Face Recognition Using Modular Bilinear Discriminant Analysis

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

[1] France Telecom R&D TECH/IRIS,
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 approach for face recognition, named Modular Bilinear Discriminant Analysis (MBDA). In a first step, a set of experts is created, each one being trained independently on specific face regions using a new supervised technique named Bilinear Discriminant Analysis (BDA). BDA relies on the maximization of a generalized Fisher criterion based on bilinear projections of face image matrices. In a second step, the experts are combined to assign an identity with a confidence measure to each of the query faces. A series of experiments is performed in order to evaluate and compare the effectiveness of MBDA with respect to BDA and to the Modular Eigenspaces method. The experimental results indicate that MBDA is more effective than both BDA and the Modular Eigenspaces approach for face recognition.

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 image vector space, where statistical analysis is costly and may be unstable. To overcome these drawbacks, Yang *et al.* [3] proposed the Two Dimensional PCA (2D PCA) method that aims at performing PCA using directly the face image matrices. It has been shown that 2D PCA is more effective [3] and robust [4] than the eigenfaces 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 on image matrices. We have shown on various face image databases that 2DoLDA is more effective than both 2D PCA and the Fisherfaces method for face recognition, and that it is more robust to variations in lighting conditions, facial expressions and head poses.

S. Bres and R. Laurini (Eds.): VISUAL 2005, LNCS 3736, pp. 24–34, 2005.

The first contribution of this paper is a new supervised feature extraction method generalizing and outperforming 2DoLDA, namely Bilinear Discriminant Analysis (BDA). This method is based upon the optimization of a generalized Fisher criterion [6, 2] computed from image matrices directly, and we call it BDA because this criterion uses bilinear projections. The second contribution of this paper is a modular classification scheme combining BDA experts trained on different regions of the face, chosen as in [7, 8] and designed to be more robust to facial expression changes.

The remainder of this paper is organized as follows. In section 2, we describe in details the principle and algorithm of the proposed BDA technique, pointing out its advantages over previous methods. In section 3, we present our multiple expert scheme named MBDA. Then, we provide in section 4 a series of two experiments performed on an international data set, demonstrating the effectiveness and robustness of MBDA and comparing its performances with respect to 2DoLDA and the Modular Eigenspaces method [7]. Finally, conclusions and closing remarks are drawn in section 5.

2 Bilinear Discriminant Analysis (BDA)

In this section, we describe the proposed BDA feature extraction technique. The model is constructed from a training set Ω containing n face images, with more than one view per each of the C registered persons. The set of images corresponding to one person is called a *class* ; class c is denoted by Ω_c. Each face image is stored as a $h \times w$ matrix X_i labelled by its class.

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:

$$X_i^{Q,P} = Q^T X_i P \tag{1}$$

where the matrix $X_i^{Q,P}$, of size $k \times k$, is considered as the *signature* of the face X_i. We are searching for the optimal pair of matrices (Q^*, P^*) maximizing the separation between signatures from different classes while minimizing the separation between signatures from the same class. As a consequence, we can consider the following generalized Fisher criterion:

$$(Q^*, P^*) = \underset{(Q,P) \in \mathbb{R}^{h \times k} \times \mathbb{R}^{w \times k}}{Argmax} \frac{|S_b^{Q,P}|}{|S_w^{Q,P}|} \tag{2}$$

$$= \underset{(Q,P) \in \mathbb{R}^{h \times k} \times \mathbb{R}^{w \times k}}{Argmax} \frac{|\sum_{c=1}^{C} n_c (\overline{X_c^{Q,P}} - \overline{X^{Q,P}})^T (\overline{X_c^{Q,P}} - \overline{X^{Q,P}})|}{|\sum_{c=1}^{C} \sum_{i \in \Omega_c} (X_i^{Q,P} - \overline{X_c^{Q,P}})^T (X_i^{Q,P} - \overline{X_c^{Q,P}})|} \tag{3}$$

where $S_w^{Q,P}$ and $S_b^{Q,P}$ are respectively the within-class and between-class covariance matrices of the set $(X_i^{Q,P})_{i \in \{1,\dots,n\}}$ of the projected samples from Ω, $\overline{X^{Q,P}}$ and $\overline{X_c^{Q,P}}$ are the mean face matrices calculated respectively over Ω and Ω_c.

The objective function given in equation (3) is biquadratic and has no analytical solution. We therefore propose an iterative procedure that we call *Bilinear*

Discriminant Analysis. Let us expand the expression (3):

$$(Q^*, P^*) = \underset{(Q,P)\in\mathbb{R}^{h\times k}\times\mathbb{R}^{w\times k}}{Argmax} \left[\frac{|\sum_{c=1}^{C} n_c (P^T(\overline{X_c}-\overline{X})^T QQ^T(\overline{X_c}-\overline{X})P)|}{|\sum_{c=1}^{C}\sum_{i\in\Omega_c}(P^T(X_i-\overline{X_c})^T QQ^T(X_i-\overline{X_c})P)|} \right] \tag{4}$$

For any fixed $Q \in \mathbb{R}^{h\times k}$ and using equation (4), the objective function (3) can be rewritten:

$$P^* = \underset{P\in\mathbb{R}^{w\times k}}{Argmax} \left[\frac{|P^T\left[\sum_{c=1}^{C} n_c(\overline{X_c^Q}-\overline{X^Q})^T(\overline{X_c^Q}-\overline{X^Q})\right]P|}{|P^T\left[\sum_{c=1}^{C}\sum_{i\in\Omega_c}(X_i^Q-\overline{X_c^Q})^T(X_i^Q-\overline{X_c^Q})\right]P|} \right] \tag{5}$$

$$= \underset{P\in\mathbb{R}^{w\times k}}{Argmax} \frac{|P^T S_b^Q P|}{|P^T S_w^Q P|} \tag{6}$$

with 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 $(X_i^Q)_{i\in\{1,\ldots,n\}}$, where

$$\forall i \in \{1,\ldots,n\},\ X_i^Q = Q^T \cdot X_i \tag{7}$$

Therefore, the columns of matrix P^* are the k eigenvectors of ${S_w^Q}^{-1} S_b^Q$ with largest eigenvalues. A stable way to compute the eigen-decomposition, by applying Singular Value Decomposition (SVD) on the covariance matrices, is given in [6]. If $Q = I_h$, the identity matrix of size $h \times h$, P^* is the projection matrix of 2DoLDA [5].

Given that, for every square matrix A, $|A^T A| = |AA^T|$ and considering the matrix $P^T(\overline{X_c}-\overline{X})^T Q$ of size $k \times k$, the objective function (3) can be rewritten:

$$(Q^*, P^*) = \underset{(Q,P)\in\mathbb{R}^{h\times k}\times\mathbb{R}^{w\times k}}{Argmax} \left[\frac{|\sum_{c=1}^{C} n_c (Q^T(\overline{X_c}-\overline{X})PP^T(\overline{X_c}-\overline{X})^T Q)|}{[\sum_{c=1}^{C}\sum_{i\in\Omega_c}(Q^T(X_i-\overline{X_c})PP^T(X_i-\overline{X_c})^T Q)|} \right] \tag{8}$$

For any fixed $P \in \mathbb{R}^{w\times k}$, using equation (8), the objective function (3) can be rewritten:

$$Q^* = \underset{Q\in\mathbb{R}^{h\times k}}{Argmax} \frac{|Q^T \Sigma_b^P Q|}{|Q^T \Sigma_w^P Q|} \tag{9}$$

Σ_w^P and Σ_b^P being the generalized within-class and between-class covariance matrices of the set $((X_i^P)^T)_{i\in\{1,\ldots,n\}}$, where

$$\forall i \in \{1,\ldots,n\},\ X_i^P = X_i \cdot P \tag{10}$$

Therefore, the columns of matrix Q^* are the k eigenvectors of $(\Sigma_w^P)^{-1}\Sigma_b^P$ with largest eigenvalues.

We can note that BDA leads to a significant reduction in the dimensionality of the signatures compared to 2DPCA and 2DoLDA: the size of a signature using BDA is k^2, versus $h \cdot k$ for 2DoLDA and 2D PCA.

2.1 Algorithm of the BDA Approach

Let us initialize $P_0 = I_w$ and $\alpha_0 = 0$. The number k of components is fixed. The choice of k will be discussed in section 4. The proposed algorithm for BDA is:
<u>While</u> $\alpha_t < \tau$

1. For $i \in \{1, \dots, n\}$, compute $X_i^{P_t} = X_i \cdot P_t$.
2. Compute $\Sigma_w^{P_t}$, $\Sigma_b^{P_t}$ and $(\Sigma_w^{P_t})^{-1} \cdot \Sigma_b^{P_t}$;
3. Compute Q_t, whose columns are the first k eigenvectors of $(\Sigma_w^{P_t})^{-1} \cdot \Sigma_b^{P_t}$;
4. For $i \in \{1, \dots, n\}$, compute $X_i^{Q_t} = (Q_t)^T \cdot X_i$.
5. Compute $S_w^{Q_t}$, $S_b^{Q_t}$, and $(S_w^{Q_t})^{-1} \cdot S_b^{Q_t}$;
6. Compute P_t, whose columns are the first k eigenvectors of $(S_w^{Q_t})^{-1} \cdot S_b^{Q_t}$;
7. Compute $\alpha_t = \sqrt{(\|P_t - P_{t-1}\|_2^2 + \|Q_t - Q_{t-1}\|_2^2}$.

It should be noted that the roles of P and Q can be switched, by initializing $Q_0 = I_h$, computing $X_i^{Q_t}$ instead of $X_i^{P_t}$ at step 1., and so on. Experimental results show similar performances for the two versions of the algorithm.

The stopping parameter τ can be determined empirically, from experiments. As P_t and Q_t are normal matrices, no drastic variation of τ is observed from one test set to another, and therefore τ can be determined easily. However, experimental results have also shown that after one iteration the recognition results are satisfying. Therefore, in the following, we will use the preceding algorithm with only one iteration, which is less costly, ensures good recognition results, and frees us from determining τ.

3 Modular Bilinear Discriminant Analysis (MBDA)

We can consider that there are basically two expert combination scenarios. In the first scenario, all the experts use the same input pattern but different feature extraction techniques, chosen to be complementary. In the second scenario, each expert uses its own representation of the input signal, but all the experts use the same feature extractor. In this paper, we focus on the second scenario, and propose a face recognition system based on three experts built by using BDA, each one being trained on a specific face template.

3.1 Description of the Face Templates

Fig. 1 shows the different facial regions from which the experts are trained. Expert 1 is trained from a face region of size 75×65 pixels containing all the facial features, centered on the position of the nose. Expert 2 is trained from a template of size 40×65 pixels containing the nose, eyes and eyebrows, while expert 3 uses a template of size 30×65 pixels containing only the eyes and eyebrows. These regions are chosen to guarantee good recognition rates in most configurations, according to the results obtained in [7, 8].

1 2 3

Fig. 1. Templates from which are trained the three BDA experts

3.2 Expert Combination

We investigated two ways of using simultaneously the three experts: multistage expert combining and expert voting.

Multistage Expert Combining. Multistage Expert Combining (see Fig. 2) requires a two-step training stage. In the first step, each expert $j \in \{1, \ldots, 3\}$ is trained separately, to build the corresponding pair of projection matrices (Q_j, P_j). In the second step, a combiner applies PCA on the concatenation of the signatures obtained in the first step. Each training sample $(X_i) \in \Omega$ is projected onto (Q_j, P_j), giving the matrix $X_i^{Q_j,P_j}$, of size $k_j \times k_j$ (see equation (1)). Then, each of the three matrices $X_i^{Q_j,P_j}$ is transformed into a vector. Next, these three vectors are concatenated to obtain a single vector $\hat{X}_i$ of large size $k_1^2 + k_2^2 + k_3^2$. A subspace $\mathcal{F}$ is built by applying PCA to the set of vectors $(\hat{X}_i)_{i \in \{1,\ldots,n\}}$, in order to reduce the dimensionality of the signatures to size $l << k_1^2 + k_2^2 + k_3^2$ while keeping most of the information contained in the set of vectors $(\hat{X}_i)_{i \in \{1,\ldots,n\}}$. Finally, all vectors $\hat{X}_i$ are projected onto $\mathcal{F}$ to provide meta-signatures $\tilde{X}_i$.

When a face image T has to be recognized, its meta signature $\tilde{T}$ is computed as for the images of the training set and compared to those of Ω using the Euclidean distance. The identity of T is determined by a simple Nearest Neighbour rule. A confidence measure can be associated to this identification result, based on the distributions of the sample classes among the K nearest neighbours of $\tilde{T}$.

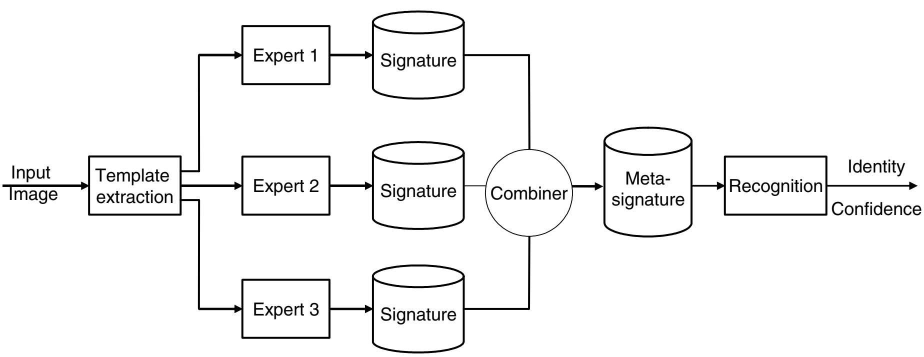

Fig. 2. Multistage Expert Combining. The signatures provided by the three experts are combined using PCA to obtain a meta-signature.

Expert Voting. In the expert voting scheme (see Fig. 3), for any query image T, the signature T^{Q_j,P_j} computed from expert j is compared to the signatures $X_i^{Q_j,P_j}$ of its own local database: for each set (T, j, c) of query image T, expert j and sample class c, the following score is computed [8]:

$$s^j(T,c) = \frac{\underset{X_i \in \Omega_c}{Max} \|T^{Q_j,P_j} - X_i^{Q_j,P_j}\|_2^{-1}}{\sum_{c=1}^{C} \underset{X_i \in \Omega_c}{Max} \|T^{Q_j,P_j} - X_i^{Q_j,P_j}\|_2^{-1}} \quad (11)$$

For each class $c \in \{1, \ldots, c\}$, the three sets of scores $(s^j(T,c))$ are combined. Two combination schemes, namely majority voting and sum rules, are evaluated. The majority voting rule consists in assigning to the query image the identity it is more frequently associated to. In case of ambiguity, expert 1 wins. The sum-rule is reported in [9] to be the best performing voting rule, despite the quite restrictive assumptions that make it optimal. Kittler *et al* [9] explained that surprising outcome by a superior resiliency to estimation errors.

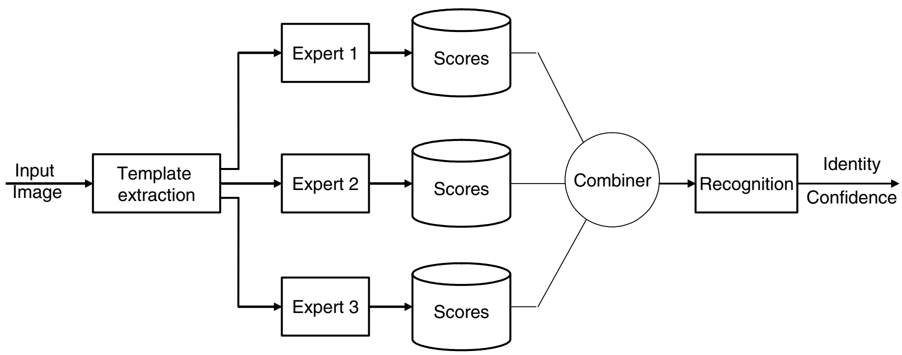

Fig. 3. Expert Voting. Recognition is performed using a rule combining the outputs of the three experts.

According to the sum-rule, a similarity measure $s(T, c)$ is obtained by adding the scores obtained from each expert:

$$s(T,c) = \sum_{j=1}^{3} s^j(T,c) \quad (12)$$

Let us call c_1 the class obtaining the highest score. The identity assigned to the query image T is c_1. If we consider that class c_2 performs the second highest score, the following confidence measure $b(T, c_1)$ is computed:

$$b(T,c_1) = log\left(\frac{s(T,c_1)}{s(T,c_2)}\right) \quad (13)$$

4 Experimental Results

In this section, we perform two series of experiments on a subset of the Asian Face Image Database PF01 [10] containing 75 persons, under neutral illumination

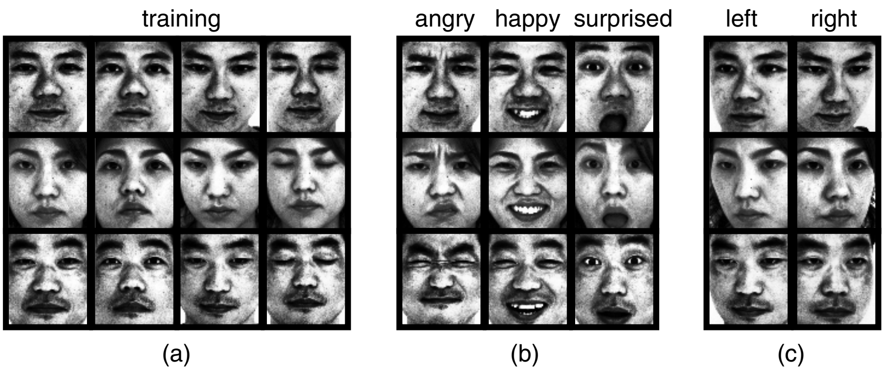

Fig. 4. Extracts of the training set (a); of the test sets used for the first experiment (b) and of the test sets used for the second experiment (c)

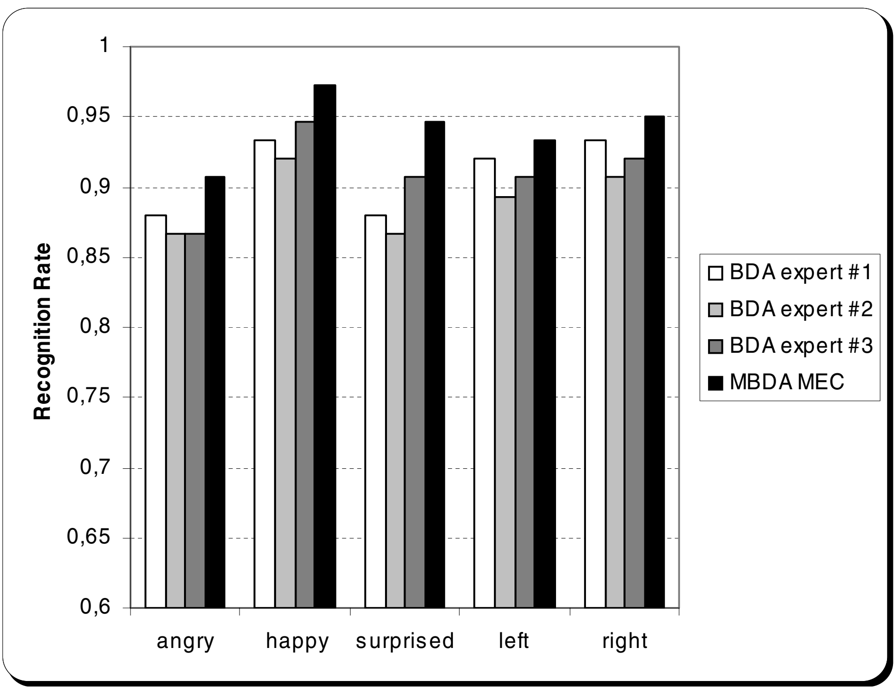

Fig. 5. Compared recognition rates of the three BDA experts, and MBDA using Multistage Expert Combining (denoted by MBDA MEC)

conditions. The experimental results show the superiority of MBDA over BDA and Modular Eigenspaces [7] in the presence of variations in the facial expressions and head poses.

The aim of the first experiment is to show that MBDA is the best performing method when dealing with drastic facial expression changes. The training set contains four views per person (see Fig. 4(a)). There are three test sets (see Fig. 4(b)): the first one contains one image per person expressing anger, the second one contains one smiling view per person, and the third one contains one surprised view per person. There are drastic variations in the facial expression from the training set to the test sets.

We design a second experiment which aims at verifying that MBDA also provides satisfying results in the presence of other sources of dissimilarities, such as head pose changes.

In the second experiment, we use the same training set as in the first experiment. There are two test sets (see Fig. 4(c)): the first one contains one image per person corresponding to a left head-pose, and the second one contains one right-head pose view per person. There are significant variations in the head pose between the training and test sets.

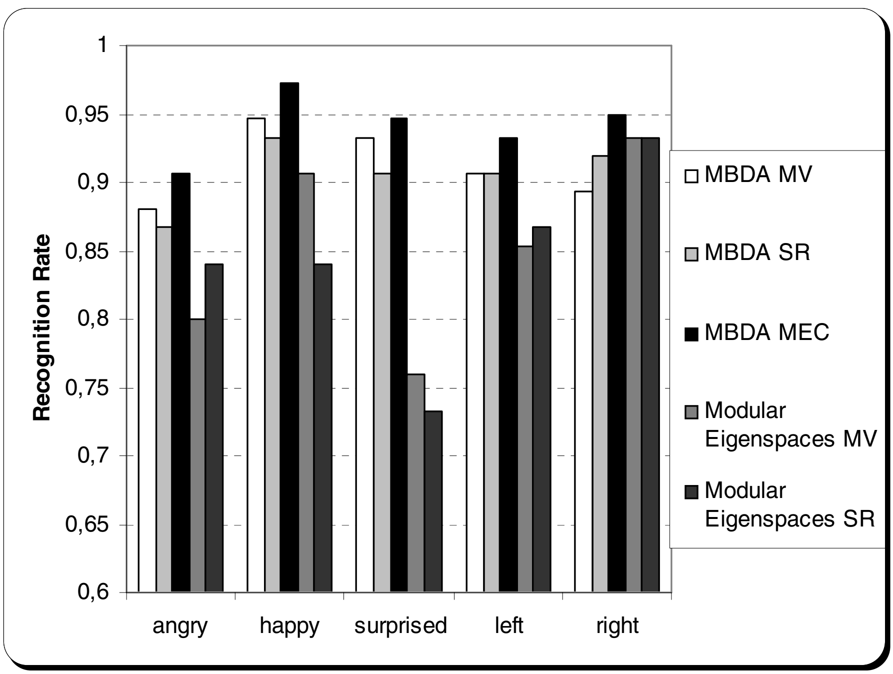

Fig. 6. Compared recognition rates of Modular Eigenspaces and MBDA using: the Majority Voting (MV) rule, the Sum Rule (SR), and the Multistage Expert Combining (MEC) scheme

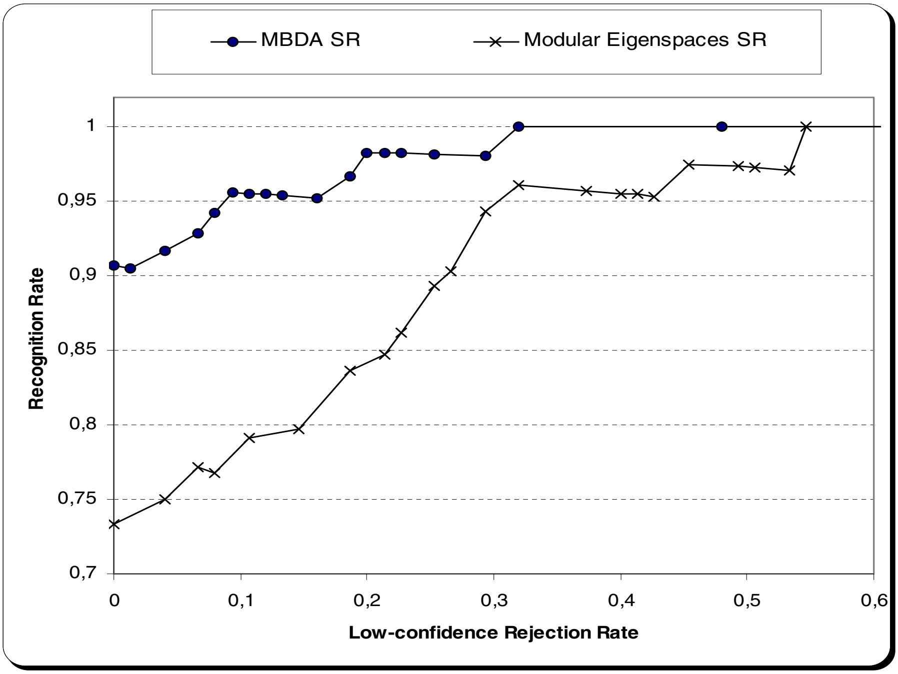

Fig. 7. Compared recognition rates of MBDA and Modular Eigenspaces versus the rate of images rejected due to low confidence, on the test set dealing with the surprised facial expression

Effect of k: The number of components can be determined by using a leave-one-out strategy ; for the considered training set the optimal k is respectively 14, 15 and 17 for experts 1, 2 and 3.

MBDA vs BDA: Fig. 5 shows the compared recognition rates for each of the three experts, and for multistage expert combining (see Fig. 2). From this figure, we can see that expert 1 outperforms the other experts in the presence of head pose variations ("left", "right"), or when the facial expression has a great impact on the aspect of the eyes and eyebrows ("angry"). However, when the facial expression results in drastic changes in the mouth aspect ("happy", "surprised"), expert 3 outperforms expert 1. In all these cases, MBDA outperforms the three experts (3.9% mean improvement of the recognition rates over the mean recognition rate of the three experts, calculated over the five test sets).

MBDA vs Modular Eigenspaces: Fig. 6 shows the compared recognition rates of MBDA and Modular Eigenspaces. From this figure, we can see that MBDA outperforms Modular Eigenspaces (with respectively 6.14% and 6.42%

mean improvement in the recognition rates, when using the Majority Voting and the Sum Rule). It has to be noted that MBDA using Multistage Expert Combining improves the recognition rates of more than 9% over both of the Modular Eigenspaces-based schemes.

Combiner Selection: From Fig. 6 we can see that for both MBDA and Modular Eigenspaces, the two expert voting schemes (see Fig. 3), namely Majority Voting (MV) and Sum Rule (SR) schemes give comparable and satisfying results. However, the Multistage Expert Combiner outperforms them, with respectively 3% and 3.52% mean improvement of the recognition rates over the majority voting and the sum rules. It has to be noted that in multistage expert combining we experimented to build $\mathcal{F}$ by applying LDA and observed that PCA provides better results.

Efficiency of the Confidence Measure: Fig. 7 gives the recognition rates of MBDA and Modular Eigenspaces, computed from the Sum Rule combiner, versus the rate of images rejected due to low confidence, on the "surprised" test set. The confidence measure used to determine the samples to reject is b, given in equation (13). The confidence measure b can be considered as effective, as most samples it rejects would be, if kept, wrongly classified. Fig. 7 also indicates that if 10% of the query images were rejected due to low-confidence, MBDA would provide more than 95.5% recognition rate, while Modular Eigenspaces would achieve less than 79.2% recognition rate.

5 Conclusion

In this paper, we have presented a Modular Bilinear Discriminant Analysis approach for face recognition. A set of experts were trained independently on specific face regions, by using a new supervised feature extractor named Bilinear Discriminant Analysis, generalizing and outperforming 2DoLDA. Then, these experts were combined to assign an identity with a confidence measure to each of the query faces.

A series of experiments was performed in order to evaluate different combination schemes, and to compare the performances of MBDA with respect to BDA and Modular Eigenspaces. The experimental results have shown that MBDA, especially with the Multistage Expert Combining scheme, is more effective than both BDA and Modular Eigenspaces, and more robust when dealing with variations in facial expressions and head poses.

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. Pentland, A.D., Moghaddam, B., Starner, T.: View-Based and Modular Eigenspaces for Face Recognition. In Proc. of the IEEE Computer Society Conference on Pattern Recognition (1994) 84–91.
8. Price, J.R., Gee, T.F.: Face Recognition Using Direct, Weighted Linear Discriminant Analysis and Modular Subspaces. Pattern Recognition. **38**(2) (2005) 209–219.
9. Kittler, J., Hatef, M., Duin, R.P.W., Matas, J.: On Combining Classifiers. IEEE Trans. on Pattern Analysis and Machine Intelligence. **20**(3) (1998) 226–239.
10. Hwang, B.W., Roh, M.C., Lee, S.W.: Performance Evaluation of Face Recognition Algorithms on Asian Face Database. In Proc. of the IEEE International Conference on Automatic Face and Gesture Recognition, Seoul, Korea (May 2004) 278–283.