Bilinear Discriminant Analysis for Face Recognition

Muriel Visani[1], Christophe Garcia[1], and Jean-Michel Jolion[2]

[1] France Telecom Research & Development,
4, rue du Clos Courtel, 35512 Cesson-Sevigne, France
{muriel.visani, christophe.garcia}@rd.francetelecom.com
[2] Laboratoire LIRIS, INSA Lyon, 20, Avenue Albert Einstein,
Villeurbanne, 69621 cedex, France
Jean-Michel.Jolion@insa-lyon.fr

Abstract. In this paper, we present a new statistical projection-based face recognition method, called Bilinear Discriminant Analysis (BDA). The proposed technique effectively combines two complementary versions of Two-Dimensional-Oriented Linear Discriminant Analysis (2DoLDA), namely Column-Oriented Linear Discriminant Analysis (CoLDA) and Row-Oriented Linear Discriminant Analysis (RoLDA). BDA relies on the maximization of a generalized bilinear projection-based Fisher criterion. A series of experiments was performed on various international face image databases in order to evaluate and compare the effectiveness of BDA to RoLDA and CoLDA. The experimental results indicate that BDA outperforms RoLDA, CoLDA and 2DPCA for face recognition, while leading to a significant dimensionality reduction.

1 Introduction

In the eigenfaces [1] (resp. fisherfaces [2]) method, the 2D face images of size $h \times w$ are first transformed into 1D image vectors of size $h \cdot w$, and then a Principal Component Analysis (PCA) (resp. Linear Discriminant Analysis (LDA)) is applied to this high-dimensional vector space, where statistical analysis is costly and may be unstable. To overcome these drawbacks, Yang *et al.* [3] proposed the Two Dimensional PCA (2DPCA) method, that aims at performing PCA directly using the face image matrices. It has been shown that 2D PCA is more effective [3] and robust [4] than the eigenfaces method when dealing with face segmentation inaccuracies, low-quality images and partial occlusions.

In [5], we proposed the Two-Dimensional-Oriented Linear Discriminant Analysis (2DoLDA) approach, that consists in applying LDA to image matrices. We have shown on various face databases that 2DoLDA provides better face recognition results than both 2DPCA and the Fisherfaces method, and that it is more robust to variations in lighting conditions, facial expressions and head pose.

In this paper, we propose a novel supervised projection method called Bilinear Discriminant Analysis (BDA) that outperforms 2DoLDA while substantially reducing the computational cost of the recognition step.

S. Singh et al. (Eds.): ICAPR 2005, LNCS 3687, pp. 247–256, 2005.

The remainder of the paper is organized as follows. In section 2, we remind the theory and algorithm of 2DoLDA. In section 3, we describe the principle and algorithm of the proposed BDA method, pointing out its advantages over previous methods. In section 4, a series of three experiments, on different international data sets, is presented to demonstrate the effectiveness and robustness of BDA and compare its performances with respect to RoLDA, CoLDA and 2DPCA. Finally, conclusions are drawn in section 5.

2 Two-Dimensional Oriented Linear Discriminant Analysis (2Do LDA)

In [5], we introduced a version of 2DoLDA that will further be called Row-Oriented Linear Discriminant Analysis (RoLDA). However, 2DoLDA may be implemented in two different ways: RoLDA and Column-oriented LDA (CoLDA). Let us first present RoLDA.

The model is constructed from a training set Ω containing n face images of C people, with multiple views per person. The set of images corresponding to one person is called a *class*. Let us denote Ω_c the set of n_c images belonging to class c. Each face image is stored as a $h \times w$ matrix X_i, labelled by its belonging class. Let us consider a $w \times k$ projection matrix P, and the following projection:

$$X_i^P = X_i \cdot P \tag{1}$$

The matrix X_i^P, of size $h \times k$, is the signature of X_i using RoLDA. Our aim is to determine, for a fixed size $h \times k$, the optimal matrix P^* jointly maximizing separation between different classes and minimizing separation between signatures from the same class. Under the assumptions of multinormality and homoscedasticity of the image matrices rows, P^* maximizes the following *generalized Fisher criterion* [5]:

$$J(P) = \frac{|P^T S_b P|}{|P^T S_w P|} \tag{2}$$

S_w and S_b being respectively the *generalized within-class* and *between-class covariance matrices* of the training set:

$$S_w = \sum_{c=1}^{C} \sum_{X_i \in \Omega_c} (X_i - \bar{X}_c)^T (X_i - \bar{X}_c) \text{ and } S_b = \sum_{c=1}^{C} n_c (\bar{X}_c - \bar{X})^T (\bar{X}_c - \bar{X}) \tag{3}$$

with $\bar{X}_c$ and $\bar{X}$ being mean images, computed respectively from Ω_c and Ω. If S_w is non-singular (which is generally verified as $w << n$), the k columns of P^* are the eigenvectors of $S_w^{-1} S_b$ with largest eigenvalues. A numerically stable way to compute them is given in [6].

Analogeously, CoLDA relies on the following projection: $X_i^Q = Q^T \cdot X_i$ (4) where Q is a $h \times k$ projection matrix, and the $k \times w$ matrix X_i^Q is the signature of X_i using CoLDA. Under the assumptions of multinormality and homoscedasticity of the image matrices columns, we can consider the following generalized Fisher criterion:

$$J(Q) = \frac{|Q^T \Sigma_b Q|}{|Q^T \Sigma_w Q|} \tag{5}$$

where Σ_w and Σ_b are respectively the within-class and between-class covariance matrices of the set $(X_i^T)_{i \in \{1 \dots n\}}$:

$$\Sigma_w = \sum_{c=1}^{C} \sum_{X_i \in \Omega_c} (X_i - \bar{X}_c)(X_i - \bar{X}_c)^T \text{ and } \Sigma_b = \sum_{c=1}^{C} n_c (\bar{X}_c - \bar{X})(\bar{X}_c - \bar{X})^T \tag{6}$$

Let us denote Q^* the optimal projection matrix of size $h \times k$, maximizing criterion (5). If Σ_w is non-singular, the columns of Q^* are the k eigenvectors of $\Sigma_w^{-1} \Sigma_b$ with largest eigenvalues.

For RoLDA and CoLDA, there are at most $C - 1$ eigenvectors corresponding to non-zero eigenvalues; their number k can be selected using the Wilks Lambda criteria, which is also known as *stepwise discriminant analysis* [7]. This analysis shows that the number k of eigenvectors required by both methods is comparable and generally inferior to 15, even if the number of classes is large, as shown in Fig. 2.(a), reporting on an experiment performed on 107 classes.

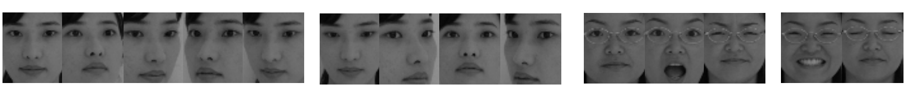
Recognition is performed by using the Euclidean distance between the signatures of the face images, and the nearest neighbour rule.

3 Bilinear Discriminant Analysis (BDA)

3.1 Why Combine CoLDA and RoLDA?

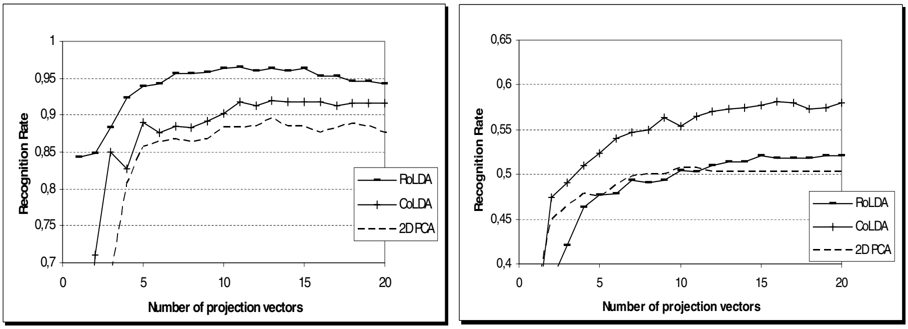
We conducted four experiments highlighting the complementarity of RoLDA and CoLDA. In the following, all the face images are centered and cropped to a size of $h \times w = 75 \times 65$ pixels.

The first two experiments are performed on subsets of the Asian Face Database PF01 [8] containing 107 people. They illustrate the fact that, depending on the training and test data, RoLDA and CoLDA outperform each other. In the first experiment, the training and test sets, illustrated in Fig. 1.(a-b), contain respectively 5 near-frontal views per person (535 images) and 4 non-frontal views per person (428 images). These two sets differ in the head pose. Fig. 2. (a) shows



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

Fig. 1. Extracts of (a) the training set and (b) the test set used for the first experiment; Extracts of (c) the training set and (d) the test set used for the second experiment.



(a) (b)

Fig. 2. Compared recognition rates of RoLDA, CoLDA and 2DPCA on a subset of the PF01 database showing (a) head pose changes and (b) facial expression changes, when varying the number k of projection vectors.

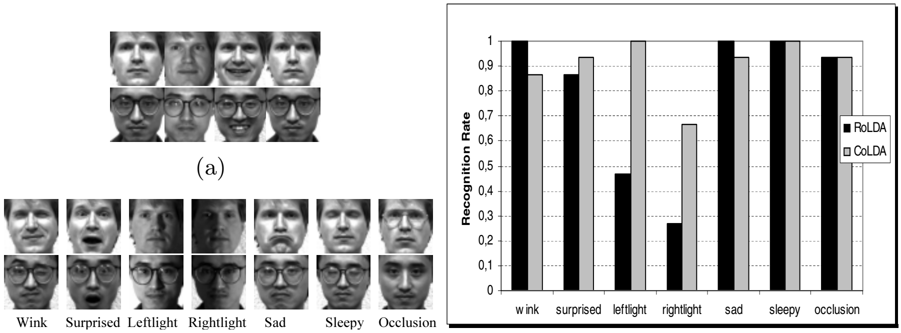
that both CoLDA and RoLDA are highly effective (recognition rates superior to 92 %), and outperform 2DPCA. However, RoLDA outperforms CoLDA, with a 4,5% improvement of the recognition rate between their respective maxima.

The second experiment is performed on a subset of the PF01 database containing 107 people, with five different facial expressions. This subset is randomly partitioned into a training set and a test set, illustrated in Fig. 1.(c-d). From Fig. 2.(b) we can see that, even if RoLDA and CoLDA are not highly performing (the recognition rates are inferior to 60%), both of them outperform 2DPCA. However, CoLDA is more effective than RoLDA, with a 5,6% improvement of the recognition rate between their respective maxima.

The third and fourth experiments provide further comparison of the performances of CoLDA and RoLDA. They are performed on the Yale Face Database [2], that contains 15 people and 11 views per person, with occlusions and variations in lighting conditions and facial expressions. In the third experiment, the Yale database is randomly partitioned into a training set containing four views per person, and a test set containing six views per person. To ensure homoscedasticity, the views of each set are consistent among the classes, e.g. all the "wink" views are included in the test set, and all the "neutral" in the training set. This operation is repeated five times. From each partition, we compute a confusion matrix with $k = C-1 = 14$ (see Table 1.) In each confusion matrix, the top left cell contains the number of faces correctly classified by both RoLDA and CoLDA. The top right entry is the number of faces correctly classified by

Table 1. Confusion matrices of RoLDA and CoLDA, computed from five random partitions of the Yale Face Database

53	10	71	11	72	8	55	5	63	2
11	16	5	3	4	6	14	16	7	18
(a)		(b)		(c)		(d)		(e)	



(a) (b) (c)

Fig. 3. Extracts (a) of the training set and (b) of the seven test sets, taken from Yale and used for the fourth experiment. Any subject not wearing eyeglasses in the training set wears eyeglasses in the "occlusion" set, and *vice-versa.* (c) Compared recognition rates of CoLDA and RoLDA, computed from the seven test sets illustrated in (b).

RoLDA, but misclassified by CoLDA. The bottom left cell contains the number of faces correctly classified by CoLDA, but misclassified by RoLDA. The bottom right entry is the number of faces misclassified by both methods. Table 1.(a) shows that, on the first random partition of the Yale database, the performances of RoLDA and CoLDA are comparable (the recognition rates are respectively $\frac{53+10}{53+10+11+16} = 70\%$ and 71,1%). However, classification results are very different: 21 samples (23,3% of the test set) are correctly classified by only one method. Moreover, $82,2\% \gg \mathtt{max}(70\%,\, 71,1\%)$ of the query faces are recognized by at least one of the two methods. Table 1.(b-c) illustrate the fact that RoLDA generally outperforms CoLDA. Table 1.(d-e) show that, in some configurations where the rate of misclassification by both methods is high -respectively $\frac{16}{90} = 17,8\%$ and 20% for partitions (d) and (e)-, CoLDA outperforms RoLDA.

The fourth experiment provides further qualitative analysis. The training set, illustrated in Fig. 3.(a), contains four views for each of the 15 subjects, with variations in lighting conditions and facial expressions. Then, seven test sets, illustrated in Fig. 3.(b) and corresponding to the remaining views, are built. Fig. 3.(c) illustrates the fact that, even if RoLDA is generally more effective than CoLDA, in some cases CoLDA drastically outperforms RoLDA, especially when the test set contains dissimetries of the image following the vertical axis ("leftlight" and "rightlight"). CoLDA can also slightly outperform RoLDA when the test set shows strong facial expression changes, e.g. "surprised". Choosing between CoLDA and RoLDA therefore requires a preliminary qualitative analysis of the training and test sets, which is a difficult task. As both RoLDA and CoLDA have high performances but give different recognition results, appropriately combining them can lead to a highly effective method.

In 2DoLDA, considering image matrices instead of vectors (as in the Fisherfaces method) when performing LDA leads to a reduced computational cost when building the model, and to a reduced storage cost [5]. But the size of the signatures is $h \times k$ for RoLDA and $k \times w$ for CoLDA, and may be large. As exposed in the following section, using BDA leads to a drastic reduction in the signatures size, and therefore reduces the computational cost during the recognition step, which is often online.

3.2 Description of Bilinear Discriminant Analysis

Let us consider two projection matrices $Q \in \mathbb{R}^{h\times k}$ and $P \in \mathbb{R}^{w\times k}$, and the following bilinear projection:

$$X_i^{Q,P} = Q^T X_i P \tag{7}$$

where the $k \times k$ matrix $X_i^{Q,P}$ is the signature of X_i using BDA. For any fixed k, let us search for the optimal pair of matrices (Q^*, P^*), maximizing the following generalized Fisher criterion:

$$(Q^*, P^*) = \underset{(Q,P)\in\mathbb{R}^{h\times k}\times\mathbb{R}^{w\times k}}{\texttt{Argmax}} \frac{|S_b^{Q,P}|}{|S_w^{Q,P}|} \tag{8}$$

$$= \underset{(Q,P)\in\mathbb{R}^{h\times k}\times\mathbb{R}^{w\times k}}{\texttt{Argmax}} \frac{|\sum_{c=1}^{C} n_c(\overline{X_c^{Q,P}} - \overline{X^{Q,P}})^T(\overline{X_c^{Q,P}} - \overline{X^{Q,P}})|}{|\sum_{c=1}^{C}\sum_{i\in\Omega_c}(X_i^{Q,P} - \overline{X_c^{Q,P}})^T(X_i^{Q,P} - \overline{X_c^{Q,P}})|} \tag{9}$$

$S_w^{Q,P}$ and $S_b^{Q,P}$ being the within-class and between-class covariance matrices of the signatures set $(X_i^{Q,P})_{i\in\{1,\ldots,n\}}$.

This objective function is biquadratic and has no analytical solution. We therefore propose an iterative procedure that we call *Bilinear Discriminant Analysis*. Let us expand the expression (9):

$$(Q^*, P^*) = \underset{(Q,P)\in\mathbb{R}^{h\times k}\times\mathbb{R}^{w\times k}}{\texttt{Argmax}} \left[\frac{|\Sigma_{c=1}^{C} n_c(P^T(\overline{X_c}-\overline{X})^T QQ^T(\overline{X_c}-\overline{X})P)|}{|\Sigma_{c=1}^{C}\Sigma_{i\in\Omega_c}(P^T(X_i-\overline{X_c})^T QQ^T(X_i-\overline{X_c})P)|}\right] \tag{10}$$

For any fixed $Q \in \mathbb{R}^{h\times k}$, using equation (10), the objective function (9) can be rewritten:

$$P^* = \underset{P\in\mathbb{R}^{w\times k}}{\texttt{Argmax}} \left[\frac{|P^T\left[\Sigma_{c=1}^{C} n_c(\overline{X_c^Q}-\overline{X^Q})^T(\overline{X_c^Q}-\overline{X^Q})\right]P|}{|P^T\left[\Sigma_{c=1}^{C}\Sigma_{i\in\Omega_c}(X_i^Q-\overline{X_c^Q})^T(X_i^Q-\overline{X_c^Q})\right]P|}\right] = \underset{P\in\mathbb{R}^{w\times k}}{\texttt{Argmax}} \frac{|P^T S_b^Q P|}{|P^T S_w^Q P|} \tag{11}$$

S_w^Q and S_b^Q being respectively the *generalized within-class covariance matrix* and the *generalized between-class covariance matrix* of the set $(X_i^Q)_{i\in\{1\ldots n\}}$, each X_i^Q being computed using (4). Therefore the columns of the matrix P^* are the k eigenvectors of ${S_w^Q}^{-1} S_b^Q$ with largest eigenvalues, obtained by applying RoLDA on the set of the projected samples X_i^Q. Let us denote $A = P^T(\overline{X_c} - \overline{X})^T Q$, matrix of size $k \times k$. Given that, for every square matrix A, $|A^T A| = |AA^T|$, the objective function (9) can be rewritten:

$$(Q^*, P^*) = \underset{(Q,P)\in\mathbb{R}^{h\times k}\times\mathbb{R}^{w\times k}}{\texttt{Argmax}} \left[\frac{|\Sigma_{c=1}^{C} n_c(Q^T(\overline{X_c}-\overline{X})PP^T(\overline{X_c}-\overline{X})^T Q)|}{[\Sigma_{c=1}^{C}\Sigma_{i\in\Omega_c}(Q^T(X_i-\overline{X_c})PP^T(X_i-\overline{X_c})^T Q)|}\right] \tag{12}$$

For any fixed $P \in \mathbb{R}^{w \times k}$, using equation (12) the objective function (9) can be rewritten $Q^* = \underset{Q \in \mathbb{R}^{h \times k}}{\mathtt{Argmax}} \frac{|Q^T \Sigma_b^P Q|}{|Q^T \Sigma_w^P Q|}$, where Σ_w^P and Σ_b^P are respectively the generalized within-class and between-class covariance matrices of the set $((X_i^P)^T)_{i \in \{1 \ldots n\}}$, each X_i^P being computed using (1). Therefore, the columns of Q^* are the k eigenvectors of $(\Sigma_w^P)^{-1} \Sigma_b^P$ with largest eigenvalues, obtained by applying CoLDA on the set of the projected samples X_i^P.

3.3 Algorithm of the BDA Approach

Let us initialize $P_0 = I_w$, the identity matrix of $\mathbb{R}^{w \times w}$, and $k_0 = C-1$. The proposed algorithm for BDA is:

1. For $i \in \{1, \ldots, n\}$, compute $X_i^{P_t} = X_i P_t$;
2. Apply CoLDA to $(X_i^{P_t})_{i \in \{1,\ldots,n\}}$: compute $\Sigma_w^{P_t}$, $\Sigma_b^{P_t}$ and, from $(\Sigma_w^{P_t})^{-1} \cdot \Sigma_b^{P_t}$, compute Q_t, of size $h \times k_t$;
3. For $i \in \{1, \ldots, n\}$, compute $X_i^{Q_t} = (Q_t)^T X_i$;
4. Apply RoLDA to $(X_i^{Q_t})_{i \in \{1,\ldots,n\}}$: compute $S_w^{Q_t}$, $S_b^{Q_t}$ and, from $(S_w^{Q_t})^{-1} \cdot S_b^{Q_t}$, compute P_t, of size $w \times k_t$;
5. Compute $\alpha = -(n - \frac{w+C}{2} - 1) \ln \left[\prod_{j=k_t+1}^{C-1} \frac{1}{1+\lambda_j} \right]$;
6. if $\alpha < p\text{-}value \left[\chi^2 \left((w-k_t)(C-k_t-1) \right) \right]$, then $t \leftarrow t+1$, $k_t \leftarrow k_{t-1}$-1, and return to step 1;
7. else $k_t \leftarrow k_{t-1}$, $Q \leftarrow Q_{t-1}$ and $P \leftarrow P_{t-1}$.

The stopping criterion (steps 5.-7.) derives from the Wilks Lambda criterion, testing the discriminatory power of the C-k_t-1 eigenvectors of $(S_w^{Q_t})^{-1} \cdot S_b^{Q_t}$ removed at step 4., by keeping in P_t only the k_t eigenvectors with highest eigenvalues $(\lambda_j)_{j \in \{1 \ldots k_t\}}$. We consider the following test: H_0: at least one of the eigenvectors k_t+1,...,C-1 is discriminative, and H_1: non H_0. Under H_0, it can be easily shown that $-(n - \frac{w+C}{2} - 1) \ln(\prod_{j=k_t+1}^{C-1} \frac{1}{1+\lambda_j})$ corresponds to a χ^2 distribution, with $(w$-$k_t)(C$-k_t-1) degrees of freedom. The p-value can be chosen at a confidence level of 5%. If $\alpha < p\text{-}value$, the C-k_t-1 last eigenvectors can be removed and the stepwise analysis goes on. If $\alpha > p\text{-}value$, the eigenvector k_t+1=k_{t-1} is discriminative and should be kept.

Recognition is performed in the BDA projection space, by using the Euclidean distance between face image signatures, and the nearest neighbour rule.

We can note that the computational cost of one comparison is $o(k^2)$ for BDA, versus $o(h \cdot k)$ for RoLDA and 2DPCA, and $o(w \cdot k)$ for CoLDA; therefore BDA drastically reduces the computational cost of the recognition step.

4 Experimental Results

Three experiments are performed on the Asian Face Database PF01 [8], the FERET [9] [1] face database, and the ORL Database [10], to assess the effectiveness of BDA and compare it with RoLDA, CoLDA and 2D-PCA.

[1] Portions of the research in this paper use the FERET database of facial images collected under the FERET program.

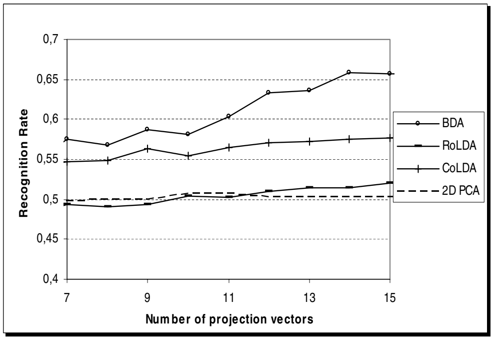
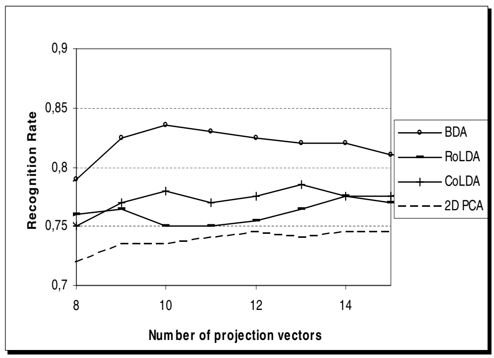


Fig. 4. Compared recognition rates of BDA, RoLDA, CoLDA and 2DPCA, on the subset of PF01 with expression variations, when varying the number k of eigenvectors.

(a) (b) (c)



(d)

Fig. 5. Extracts (a) of the training set and (b-c) of the two test sets to be matched. (d) Compared recognition rates of BDA, RoLDA, CoLDA and 2DPCA, when matching the test sets (b) and (c).

The training and test sets used for the first experiment, differing in the facial expressions, were used for the second experiment reported in section 3.1 and are illustrated in Fig. 1.(c-d). From Fig. 4., we can see that BDA strongly outperforms RoLDA, CoLDA and 2DPCA.

The second experiment, performed on FERET, aims at evaluating the generalization power of BDA. Indeed, LDA-based methods are known to be more effective when comparing faces of known people, but provide worse generalization results than unsupervised methods. The training set, illustrated in Fig. 5.(a), contains 818 images of 152 people with at least four views per person, taken on different days and under different lighting conditions. Two test sets, each one containing 200 people with one view per person and illustrated in Fig. 5.(b-c), are compared. The test sets are taken from FERET, but none of the 200 people is registered in the training set. From one test set to the other, the facial expres-

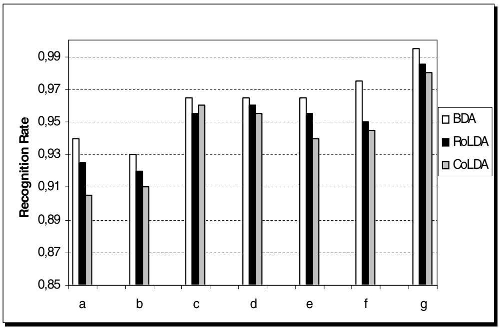


Fig. 6. Compared recognition rates of BDA, RoLDA and CoLDA on 7 random partitions of the ORL database

		$\cap$ BDA
RoLDA $\cap$ CoLDA	1297	1292
RoLDA $\cap$ $\rceil$CoLDA	33	24
$\rceil$RoLDA $\cap$ CoLDA	22	14
$\rceil$RoLDA $\cap$ $\rceil$CoLDA	48	17
Total	1400	1347

Table 2. Contingency table summed up over 7 random partitions of ORL

sions vary. From Fig. 5.(d) we can conclude that, when the training set contains many classes with important variations inside the classes, BDA provides better generalization than the other methods.

For the third experiment, the ORL database is randomly partitioned into a training set containing five views, and a test set containing the five remaining views, for each of the 40 persons. This operation is repeated seven times and BDA, RoLDA and CoLDA are applied. Fig. 6. shows that BDA provides better recognition rates than RoLDA and CoLDA on all the random partitions, whenever RoLDA outperforms CoLDA (partitions (a-b) and (d-g)) or CoLDA outperforms RoLDA (partition (c)). The results are computed from the optimal number of projection vectors, which is $k = 14$ for the three methods. For further analysis, the contingency table summed up over partitions (a-g) is given in Table 2. The total number of query faces is $7 \cdot 200 = 1400$. The logical symbol "$\rceil$" stands for "not", i.e. the entry in the second row and first column of the table is the number of faces recognized by RoLDA, but misclassified by CoLDA. The logical symbol "$\cap$" stands for "and": the entry in the second row and second column is the number of samples correctly classified by RoLDA and BDA, but misclassified by CoLDA. From Table 2. we can see that BDA correctly classifies $\frac{1292}{1297} = 99,6\%$ of the samples that were recognized by both RoLDA and CoLDA. Moreover, it recognizes the major part of the samples that were recognized by only one of the two methods (72,7% for RoLDA and 63,6% for CoLDA). It also correctly classifies 35,4% of the samples that were misclassified by both methods, which shows the effectiveness of the BDA iterative algorithm. It should be noted that, as the face images have been cropped to a size of 75×65 pixels, the size of one sample signature is $75 \cdot 14 = 1050$ for RoLDA, $65 \cdot 14 = 910$ for CoLDA, and only $14^2 = 196$ for BDA.

5 Conclusion

In this paper, we have proposed a new supervised statistical projection based technique, named Bilinear Discriminant Analysis, that can be successfully applied to face recognition. This method effectively combines two complementary versions of 2DoLDA, through an iterative algorithm maximizing a generalized Fisher criterion relying on bilinear projections.

A series of experiments, performed on various international databases, have shown the complementarity of the two versions of 2DoLDA and highlighted that the proposed iterative algorithm outperforms 2DoLDA and 2DPCA; as a consequence it also outperforms the fisherfaces and eigenfaces methods. Moreover, BDA provides image signatures of reduced size compared to 2DoLDA and 2DPCA, which results in an important computational gain during the recognition step.

Acknowledgement. This research was supported by the European Commission under contract FP6-001765 aceMedia.

References

1. Turk, M.A., Pentland, A.D.: Eigenfaces for Recognition. Journal of Cognitive Neuroscience. **3**(1) (1991) 71–86.
2. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs Fisherfaces : Recognition Using Class Specific Linear Projection. IEEE Trans. on Pattern Analysis and Machine Intelligence, Special Issue on Face Recognition. **19**(7) (1997) 711–720.
3. Yang, J., Zhang, D., Frangi, A.F., Yang, J.Y.: Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition. IEEE Trans. on Pattern Analysis and Machine Intelligence. **26**(1) (2004) 131–137.
4. Visani, M., Garcia, C., Laurent, C.: Comparing Robustness of Two-Dimensional PCA and Eigenfaces for Face Recognition. In Proc. of the International Conference on Image Analysis and Recognition (ICIAR 04). Springer Lecture Notes in Computer Science (LNCS 3211), A. Campilho, M. Kamel (eds) (Sept. 2004) 717–724.
5. Visani, M., Garcia, C., Jolion, J.M.: Two-Dimensional-Oriented Linear Discriminant Analysis for Face Recognition. In Proc. of the International Conference on Computer Vision and Graphics (ICCVG 2004). To appear in Computational Imaging and Vision series (Sept. 2004).
6. Swets, D.L., Weng, J.: Using Discriminant Eigenfeatures for Image Retrieval. IEEE Trans. on Pattern Analysis and Machine Intelligence. **18**(8) (1996) 831–836.
7. Jenrich, R.I.: Stepwise Discriminant Analysis. In A. Enslein, A. Ralston, and H. S. Wilf (eds.), Statistical Methods for Digital Computers. Wiley Interscience, New York, N.Y, (1977) 76–95.
8. Hwang, B.W., Roh, M.C., Lee, S.W.: Performance Evaluation of Face Recognition Algorithms on Asian Face Database. IEEE International Conference on Automatic Face and Gesture Recognition. Seoul, Korea, (May 2004) 278–283.
9. Phillips, P.J., Wechsler, H., Huang, J., Rauss, P.: The FERET Database and Evaluation Procedure for Face Recognition Algorithms. Image and Vision Computing. **16**(5) (1998) 295–306.
10. Samaria, F., Harter, A.: Parameterisation of a Stochastic Model for Human Face Identification. Proceedings of the IEEE Workshop on Applications of Computer Vision. Sarasota, Florida, (Dec. 1994).