the corresponding descriptors, the similarity measure is defined as follows:

$$s(\mathbf{z}_q, \mathbf{z}) = \sqrt{\sum_{i=1}^{k} \tilde{\lambda}_i (z_{q,i} - z_i)^2}, \tag{17}$$

where $\tilde{\lambda}_i$ is a value related to the LDA eigenvalue λ_i, and k is 13 for the MPEG-7 S4 test set.

[10] The original images are color images (JPEG); we converted them into black-white in our experiment.

[11] Currently this is done manually.

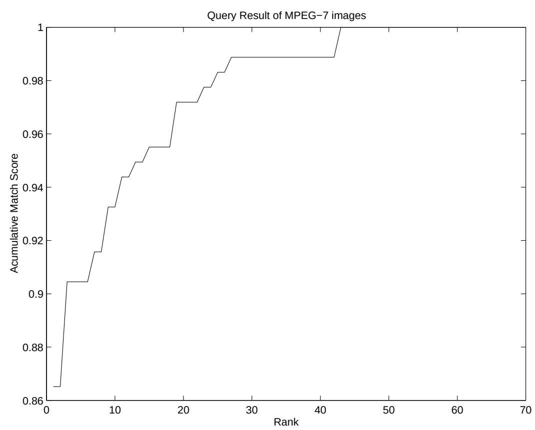
3.4.3 Query Experiment

Computing W

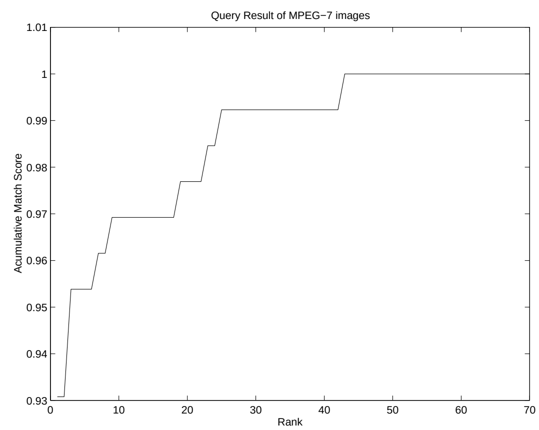
For each of the 14 classes, 5 images were selected (for a total of 70 images) for computing the matrix W. One of each set of 5 images was selected to be a non-frontal view.

Image Query (Two scenarios)

- **Full Querying** All 70 images used in the LDA training stage were stored in the database. Each of the available 178 images was used as a query image, and retrieval from the database was performed using subspace LDA. The set of retrieved images for each query image was examined; the images in each retrieved set were ordered based on the weighted LDA distance measure. Using the criterion that the top-ranked retrieved image must correspond to the correct class, an overall correct retrieval rate of 86.5% was observed. Using the criterion that one of the best three retrieved images must belong to the correct class, a correct retrieval rate of 90.4% was observed. The cumulative match curve for this test case is shown in Figure 17(a).



(a)



(b)

Figure 17: (a) Query performance on the 178 MPEG7 images. (b) Query performance comparison on the 130 frontal-view MPEG7 images.

- **Frontal-view Querying** Again all 70 training images were stored in the database. Each of the available 130 frontal-view images was used as a query image, and retrieval from the database was performed using subspace LDA. Using the criterion that the top-ranked retrieved image must correspond to the correct class, an overall correct retrieval rate of 93.1% was observed. Using the criterion that one of the best three retrieved images must belong to the correct class, a correct retrieval rate of 95.4% was observed. The cumulative match curve for this test case is shown in Figure 17(b).

Some query images and the retrieved database images are displayed in Figure 18.

4 Discussion and Conclusions

We have presented a face recognition system based on subspace LDA. We have shown that LDA is a good choice for the face recognition/verification problem with very limited training samples per class. For the specific task of face recognition, we utilized the existence of a face subspace.

Figure 18: Query examples for the MPEG-7 image database with each row being a query example: the first column represents the query image, and the following columns are the retrieved images from the database.

By adopting the concept of a face subspace, we showed that subspace LDA provides a very good solution to the *generalization/overfitting problem* for new test samples from the trained classes and samples from non-trained classes. Using an eigenvalue-guided distance metric, the improved performance of subspace LDA over other classifiers is demonstrated through extensive experiments on various face datasets. The improved performance of the subspace LDA face recognition system has also been demonstrated using the FERET test. More recently, the efficiency of subspace LDA has been demonstrated in an MPEG-7 query test.

4.1 Discussion

Despite the successes of the described subspace LDA face recognition system, some issues still need to be addressed.

First of all, there is an issue regarding the extraction of the subspace via PCA when the SNR (signal noise ratio) is not high enough. In an extremely bad situation, the whitening procedure will produce a wrongly-ordered matrix, and we could choose a value of m so that noise signals are included in the subspace while real signals are discarded. The choice of a perfect m becomes more difficult when the sample number is quite limited compared to the space dimensionality because the covariance matrix is perturbed. One practically useful approach is to look at the bases (eigenvectors) when the associated eigenvalues are very small compared to the major eigenvalues. For example, when the subspace is a face-space, then the useful eigenvectors are face-like images. When the eigenvalues are very small and the eigenvectors do not embed any face-related information, they should be discarded (Figures 7, 9, 8).

Second, the current system cannot handle inputs with significant illumination changes. We tested our system on the Weizmann face database from the Weizmann Institute [12] and found that it is sensitive to significant illumination change. To handle illumination variations, researchers have proposed various methods to solve this fundamental and hard problem. With the assumption of Lambertian surfaces without shadowing and three aligned images under different lighting directions per face, a 3D linear subspace of the high-dimensional image space can be constructed for a fixed viewpoint [34, 12, 38, 39]. Thus recognition based on the 3D linear subspace would be perfectly carried out under ideal assumptions for images under different illuminations. Nevertheless, for a limited training and testing database, LDA turns out to be a better alternative [34]. When multiple images are not available, one possible solution is to apply shape from X techniques to estimate the lighting direction and the 3D surface of the object from just one image. But it is unrealistic to say that these techniques may work well enough under non-ideal conditions. Within the eigen-subspace domain, it has also been suggested that by discarding the three most significant principal components, variations due to lighting can be reduced. However in order to maintain the system performance for normally lighted images and improve the performance for images under variable illumination, we must make an unrealistic assumption: the first three principal components *only* capture the variations due to lighting. In general, handling illumination has proven to be a difficult problem for face recognition [4].

Third, the current system is not very good at handling face images rotated out of the image plane. One reason is that in our training set, very few rotated images are available (only 26 out of a total of 1038 images). Also, heavily rotated images make the rationale of applying subspace LDA questionable.

Finally, the current system assumes that face detection and segmentation are perfect, which may not be true in reality. We deliberately fed the system some poorly segmented/normalized face images and found the system to be sensitive to segmentation/detection errors.

[12] Internet address: ftp://eris.wisdom.weizmann.ac.il/pub/FaceBase/

4.2 Future Directions

We are currently addressing the issues of illumination change based on heuristics, and poor segmentation/normalization based on the concept of 2D distorted subspaces [30].

Future work in this area includes extending the methodology to other image classification domains and investigating the possibility of combining different subspace features such as a wavelet based decomposition with LDA [29]. We also need to develop new methods other than PCA to effectively extract the subspace in the case of low SNR. For example, Black *et al.* described elegant work on how to reformulate the standard mean square problem as a robust estimation problem [41]. Finally, for the challenging illumination problem, we are looking into some recent work by Jacobs, Georghiades and others [36, 37].

Acknowledgments

We would like to thank Saad Sirohey for participating the FERET test, and Hyeonjoon Moon for providing us the FERET test plots. We also thank Dr. Qinfen Zheng for insightful discussion. It has been a pleasant experience to work with Dr. N. Nandhakumar and D. Bhat at LGERCA on our MPEG-7 proposal. Finally we thank the Heinrich Hertz Institute of Germany for providing the MPEG-7 content set S4.

References

[1] R. Brunelli and T. Poggio, "Face Recognition: Features versus Templates," *IEEE Trans. on PAMI*, Vol. 15, pp. 1042-1052, 1993.

[2] R. Chellappa, C.L. Wilson and S. Sirohey, "Human and Machine Recognition of Faces, A Survey," *Proc. of the IEEE*, Vol. 83, pp. 705-740, 1995.

[3] K. Fukunaga, *Statistical Pattern Recognition,* New York: Academic Press, 1989.

[4] Y. Adini, Y. Moses, and S. Ullman, "Face Recognition: The problem of Compensating for Changes in Illumination Direction," *IEEE Trans. on PAMI*, Vol. 19, pp. 721-732, 1997.

[5] L. Sirovich and M. Kirby, "Low-dimensional Procedure for the Characterization of Human Faces," *Journal of the Optical Society of America A,* Vol. 4, pp. 519-524, 1987.

[6] M. Turk and A. Pentland, "Eigenfaces for Recognition," *Journal of Cognitive Neuroscience*, Vol. 3, pp. 72-86, 1991.

[7] B. Moghaddam and A. Pentland, "Probabilistic Visual Learning for Object Representation," *IEEE Trans. on PAMI*, Vol. 19, pp. 696-710, 1997.

[8] N. Costen, I. Craw, T. Kato, G. Robertson and S. Akamatsu, "Manifold Caricatures: On the Psychological Consistency of Computer Face Recognition," in *Proc. Second International Conference on Automatic Face and Gesture Recognition*, pp. 4-9, Killington, Vermont, 1996.

[9] K. Etemad and R. Chellappa, "Discriminant Analysis for Recognition of Human Face Images," *Journal of the Optical Society of America A*, Vol. 14, pp. 1724-1733, 1997.

[10] D.L. Swets and J. Weng, "Using Discriminant Eigenfeatures for Image Retrieval," *IEEE Trans. on PAMI*, Vol. 18, pp. 831-836, 1996.

[11] B.S. Manjunath, R. Chellappa and C.V.D. Malsburg, "A Feature Based Approach to Face Recognition," in *Proc. Conference on Computer Vision and Pattern Recognition*, pp. 373-378, Urbana-Champaign, IL, 1992.

[12] S.K. Nayar, S. A. Nene, and H. Murase, "Subspace Methods for Robot Vision," Technical Report CUCS-06-95, Columbia University, New York, 1995.

[13] E. Oja, *Subspace Methods of Pattern Recognition*, Hertfordshire: Research Studies Press, 1983.

[14] P.J. Phillips, H. Moon, P. Rauss, and S.A. Rizvi, "The FERET Evaluation Methodology for Face-Recognition Algorithms," in *Proc. Conference on Computer Vision and Pattern Recognition*, pp. 137-143, San Juan, PR, 1997.

[15] S.A. Rizvi, P.J. Phillips, and H. Moon, "The FERET Verification Testing Protocol for Face Recognition Algorithms," in *Proc. Third International Conference on Automatic Face and Gesture Recognition*, Nara, Japan, 1998.

[16] H. Moon and P.J. Phillips, "Analysis of PCA-based Face Recognition Algorithms," *Empirical Evaluation Techniques in Computer Vision,* K.W. Bowyer and P.J. Phillips, eds, IEEE Computer Society Press, Los Alamitos, CA, 1998.

[17] S.S. Wilks, *Mathematical Statistics,* New York: Wiley, 1962.

[18] G.W. Stewart and J.G. Sun, *Matrix Perturbation Theory,* San Diego: Academic Press, 1990.

[19] M.H. Kalos and P.A. Whitlock, *Monte Carlo Methods,* New York: Wiley, 1986.

[20] M. Gu and S.C. Eisenstat, "A Stable and Fast Algorithm for Updating the Singular Value Decomposition," Research Report DCR/RR-966, Yale University, New Haven, CT, 1994.

[21] S. Chandrasekaran, B.S. Manjunath, Y.F. Wang, J. Winkeler, and H. Zhang, "An Eigensapce Update Algorithm for Image Analysis,", *Graphical Models and Image Processing,* Vol. 59, pp. 321-332, 1997.

[22] W. Zhao, A. Krishnaswamy, R. Chellappa, D.L. Swets, and J. Weng, "Discriminant Analysis of Principal Components for Face Recognition," *Face Recognition: From Theory to Applications*, Eds. H. Wechsler, P.J. Phillips, V. Bruce, F.F. Soulie and T.S. Huang, Berlin: Springer-Verlag, pp. 73-85, 1998.

[23] W. Zhao, R. Chellappa, and A. Krishnaswamy, "Discriminant Analysis of Principal Components for Face Recognition," in *Proc. Third International Conference on Automatic Face and Gesture Recognition*, pp. 336-341, Nara, Japan, 1998.

[24] W. Zhao, R. Chellappa, and N. Nandhakumar, "Empirical Performance Analysis of Linear Discriminant Classifiers," in *Proc. Conference on Computer Vision and Pattern Recognition*, pp. 164-169, Santa Barbara, CA, 1998.

[25] L. Wiskott, J.-M. Fellous, and C. von der Malsburg, "Face Recognition by Elastic Bunch Graph Matching," *IEEE Trans. on PAMI*, Vol. 19, pp. 775-779, 1997.

[26] L. Wiskott, J.M. Fellous, N. Kruger, and C.v.d. Malsburg, "Face Recognition and Gender Determination," in *Proc. First International Workshop on Automatic Face and Gesture Recognition*, Zurich, Switzerland, 1995.

[27] W. Zhao, "Subspace Methods in Object/Face Recognition," Unpublished manuscript.

[28] W.H. Press *et al.*, *Numerical Recipes in C: The Art of Scientific Computing,* Cambridge University Press, 1992.

[29] W. Zhao and N. Nandhakumar, "Linear Discriminant Analysis of MPF for Face Recognition," in *Proc. International Conference on Pattern Recognition*, Brisbane, Australia, pp. 185-188, 1998.

[30] W. Zhao, "Improving the Robustness of Face Recognition," in *Proc. 2nd International Conference on Audio- and Video-based Person Authentication*, Washington DC, pp. 78-83, 1999.

[31] N. Nandhakumar, D. Bhat, W. Zhao, and R. Chellappa, "Descriptor for Human Face Image Objects in Multimedia Databases," MPEG-7 proposal P551.

[32] "Description of MPEG-7 Content Set", N2467, October 1998.

[33] "Call for Proposals for MPEG-7 Technology", N2469, October 1998.

[34] P.N. Belhumeur, J.P. Hespanha and D.J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Trans. on PAMI*, Vol. 19, pp. 711-720, 1997.

[35] P.J. Phillips and Y. Vardi, "Data-driven Methods in Face Recognition," in *Proc. First International Workshop on Automatic Face and Gesture Recognition*, pp. 65-69, Zurich, Switzerland, 1995.

[36] D.W.Jacobs, P.N. Belhumeur, and R. Basri, "Comparing Images under Variable Illumination," in *Proc. Conference on Computer Vision and Pattern Recognition*, pp. 610-617, Santa Barbara, CA, 1998.

[37] A.S. Georghiades, D.J. Kriegman, and P.N. Belhumeur, "Illumination Cones for Recognition Under Variable Lighting: Faces," in *Proc. Conference on Computer Vision and Pattern Recognition*, pp. 52-58, Santa Barbara, CA, 1998.

[38] P. Hallinan, "A Low-Dimensional Representation of Human Faces for Arbitrary Lighting Conditions," in *Proc. Conference on Computer Vision and Pattern Recognition*, pp. 995-999, Seattle, WA, 1994.

[39] A. Shashua, "On Photometric Issues in 3D Visual Recognition from a Single 2D image," *Int. Journal on Computer Vision*, Vol. 21, pp. 99-122, 1997.

[40] S. Gutta, J. Huang, D. Singh, I. Shah, B. Takacs, and H. Wechsler, "Benchmark Studies on Face Recognition," in *Proc. First International Workshop on Automatic Face and Gesture Recognition*, Zurich, Switzerland, 1995.

[41] M. Black and A. Jepson, "Eigentracking: Robust matching and tracking of articulated objects using a view-based representation," in *Proc. European Conference on Computer Vision*, pp. 329-342, Berlin: Springer, 1996.