Discriminant Analysis of Principal Components for Face Recognition

Wenyi Zhao[1], Arvindh Krishnaswamy[2], Rama Chellappa[1]
Daniel L. Swets[3], John Weng[4]

[1]Center for Automation Research, University of Maryland
College Park, MD 20742-3275, USA
Email:{*wyzhao, rama*}@*cfar.umd.edu*
[2]Electrical Engineering Dept, Stanford University
Stanford, CA 94305, USA
[3]Computer Science Department, Augustana College
Sioux Falls, SD 57197, USA
[4]Computer Science Dept, Michigan State University
East Lansing, MI 48824-1027, USA

Abstract. In this paper we describe a face recognition method based on PCA (Principal Component Analysis) and LDA (Linear Discriminant Analysis). The method consists of two steps: first we project the face image from the original vector space to a face subspace via PCA, second we use LDA to obtain a linear classifier. The basic idea of combining PCA and LDA is to improve the generalization capability of LDA when only few samples per class are available. Using FERET dataset we demonstrate a significant improvement when principal components rather than original images are fed to the LDA classifier. The hybrid classifier using PCA and LDA provides a useful framework for other image recognition tasks as well.

1 Introduction

The problem of automatic face recognition is a composite task that involves detection and location of faces in a cluttered background, normalization, recognition and verification. Depending on the nature of the application, e.g. sizes of training and testing database, clutter and variability of the background, noise, occlusion, and finally speed requirements, some of the subtasks could be very challenging. Assuming that segmentation and normalization haven been done, we focus on the subtask of person recognition and verification and demonstrate the performance using a testing database of about 3800 images.

There have been many methods proposed for face recognition. And one of the key components of any methods is facial feature extraction. Facial feature could be a gray-scale image, a low-dimensional abstract feature vector, and it could be either global or local. There are two major approaches to facial feature extraction for recognition, holistic template matching based systems and geometrical local feature based schemes [1]. The algorithm we present belongs to the first category.

2 LDA of Principal Components face recognition system

2.1 PCA and LDA

Principal Component Analysis is a standard technique used to approximate the original data with lower dimensional feature vectors [2]. The basic approach is to compute the eigenvectors of the covariance matrix, and approximate the original data by a linear combination of the leading eigenvectors. The mean square error (MSE) in reconstruction is equal to the sum of the remaining eigenvalues. The feature vector here is the PCA projection coefficients. PCA is appropriate when the samples are from one class or group(super-class). In real implementation, there are two ways to compute the eigenvalues and eigenvectors: SVD decomposition and regular eigen-computation. For efficient way to compute or update the SVD, please refer to [4, 3]. 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 $C + \kappa I$ instead of C, where κ is a positive number. This is based on the following lemma:

Lemma 1 *Matrices C and $C+\kappa I$ have same eigenvectors but different eigenvalues with the relationship: $\lambda_{C+\kappa I} = \lambda + \kappa$ as long as $\lambda + \kappa$ is not equal to zero.*

On the other hand, LDA produces an optimal linear discriminant function $\mathbf{f}(\mathbf{x}) = W^T\mathbf{x}$ which maps the input into the classification space in which the class identification of this sample is decided based on some metric such as Euclidean distance [17, 12, 13]. A typical LDA implementation is carried out via scatter matrices analysis [2]. We compute the within and between-class scatter matrices as follows:

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

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

Here S_w is the *Within-class Scatter Matrix* showing the average scatter Σ_i of the sample vectors $\mathbf{x}$ of different class C_i around their respective mean $\mathbf{m_i}$:

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

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

Various measures are available for quantifying the discriminatory power [2], the commonly used one being,

$$\mathcal{J}(W) = \frac{\| W^T SwW \|}{\| W^T SbW \|}. \tag{4}$$

Here W is the optimal discrimination projection and can be obtained via solving the generalized eigenvalue problem [10]:

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

The distance measure used in the matching could be a simple Euclidean, or a weighted Euclidean distance. It has been suggested that the weighted Euclidean distance will give better classification than the simple Euclidean distance [8], where the weights are the normalized versions of the eigenvalues defined in (5). But it turns out that this weighted measure is sensitive to whether the corresponding persons have been seen during the training stage or not. To account for this, we devised a simple scheme to detect whether the person in the testing image has been trained or not and then use either a weighted Euclidean distance or a simple Euclidean distance respectively.

2.2 LDA of Principal Components

Both PCA and LDA have been used for face recognition [5, 6, 7, 8, 15, 16, 11]. With PCA, the input face images usually needed to be warped to a standard face because of the large within-class variance [6, 7]. This preprocessing stage reduces the within-class variance dramatically, thus improving the recognition rate.

We first built a simple system based on pure LDA [8], but the performance was not satisfactory on a large dataset of persons not present in the training set. The idea of combining PCA and LDA has been previously explored by Weng *et al* [15].

Although the pure LDA algorithm does not have any problem discriminating the trained samples, we have observed that it does not perform very well for the following three cases:

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

Basically this is a generalization 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 2 or 4 samples per class!

Combining PCA and LDA, we obtain a linear projection which maps the input image $\mathbf{x}$ first into the face-subspace $\mathbf{y}$, and then into the classification space $\mathbf{z}$:

$$\mathbf{y} = \Phi^T \mathbf{x} \tag{6}$$

$$\mathbf{z} = W_y^T \mathbf{y} \tag{7}$$

$$\mathbf{z} = W_x^T \mathbf{x} \tag{8}$$

where Φ is the PCA transform, W_y is the best linear discriminating transform on PCA feature space, and W_x is the composite linear projection from the original image space to the classification space. After this composite linear projection, recognition is performed in the classification space based on some distance measure criterion.

3 Experiments

To process the face images, we manually locate the eyes and then perform geometric normalization with the eye locations fixed and perform intensity normalization, histogram equalization or zero mean unit variance. The normalized image size is chosen to be 48×42 since similar performance has been observed with the image size 96×84 in our experiments.

To obtain the principal components, we used 1038 FERET images from 444 classes (These images are so-called *training* set which was distributed to participants prior to the FERET test, the *gallery* set and the *probe* set were either constructed by ourselves for our own experiments or distributed during the test for the FERET test). 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 3, 4). A wrong choice of this number will result in bad performance. We have tested the algorithm that performs LDA on principal components using the first 15 eigenvectors and 1000 eigenvectors on both USC dataset and Stirling dataset. Both choices produced lower scores while the latter choice did better than the pure LDA algorithm. Since an *orthonormal linear* projection can be viewed as projection onto a set of bases, we can visualize these bases. Three different sets of bases from three different linear projections are shown here: (1) pure LDA projection W (figure 1), (2) pure PCA projection Φ (figure 3), and (3) PCA + LDA projection W_x (figure 2). All these bases are computed using the FERET *training* set, the PCA + LDA bases being based on the first 300 PCA bases.

3.1 Our experiments

All the experiments conducted here are similar to the FERET test: we have a gallery set and a probe set. In the prototyping stage, the weights that characterize the projections of images in the gallery set are computed. In the testing stage, the weights that characterize the projections of images in the probe set are calculated. Using these weights and the nearest-neighbor criterion, for each image in the probe set a rank ordering of all the images in the gallery set is produced. The cumulative match score in figure 6 is computed the same way as in FERET test [9].

3.1.1 Comparison of LDA and LDA of Principal Components

To test our system (figure 5), we constructed a gallery set which contains 738 images, with 721 from the FERET training set and 17 from the USC dataset [18].

The probe set has 115 images with 78 images in the training set, 18 images from the FERET data set but not trained, and 19 images from the USC dataset.

For the 78 trained images, both system works perfectly even though most of these images do no appear in the gallery set. But for the other 18 and 19 images from the FERET and USC datasets, the performance between these two methods is quite different.

Figure 6 shows the performance comparison between pure LDA with different intensity preprocessing and LDA of principal components with histogram equalization preprocessing.

3.1.2 Sensitivity test of LDA of Principal Components

In addition to the above experiments, we also conducted a sensitivity test. We took one original face image, and then electronically modified the image by creating occlusions, applying Gaussian blur, randomizing the pixel location, and adding artificial background. Figure 7 shows the various electronically-modified face images which have been correctly identified.

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. The detailed description of the FERET test can be found at [9]. In March 1997, we re-took the FERET test to test the effect of different intensity preprocessing for LDA and also to test the improvement due to LDA of Principal Components. Figure 8 shows a significant improvement of LDA of principal components approach over LDA in every category [1]. More recently, some preliminary results show that our system's performance for the task of *person verification* is very competitive.

3.3 Faces and Other Objects Combined

In order to test the performance when objects include more than faces, we experimented with image database that include human faces as well as other natural objects. The face part used in this combination test was organized by individual; each individual had a pool of images from which to draw training and test data sets. Each individual had at least two images for training with a change of expression. The images of 38 individuals (182 images) came from the Michigan State University Pattern Recognition and Image Processing laboratory. Images of individuals in this set were taken under uncontrolled conditions, over several days, and

[1] Even though the zero-mean-unit-variance preprocessing showed better results for pure LDA approach than histogram-equalization on the experiment reported in figure 6, the FERET test showed inferior performance. The plots here are only for the histogram-equalization preprocessing case.

No. of training images	1316 from 526 classes
No. of test images	298 from 298 classes
No. nodes in the tree	2388
No. of explored paths	10
Top one	95.0%
Top 10	99.0%

Table 1: Summary of experiment for faces and other objects.

under different lighting conditions. 303 classes (654 images) came from the FERET database. All of these classes had at least two images of an individual taken under controlled lighting, with a change of expression. 24 of these classes had additional images taken of the subjects on a different day with very poor contrast. Sixteen classes (144 images) came from the MIT Media lab under identical lighting conditions (ambient laboratory light). Twenty-nine classes (174 images) came from the Weizmann Institute, and are images with three very controlled lighting conditions for each of two different expressions. The nonface objects includes a wide range of scenes, ranging from street signs to aerial photographs. A small sample of images from the classes learned is given in Figure 9. The views differ in expression, viewing angle, lighting, etc.

Most classes in the database were represented by two images, and 19% of the classes had three or more images, up to twelve for some objects (e.g., fire hydrant). Each image consisted of a well-framed object of interest. The different images from each class were taken either in a different setting or from a different angle; where possible a change in the lighting arrangement was used to provide variation in the training images.

Following training, the PCA+LDA method was tested using a test set completely disjoint from the training set of images. In this version, the tree index method explained in [14] was also used. At each node of the tree, a new projection matrix is computed based the training samples belonging to the node. Each projection matrix represents PCA projection followed by LDA projection. The subspace created by the projection matrix is used to determine the further partition of the sample space, one child node is assigned with the samples falling into the region represented by the child node. Such recursive partition is carried on for every node until the samples assigned to the node belong to a single class. In order words, the PCA+LDA projection is applied recursively to smaller and smaller sets of samples, and thus better separating classes while the number of classes become small deep down the tree. The data that show the improvement of recognition rate and the speed gain due to this recursive PCA+LDA tree partition can be found in [14]. A summary of the results using this tree-based PCA+LDA system is shown in Table 1.

In the experiments further conducted, we trained the system using training samples artificially generated from the original training samples to vary in (a) 30% of size, (b) positional shift of 20% of size and 20% of size; (c) 3D face orientation by about 45 degrees and testing with 22.5 degrees. Training and test data sizes are similar to that in Table 1. The top 1 and top 10 correct recognition rates were, respectively, (a) 93.3% and 98.9%, (b) 93.1% and 96.6%, (c) 78.9% and 89.4%.

4 Conclusions

We have presented in this paper a face recognition system which combines PCA and LDA. Performance improvement of this method over pure LDA based method is demonstrated through our own experiments and FERET test. We believe that by combining PCA and LDA, using PCA to construct a task-specific subspace and then applying LDA on that subspace, other image recognition systems such as fingerprint, optical character recognition can be improved.

Acknowledgments

We would like to thank Saad Sirohey for participating in the FERET test, and Hyeonjoon Moon for providing us the FERET test plots. Zhao and Chellappa are supported by the ONR MURI contract N00014-95-1-0521, under ARPA order C635. Weng's work was supported by NSF Grant No. IRI 9410741 and ONR grant No. N00014-95-1-0637.

References

[1] 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.

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

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

[4] S. Chandrasekaran, B.S. Manjunath, Y.F. Wang, J. Winkeler, and H. Zhang, "An Eigensapce update algorithm for image analysis,", to appear in the journal *Graphical Model and Image Processing*, 1997.

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

[6] B. Moghaddam and A. Pentland, "Probabilistic Visual Learning for Object Detection," in *Proc. International Conference on Computer Vision*, Boston, MA, 1995, pp. 786-793.

[7] 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, Killington, Vermont, 1996, pp. 4-9.

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

[9] P.J. Philips, H. Moon, P. Rauss, and S.A. Rizvi, "The FERET Evaluation Methodology for Face-Recognition Algorithms," in *Proc. of Computer Vision and Pattern Recognition*, Puerto Rico, 1997, pp. 137-143.

[10] S.S. Wilks, *Mathematical Statistics,* New York: Wiley, 1963.

[11] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," in *Proc. European Conf. on Computer Vision*, April 1996.

[12] R.A. Fisher, "The statistical utilization of multiple measurements," *Annals of Eugenics*, 8:376–386, 1938.

[13] A.K. Jain and R.C. Dubes, *Algorithms for Clustering Data*, Prentice-Hall, New Jersey, 1988.

[14] D.Swets and J.Weng, "Discriminant analysis and eigenspace partition tree for face and object recognition from views," in *Proc. Int'l Conference on Automatic Face- and Gesture-Recognition*, pages 192–197, Killington, Vermont, Oct. 14-16 1996.

[15] J.Weng, "On comprehensive visual learning," in *Proc. NSF/ARPA Workshop on Performance vs. Methodology in Computer Vision*, pages 152–166, Seattle, WA, June 1994.

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

[17] J.H. Wilkinson, *The Algebraic Eigenvalue Problem*, Oxford University Press, New York, NY, 1965.

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

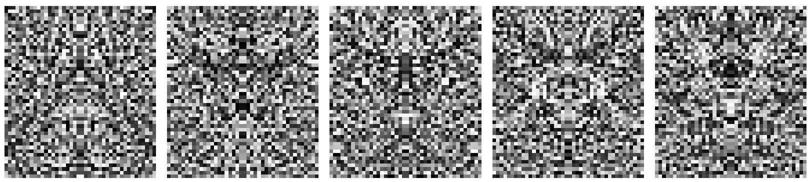


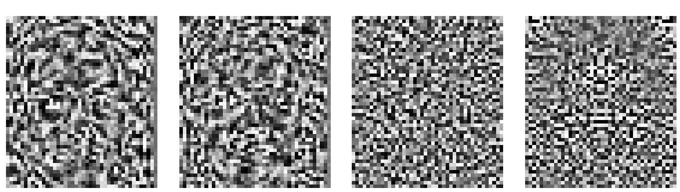
Figure 1: The first five pure LDA bases

Figure 2: The first five PCA + LDA bases

The average face and first four eigenfaces

Eigenfaces 15, 100, 200, 250, 300

Figure 3: Useful eigenfaces



Eigenfaces 400, 450, 1000, 2000

Figure 4: Suspicious eigenfaces: statistically insignificant.

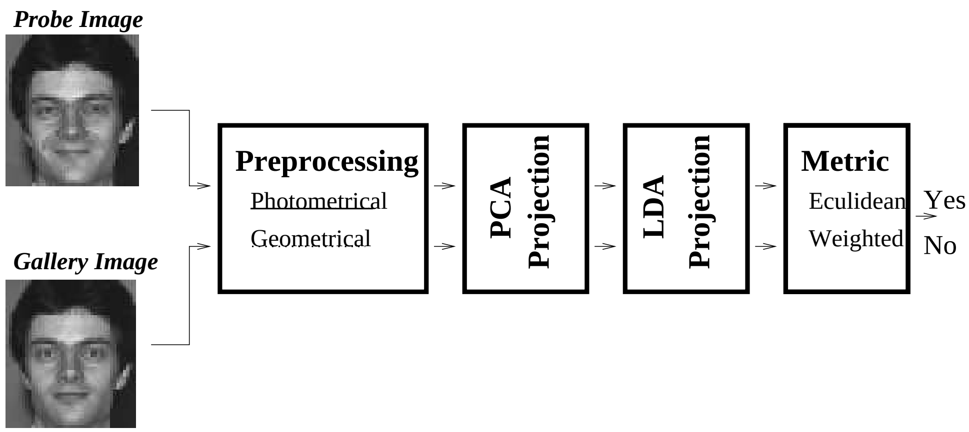
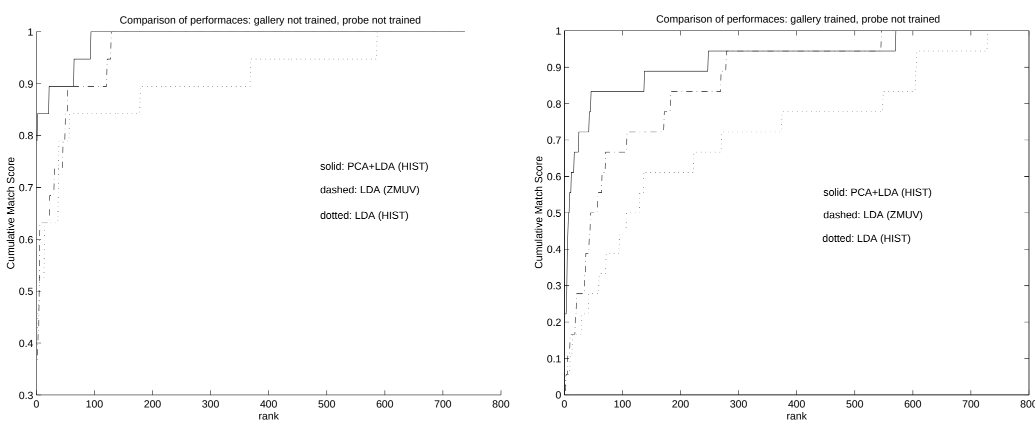


Figure 5: The generalized LDA face recognition system



(a) (b)

Figure 6: (a) Performance comparison on the 19 images from USC dataset, (b) Performance comparison on the 18 images from FERET dataset but not included in the training set. Legends used in figure: HIST is the abbreviation for HISTogram equalization preprocessing, while ZMUV is for Zero Mean Unit Variance preprocessing.

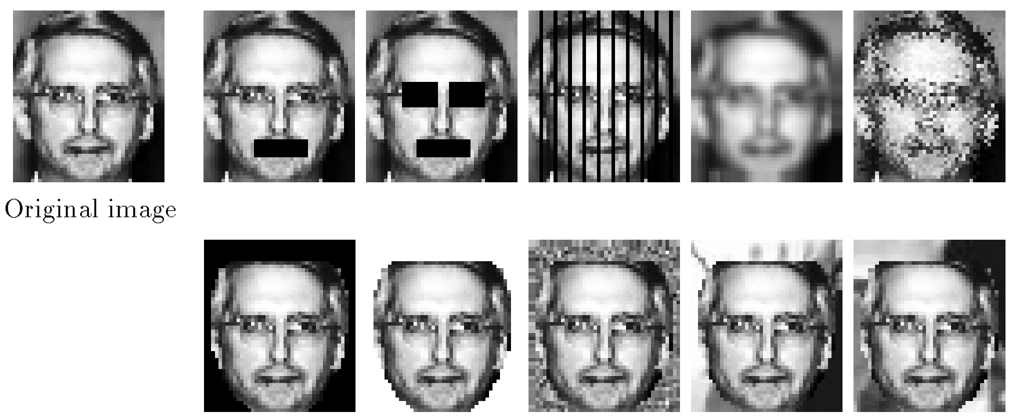
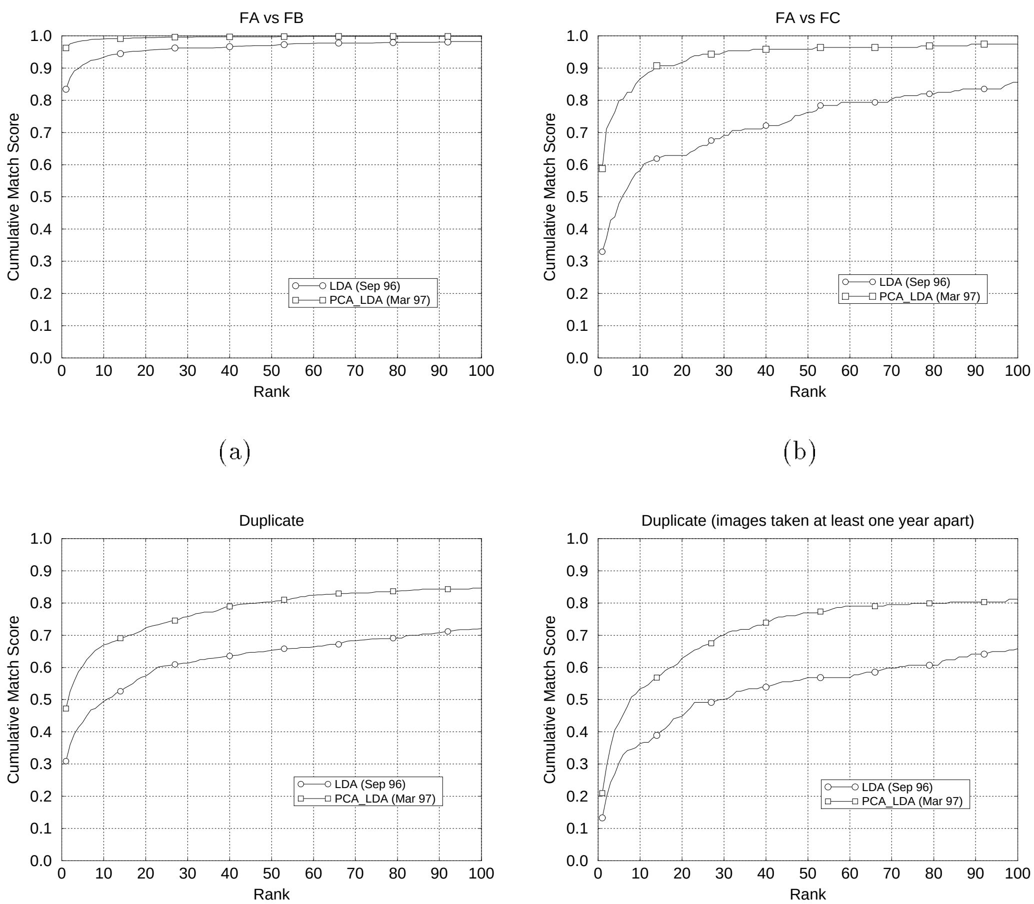


Figure 7: Electronically-modified images which have been correctly identified.



(a) (b) (c) (d)

Figure 8: 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 Lab*)

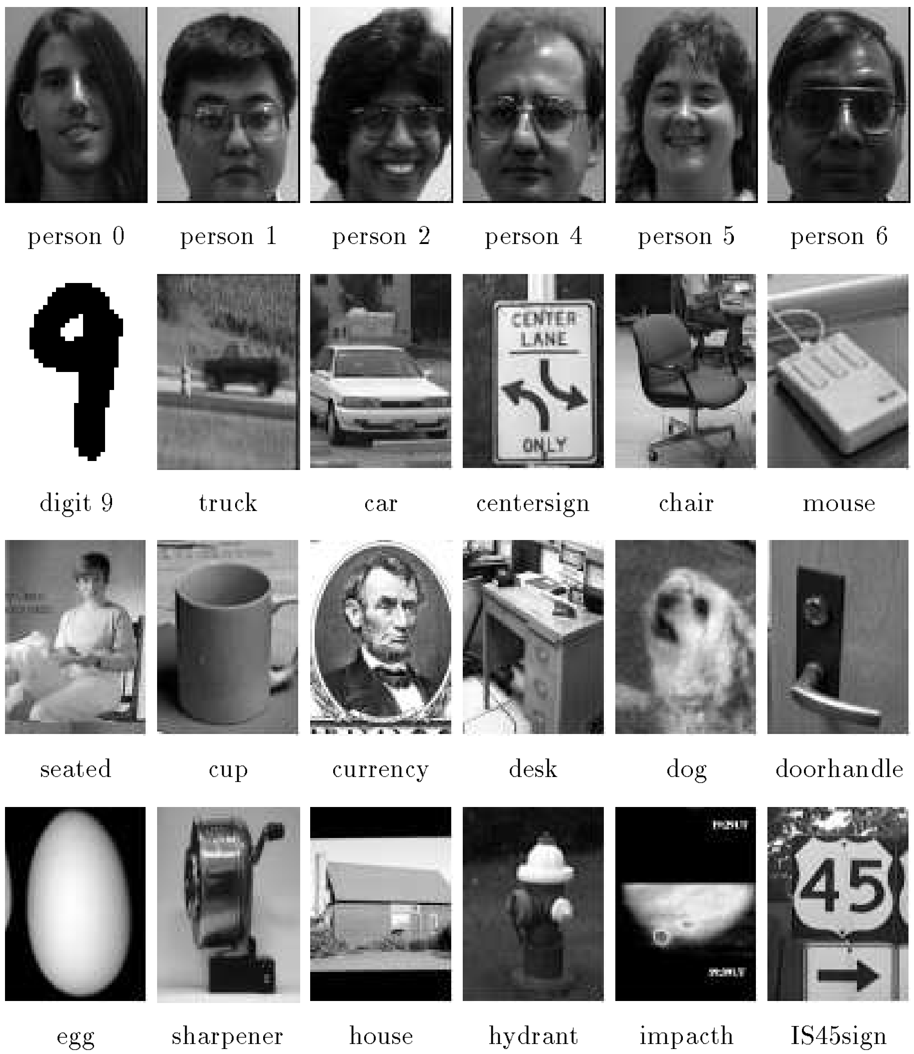


Figure 9: Representative images from the different classes. Images are shown without the pixel weighting, which applying a different weight to each pixel, with decreasing weight from the center to the periphery. This pixel weighting tends to suppress the background around the periphery.