IOSR Journal of Engineering
Apr. 2012, Vol. 2(4) pp: 685-691

A new approach for Face Recognition Based on PCA & Double LDA Treatment combined with SVM

A.HAJRAOUI [1], M.SABRI[2], O.BENCHAREF[3] and M.FAKIR[4]

[1]*Depart. of Information Processing and Telecommunications, Faculty of Science and Technology, Beni Mellal, Morocco*
[2]*Depart. of Information Processing and Telecommunications, Faculty of Science and Technology, Beni Mellal, Morocco*
[3]*Depart. of Information Processing and Telecommunications, Faculty of Science and Technology, Beni Mellal, Morocco*
[4]*Depart. of Information Processing and Telecommunications, Faculty of Science and Technology, Beni Mellal, Morocco*

ABSTRACT

In this paper we propose a faces recognition system. This system does not directly reproduce human vision on machine, but it seeks to find algorithms to achieve similar results by identifying a person using 2D image of his face. The descriptors used for features extraction, combine two algorithms: Principal Component Analysis (PCA) and a double Linear Discriminate Analysis (LDA) treatment. We chose the Support Vector Machine as an output classifier. Our approach has ensured a satisfactory recognition rate and a gain in terms of memory.

***Keywords*: Face recognition; Principal Component Analysis; Linear Discriminate Analysis; SVM.**

I. INTRODUCTION

Face recognition still one of the most appealing fields of computer vision. Despite the numbers of researches achieved in this area, recognizing a face remains to be a difficult problem. Existing methods are effective when the shooting conditions for test images are similar to training images. Also various constraints must be considered: the lighting, facial expression, head orientation, networking, aging etc...

A recognition system comprises two essential steps: features extraction and classification.

In this paper, we adopt a prototype recognition system based on PCA followed by a double LDA treatment on the extraction part and the Support Vector Machine SVM for classification.

II. EXTRACTION PROCESS

Extraction is the key step of the recognition process, since the performance of the entire system depends on it. In this step also known as indexing or modeling. We extract from the face image the information that enables the modelisation of the person's face, by a vector of values that characterizes the face (Feature vector).

To do this extraction, many methods have been developed. They can be grouped into three categories.

2.1. The local methods

They are also called methods in facial features, local features, or analytical methods. The analysis of the human face is given by the description of its individual parts and their geometrical relationships [1] [2] [3].

2.2. Global Methods

This class contains methods that enhance the overall properties of the face: the face is treated as a whole. They are mainly based on pixel information. These methods are:

The principal component analysis PCA, also known as the" Eigenfaces" [4] [5] [6] [7], which is a statistical projection technique to extract a sub-optimal reduced space and. And there are also the two-dimensional version PCA2D [8].

The Linear Discriminate Analysis LDA, also known as the " Fisher-faces" [9], reduces the dimensionality of space by optimizing the discrimination factor between classes. There is also the two-dimensional version LDA2Do [10].

Other algorithms have been developed using the discrete transforms such as the Discrete Cosine Transformed (DCT) [11] and Transformed Discrete Wavelet (DWT) [12].

Apart from these methods that use a linear projection, there is a strategy that exploits the multi-linear analysis of face images named Tensor-Faces [13].

2.3. The Hybrid methods

Those methods that have focused on merging the two previous methods (local and global) as: Local Feature Analysis (LFA) which is based on PCA and analysis of local characteristics.

III. DEVELOPED FACE RECOGNITION SYSTEM

The face recognition system developed and illustrated by Fig. 1 consists of two phases: the first for learning, and the second for recognition.

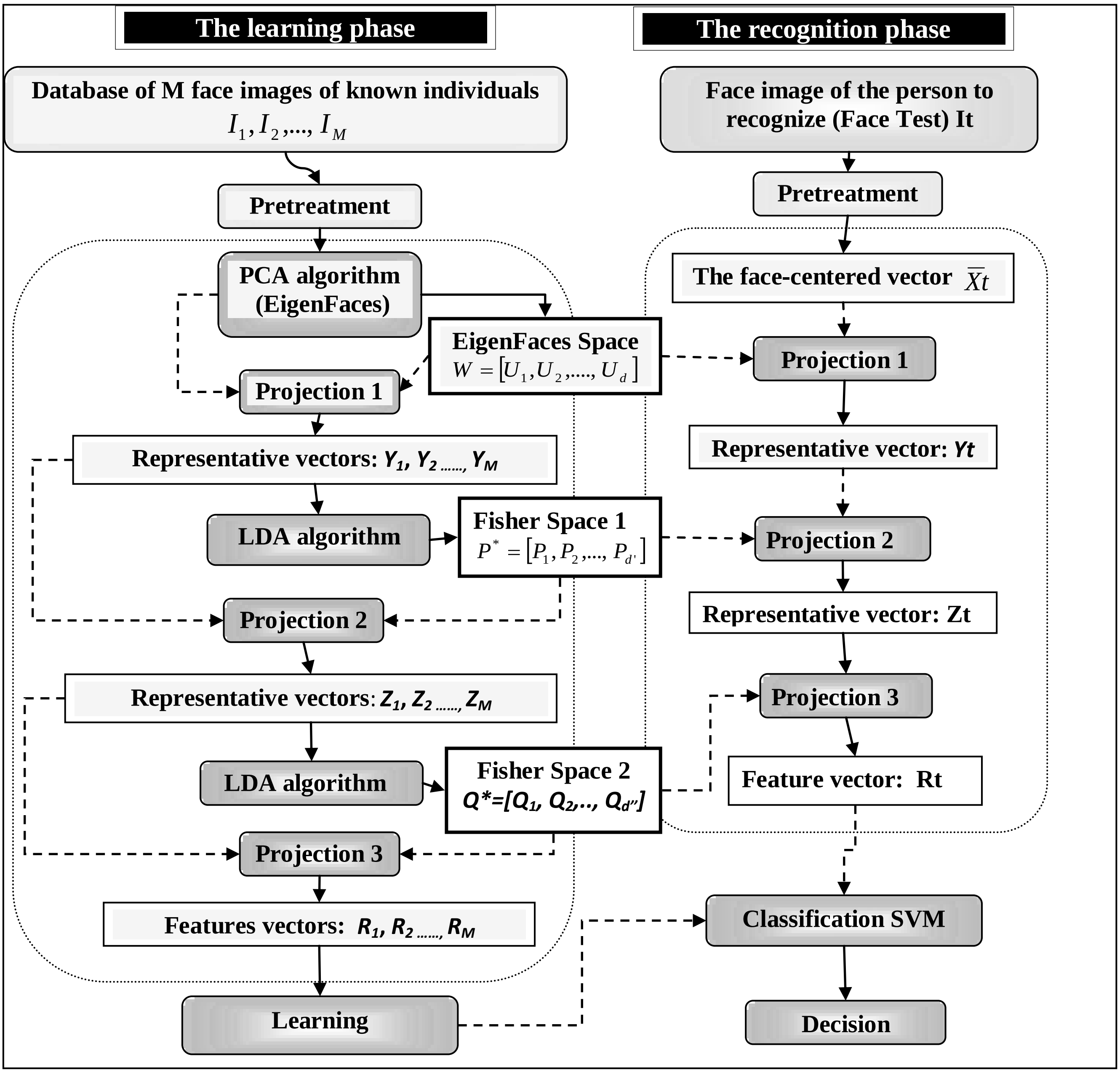

Fig. 1: Block diagram of the face recognition system developed

Our working hypotheses:

- A database Ω containing **M** face images with multiple views per person.
- All images that match the same person compose a class Ω_c and the database contains **C** classes.
- Each class Ω_c contains n_c image of size **n×m**.

3.1. Learning Process

This process provides the steps required to have efficacy learning like: pretreatment, extraction, elimination of redundancies

3.1.1. Pretreatment

The images of the database are subject to a pretreatment. If the images are in color, they must be converted to grayscale. Then they will be resized to a sufficient and identical size.

3.1.2. Extraction of feature vectors

The extraction of feature vectors of faces of the database is obtained by three successive projections of face images in three new areas:

a. Projection 1: projection in Eigenfaces space

The PCA method (Eigenfaces) [6] and [7] provides a new projection space generated by the Eigen faces, wherein an image can be reduced to a vector of dimension much lower

that ensures optimal reconstruction in the opposite direction.
To determine the Eigenfaces, a treatment is carried out as follows:
From images of faces $I_1, I_2,...,I_M$, we construct the data matrix of dimension (N, M): $T=(X_1, X_2,.., X_i,.., X_M)$.
Where each X_i is a column vector of dimension N (N = n × m) representing the image i-face after concatenation of its n rows (or columns m). We determine the average vector of the training set by the following expression:

$$\mu = \frac{1}{M}\sum_{i=1}^{M} X_i \qquad (1)$$

This vector is subtracted from each vector image to determine the vector-centered face given by the following expression:

$$\overline{X}_i = X_i - \mu \qquad (2)$$

We note $A = (\overline{X}_1, \overline{X}_2,..., \overline{X}_i ... \overline{X}_M)$ the matrix-vector of centered faces.
We compute the eigenvectors v_i of the covariance matrix S_T :

$$S_T = A^T.A \qquad (3)$$

Then, the Eigen faces Ui are determined by the expression:

$$U_i = \sum_{k=1}^{M} \overline{X}_k . v_{ik} \qquad (4)$$

With: v_{ik} denotes the k^{ith} component of the i^{ith} eigenvector of S_T.
After normalization of the Eigen faces Ui:

$$U_i^T.U_j = 0, i \neq j, i, j = 1,\ldots,M \qquad (5)$$

We choose the *d* eigenfaces of the corresponding dominant eigenvalues λi, to build unique space faces defined by the projection matrix:

$$W = [U_1, U_2,...., U_d] \qquad (6)$$

The dimension d of the space is determined in a way to minimize the size of the new space without losing too much information. The solution generally adopted is to select the number of vectors as the fraction of the total variance that represents a given percentage of information. This fraction is given by:

$$q_d = (\sum_{i=1}^{d} \lambda_i)/(\sum_{i=1}^{M} \lambda_i) \qquad (7)$$

Finally, each face-centered vector $\overline{X}_i$ of dimension N is reduced to a vector Y_i in the new space of dimension *d*, by the following projection:

$$Y_i = [y_1, y_2,..., y_d]^T = W^T.\overline{X}_i \quad i = 1,\ldots,M \qquad (8)$$

Where Y_i denotes the vector representing the vector associated with the image face I_i.

b. Projection 2: projection in the Fisher space 1

Now we determine a second projection space (Fisher space). To do that, we apply the algorithm of linear discriminate analysis (LDA) [9] on the set of vectors Y_i ($i = 1,\ldots,M$) obtained after a projection. This algorithm calculates a projection space (Fisher Space) that maximizes the distance between different classes while minimizing the distance between elements from the same class.
It comes to compute the matrix P* that maximizes the following generalized Fisher criterion.

$$P^* = \underset{P \in R^{m \times d3}}{Arg \max} \cdot \frac{|P^T S_b P|}{|P^T S_W P|} \qquad (9)$$

Where S_w and S_b denote the covariance matrices intra and inter classes of generalized database containing the vectors $Y_1, Y_2,\ldots.., Y_M$:

$$S_W = \sum_{c=1}^{C} \sum_{Yi \in \Omega c} (Y_i - \overline{Y}_c).(Y_i - \overline{Y}_c)^T \qquad (10)$$

$$S_b = \sum_{c=1}^{C} n_c (\overline{Y}_c - \overline{Y}).(\overline{Y}_c - \overline{Y})^T \qquad (11)$$

$\overline{Y}_c$ Denotes the mean vector of n_c vectors Y_i in class Ω_c and the mean vector of all vectors Yi of the database Ω.
Under the assumption that Sw is invertible (it is easy to show that this assumption is generally verified), and of columns of the matrix P * are the ***d'*** first eigenvectors of the matrix $S_w^{-1}. S_b$ (That means those associated with the largest eigenvalues).
After determination of Fisher space, defined by the projection matrix P * = [P1, P2, .., Pd '], we apply the linear projection of vectors Yi of size D, using the following formula:

$$Z_i = [z_1, z_2,..., z_{d'}]^T = P^{*T}.Y_i \quad i = 1,\ldots,M$$

(12)
Where the vector Z of size d' means the signature vector associated with the vector Yi (by transitivity, associated with the image face-Ii).

c. Projection 3: projection in the Fisher space 2

The third projection, with the same algorithm LDA is made in order to improve further the discrimination between the data. And this effect will be justified by the results presented in the following paragraph.
By repeating the same operations of the projection 2, but this time on vectors Zi () result of the projection 2, we obtain the third projection space (space Fisher 2) defined by the projection matrix: $Q^*=[Q_1, Q_2,.., Q_{d''}]$.
The projection of the vectors Zi in this third space is given by the expression:

$$R_i = [r_1, r_2,..., r_{d''}]^T = Q^{*T}.Z_i \quad i = 1,\ldots,M \qquad (13)$$

Give the feature vectors of the faces-images I_i (of the database) which are the R_i vectors of dimension: ***d''***: $R_1, R_2,\ldots.,R_M$.

3.1.3. Learning

This is the final step in the learning phase, which consists of storing the feature vectors R_i of the faces-images I_i for use in the classification stage.

It should also memorize the projection matrices W, P* and Q*.

3.2. Recognition phase (tests)

In the recognition phase (tests), the face image "***It***" of the person to recognize is subject to the same pre-treatment applied in the learning phase. Then, we extract the image feature vector *Rt.*

3.2.1. Extraction of the characteristic vector of the face image tests

The determination of the vector Rt is obtained by the submission of the image *It* (after conversion to the column vector Xt) to the three following projections:

a. **Projection 1 :**

$$Yt = W^T.\bar{X}t \qquad (14)$$

Where the vector $\bar{X}t$ denotes the centered vector-face image of It (Equation 2).

b. **Projection 2 :**

$$Zt = P^{*T}.Yt \qquad (15)$$

c. **Projection 3 :**

$$Rt = Q^{*T}.Zt \qquad (16)$$

3.2.2. Classification

Classification is the assignment of a specific class to the face test: class here represents a person with face images in the database. This assignment requires the introduction of a similarity measure. In this work, we propose the integration of the Support Vector Machine classifier (SVM).

The support vector machine (SVM) is a universal constructive learning procedure based on the statistical learning theory. Originally it was worked out for linear two-class classification with margin, where margin means the minimal distance from the separating hyper-plane to the closest data points. SVM learning machine seeks an optimal separating hyper-plane, where the margin is maximal. An important and unique feature of this approach is that the solution is based only on those data points, which are at the margin. These points are called support vectors. The linear SVM can be extended to nonlinear one when the problem is transformed into a feature space using a set of nonlinear basis functions. In the feature space - which can be very high dimensional - the data points can be separated linearly [14] and [15].

IV. Results and Discussion

4.1. The database of faces

The faces database used is the AT & T data base (formerly ORL) [16]. This database contains face images of 40 people, with 10 images for each (total 400 images). For most subjects, the images were taken at different times, ranging from lighting, facial expressions (open / closed eyes, smiling / not smiling), head of the poses and facial details (glasses / no glasses). The image size is 92x 112.Un extract of this database is given in Fig. 2.

Fig. 2: Extract from the database of AT & T data base.

To evaluate the developed approach we took in the learning phase, 5 images for each (total 200 images) and the rest (200 images) for the recognition phase (test).

So we have:

- A database Ω, containing 200 image-faces (M = 200).
- 40 classes (C = 40) and each class Ωc, contains 5 images (n_c = 5).

4.2. Extraction of feature vectors

a. Projection in the eigenfaces space

After determination of the 200 eigenfaces used in the learning process (Fig. 3). We had the choice of the dimension "d" of the space (the number of faces that will generate this space).

Fig. 3: The 17 first eigenfaces and average face.

To get an idea of the dimension to choose, we have drawn the curve (Fig. 4) described by formula (7).

From this curve "d = 14". First, because it ensures a percentage of 67.49% of information enough to represent the information and secondly, if we evaluate the recognition system using only the PCA, we obtain a recognition rate of 71.5, taking as the space dimension d = fourteen. This result is derived from the curve in Fig. 5.

Therefore, after projection of the vectors-face of size (112x92) in the new space, we will obtain the representative vectors Yi (i = 1... 200) of dimension 14.

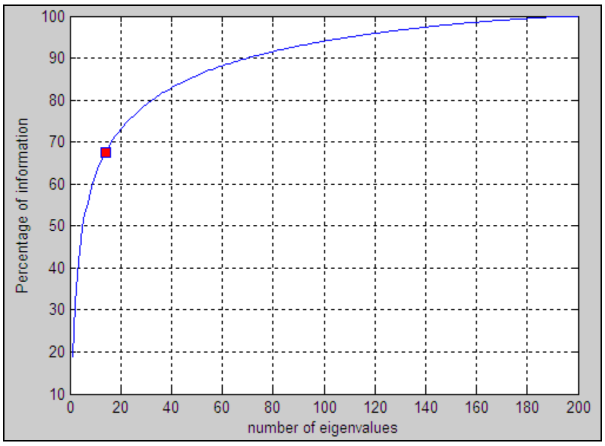

Fig. 4: Percentage of information according to the number of eigenvalues

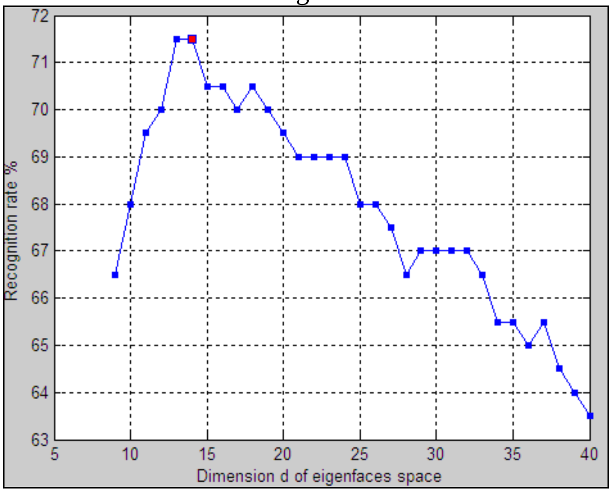

Fig. 5: Variation of recognition rate depending on the dimension d of eigenfaces space.

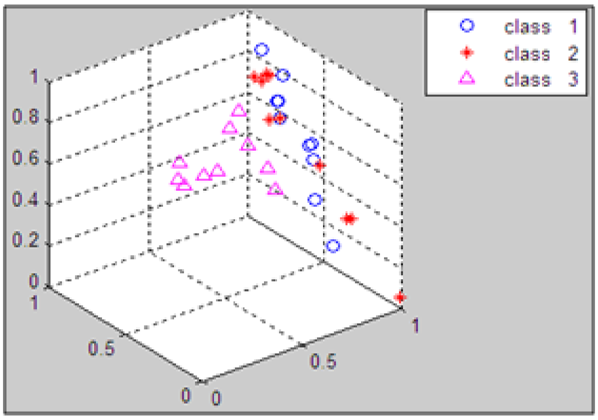

Fig. 6: Examples of face images projected in the eigenfaces space (10 images per class)

b. Projection into the first Fisher space

If the power of CPA in reducing the representation space without loss of information is justified, the discrimination between classes is not optimal. Indeed, there is an overlap between the vectors Yi representative of different classes of the database. To illustrate this non-discrimination, Fig. 6 represents the projection of face images of three different classes (3 different people). We take arbitrary 3-axes-space.

To resolve this problem, we conducted a second projection of the representative vectors Yi (i = 1... 200) in the Fisher space. Figure 7 illustrates the projection in the Fisher space of the vectors corresponding to the same Yi face used in Fig. 6. Fig. 7 shows the discriminating power of the LDA treatment.

The Determination of the dimension d' of the first Fisher space is made empirically. We trace the curve in Fig. 8 that gives statistics of recognition rate depending on the dimension d' of the space for an FRS system using only the PCA + LDA process.

From this curve, we see that the recognition rate is maximal at 87.5%, for a dimension d' = 13.

After calculation of the first Fisher space ($P^* = [P_1, P_2,.., P_{13}]$), we apply the linear projection of the vectors Yi (size 14) in this space, to obtain the representative vectors Zi ($i = 1, \ldots, 200$) of size 13.

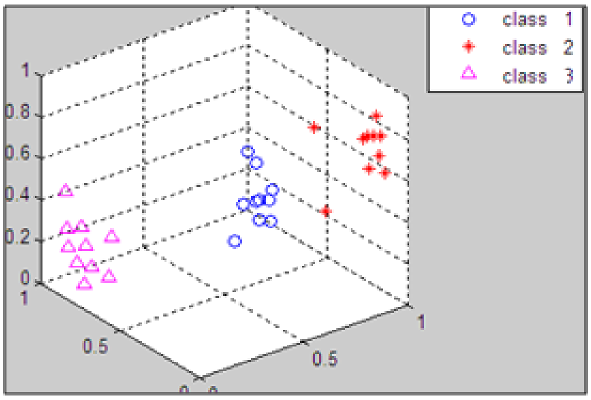

Fig. 7: Examples of projection of the Yi vectors in the Fisher space 1

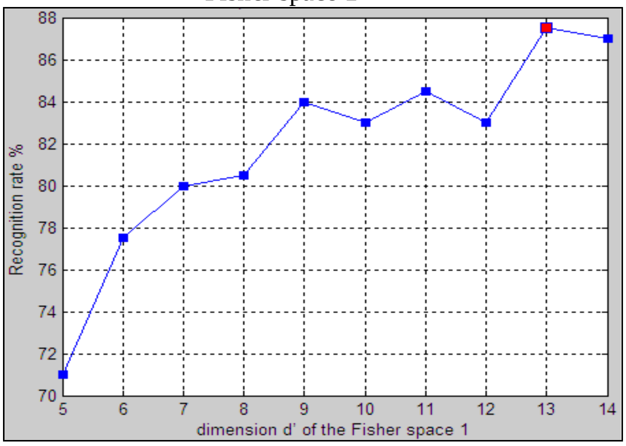

Fig. 8: Variation of recognition rate depending on the dimension d' of the first Fisher space.

c. Projection in the second Fisher space.

This projection is intended to improve discrimination between data. This is justified by Fig. 9 which gives the result of the projection in the first Fisher space of the Zi vectors corresponding to the same vectors Yi in Fig. 7 (by transitivity, corresponding to the same classes of face images of Fig. 6).

So, we clearly observe the effect of this projection which increases the distance between two different classes and decreases the distance between the data in the same class.

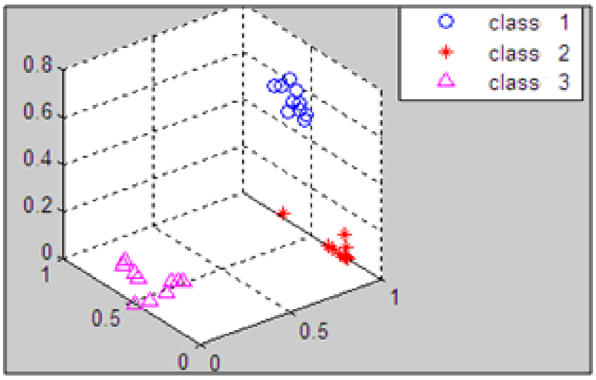

Fig. 9: Examples of projection of the Zi vectors in the Fisher space 2

To complete the extraction step. We perform the projection of the Zi vectors in the second Fisher space. To finally have the feature vectors Ri ($i = 1, \ldots 200$).

The size of the vectors Ri (implicitly the dimension of first Fisher space') is made from the curve in Fig. 10 that plots the variation of recognition system developed according to the dimension d".

The maximum rate is 92.5 %, for a dimension d" = 10.

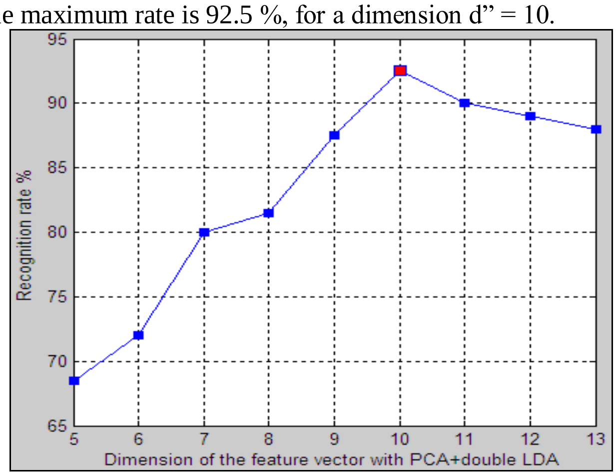

Fig. 10: Variation of recognition rate depending on the dimension of the feature vector"

4.3. Recapitulation of tests results

To test and validate of the developed recognition system (PCA + double LDA + SVM), we tried to evaluate the performance of this system compared to those of two other recognition systems. The first using only the PCA as extraction method and the second PCA + LDA. This assessment is summarized in Table 1.

From this table, we see that the developed system provides the best performance:

- A recognition rate of (92.5 %).
- A small size of the feature vector (gain in memory space at the time of storage).
- A slight increase in the learning time and recognition time (test).

Table 1: Performance comparison of three different appointment systems

	PCA	PCA + LDA	PCA + double LDA + SVM
Recognition rate	71.5 %	87.5%	92.5 %

Dimension of feature vector	14	13	10
Learning time	1.779 s	1,804 s	1,833 s
Recognition time	0.092 s	0.107 s	0.109 s

The values of learning time and recognition time are taken on a computer with the characteristics:

- CPU speed: 2.4 GHz
- Capacity of the RAM: 1.99 Go

V. CONCLUSION

The approach presented in this paper has led to good performance in terms of recognition rate and memory requirements. And this, thanks to the effect of good discrimination between features vectors performed by double LDA and the two high classification of SVM

In perspective, we propose to apply it to other faces databases. This will validate the robustness of the algorithm and establish decision rules for rejecting faces which are not registered in the database. Therefore minimizing the rate of bad detections while maximizing the recognition rate.

REFERENCES

[1] Khashman A. Intelligent Local Face Recognition. pp. 236. Recent Advances in Face Recognition I-Tech. Vienna, Austria (2008).

[2] Yu S, Shiguang S, Xilin C et Wen G. Patch-Based Gabor Fisher Classifier for Face Recognition. IEEE (2006)

[3] Campadelli P, Lanzarotti R et Lipori G. Automatic Facial Feature Extraction for Face Recognition. Recent Advances in Face Recognition, I-Tech, Vienna, Austria. (2007).

[4] Sirovich L et Kirby M. A Low-Dimensonal Procedure for the characterisation of Human Faces. Journal of Optical Society of America, vol ,no 3, p.: 519-524. (1987).

[5] Sirovich L et Kirby M. Application of the Karhunen-Loève for the characterisation of Human Faces. IEEE Transactions on Pattern and Machine intelligence, vol 12, no 1,p.:103-108. (1990).

[6] Turk M. et Pentland A.. Eigenfaces for Recognition. Journal of Cognitive Neuroscience, vol 3no 1, p. :71–86. (1991).

[7] Turk M et Pentland A. Face Recognition Using Eigenfaces. IEEE. P.:586-591. (1991).

[8] Yang J., Zhang D. et Frangi D. Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition. IEEE Trans. On PAMI, 26(1) :131–137. (2004).

[9] Belhumeur P.N., Hespanha J.P. et Kriegman D.J. Eigenfaces vs Fisherfaces: Recognition Using Class Specific Linear Projection. IEEE Trans. on PAMI, 19(7), p. :711–720. (1997).

[10] Visani M., Garcia C. et Jolion J.M. Two-Dimensional-Oriented Linear Discriminant Analysis for Face Recognition. Proc. of the Int.

Conf. On Computer Vision and Graphics ICCVG’04, Computational Imaging and Vision, Varsovie, Pologne. (2004).
[11] Kresimir D., Sonja G. and Mislav G. Image Compression in Face Recognition - a Literature Survey. Recent Advances in Face Recognition , I-Tech, Vienna, Austria. (2008).
[12] Heng F.L., Kah P.S., Li-Minn A. et Siew W.C. New Parallel Models for Face Recognition. Recent Advances in Face Recognition , I-Tech, Vienna, Austria. (2008).
[13] Vasilescu,M.A.O., et Terzopoulos,D. Multilinear Analysis of Image Eensembles: TensorFaces. Proceedings of European. (2002).
[14] V.N. Vapnik, "The Nature of Statistical Learning Theory", Springer-Verlag, New York, ISBN 0-387-94559-8, 1995.
[15] B. Scholkopf, C. Burges, and A. J. Smola. Support Vector Learning, MIT Press, Cambridge, Massachusetts, chapter 12, pp 185-208, 1999.
[16] www.cl.cam.ac.uk/Research/DTG/attarchive:pub/data/