Linear Discriminant Analysis for Face Recognition

Fatma zohra CHELALI
Speech communication and signal processing laboratory
Electronics and computer Sciences Faculty
Houari Boumedienne University of sciences and technologies,USTHB
Box n°:32 El Alia, 16111, Algiers, Algeria
Fax : (213) 21247187,email: Chelali_zohra@yahoo.fr

A. DJERADI and R. DJERADI
Speech communication and signal processing laboratory
Electronics and computer Sciences Faculty
Houari Boumedienne University of sciences and technologies,USTHB
Box n°:32 El Alia, 16111, Algiers, Algeria
r_djeradi@yahoo.fr

Abstract— **Face is the most common biometric identifier used by humans. During the past thirty years, a number of face recognitign techniques have been proposed, all of these methods focus on image-based face recognition that use a still image as input data.**

In this paper, Linear Discriminant Analysis (LDA) which is also called fisherface is an appearance-based technique used for the dimensionality reduction and recorded a great performance in face recognition. This method works on the same principle as the eigenface method (PCA).it performs dimensionality reduction while preserving as much of the class discriminatory information as possible.

LDA makes use of projections of training images into a subspace defined by the fisher faces known as fiherspace. Recognition is performed by projecting a new face onto the fisher space, The KNN algorithm is then applied for identification.

Keywords- ***Face recognitIon, LDA, PCA, fisher face, pattern recognition.***

I. INTRODUCTION

Face recognition from images is a sub-area of the general object recognition problem. It is of particular interest in a wide variety of applications [1].Recognition has a lot of applications such us: access control, video surveillance, verification for personal identification such us driver's licence and credits cards [1].

A genaral statement of the problem of machine recognition faces can be formulated as follows: given still or video images of a scene, identify or verify one or more persons in the scene using a stored database [2].

Recognition of human faces is a very complex problem: lighting conditions, facial expressions vary from time to time; faces may appear at different scales, positions and orientations, facial hair [1]. So it presents difficulties like interclass similarity and intraclass variability due to head pose, illumination conditions, and expressions as shown in figure (1).

Since face recognition in general settings is difficult, an application system typically restricts one of many aspects, including the environement in which the recognition system will take place (fixed location, fixed lighting, uniform background, single face,......etc), the allowable face change (neutral expressions, negligible aging), the number of individuals to be matched against, and the viewing condition (frontal view, no occlusions....etc) [1].

Figure1. Intraclass variability

A wide variety approaches to machine recognition of faces has been published in the litterature. A lot of methods proposed for face recognition; there are divided into three categories: local feature matching method (analytic or geometric), holistic

978-1-4244-3757-3/09/$25.00 ©2009 IEEE

matching method (appearance based methods) and hybrid methods.

The geometric feature based method uses properties of facial features such us eyes, nose, mouth and their relations for face recognition descriptors (fig 2), this was the first approach for face recognition, and a large number of geometric feature based methods have been proposed in 1980.

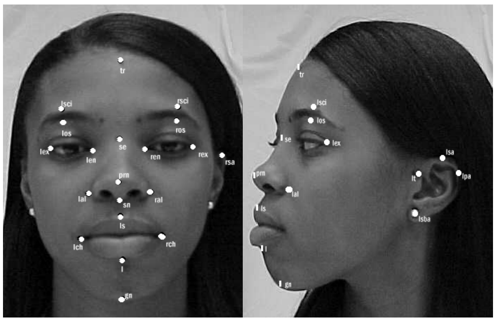


Figure2. The geometric feature based methods

The "Appearance based methods", also known as template-based methods are based on the extraction of holistic features. These methods" have been achieved since 1990.

Template based technique such us PCA, LDA, ICA follow the subspace method called eigenface originated by M.Turk &A. Pentland [2] [3]. This method is based on the Karhunen-loeve transformation (KLT), which is known as PCA, Principal Component Analysis.

This approach operates directly on an image-based representation. It extracts features into a subspace derived from training images [2].Subspace analysis is done by projecting an image into a lower dimensional space and after that recognition is performed by measuring the distance between known images and the image to be recognized [2].

The most popular appearance based sub space projections methods are: Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) and Independant Component Analysis (ICA) [2].

Principal Component Analysis is one of the most successful technique that have been used in image recognition and compression, M.Turk &A. Pentland proposed in 1991 : the PCA technique for face recognition [3], known as eigen face method , use the Karhunen-loeve transform (KLT) for the representation and recognition of face.

the purpose of PCA is to reduce the large dimensionnality of the data space (observed variables or training image) to the smaller dimensionnality of feature space (eigenfaces or independant variables).it's used to determine the most representative features among data points onto the subspace that has the largest variance among all-p dimensional subspace [5][6][14].

Another technique has been studied in recent year (1997) is LDA [6] [9]; which is also known as Fisher's discriminant Analysis.

LDA unlike PCA, [9] [10] uses the class information and finds a set of vectors that maximize the between-class scatter while minimizing the within-class scatter.

The third appearance-based statistical method is Independant Component Analysis (ICA) which is a method for transforming an observed multidimensional random vector into components that are statistically as independant from each oher as possible [6].

We focus on the second based-apperance methods LDA: this approach includes two phases: training and classification. In the training phase, a fisher space is established from the training samples and the training faces are projected onto the same subspace. The optimal projection (transformation) can be readily computed by applying the Eigen decomposition on the scatter matrices. In the classification phase, an input face is projected into the fisher space and classified using the euclidean distance as a similarity measure.

The LDA algorithm is tested using the computer vision Research projects dataset [16] .it contains 395 individuals; the images were collected between 1994 and 1996.

The paper is organized as follows: section II presents the face space in comparison to image space; section III describes the PCA and LDA algorithms, section IV presents experiment results of LDA. And we draw some preliminary conclusion in sectionV.

II. FACE SPACE

The two projections methods are called sub-space analysis methods. In this section we'll describe the sub space projection technique widely used for face recognition.

A two dimensional image [10] image $\Gamma(x, y)$ of size (m by n) pixels can be viewed as a vector (or a point) in high dimensional space. We create a vector from an array by concatenating its columns, thus getting a vector $X = [x_1 \ x_2 \ \ldots\ldots\ldots x_N]^T$, where N=m*n. Each pixel of the image corresponds to a coordinate in N-dimensional space. We'll refer to this space as space image. Since a space is highly dimensional, recognition in it is unfeasible [10].

However, if an image of an object is a point in image space, a collection of M images of the same sort of an object represents a set of points in the same subspace of the original image space.

Appearance-based object recognition deals with the following questions [10] [12]:

What is the relashionship between points in image space that correspond to all images of a particular face? Is it possible to efficiently characterize this subset of all possible images? What is the "shape" of this subset? [10].

An exemple of building a general subspace appearance-based face recognition system can be seen in figure (3).

978-1-4244-3757-3/09/$25.00 ©2009 IEEE

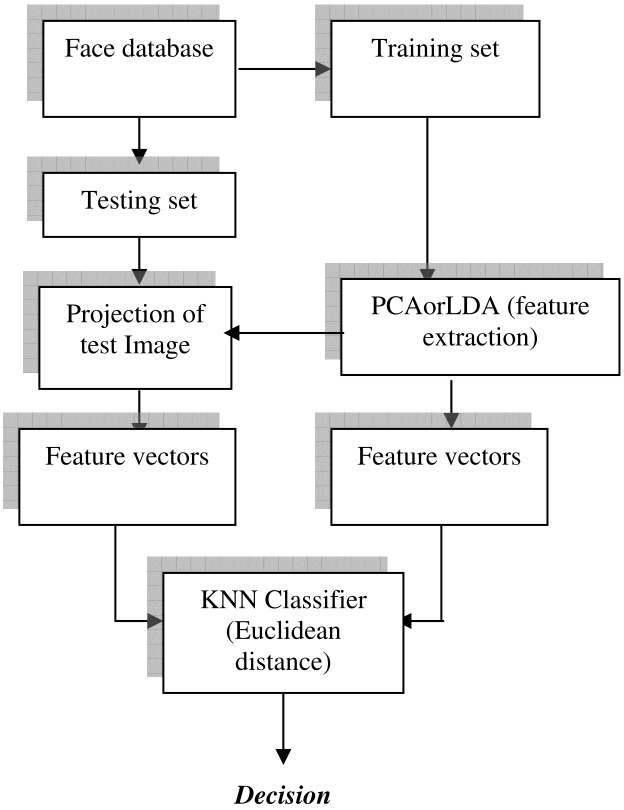


Figure 3. PCA and LDA approach for face recognition [12]

III. ALGORITHMS

A. *Principal Component Analysis (eigenface approach)*

Principal Component Analysis (PCA) is a dimensionality reduction which is used for compression and recognition problems. It is also known as eigenspace projection and closely related to popular signal processing technique known as the karhunen loeve transformation KLT [3] [9] [10].

PCA is a method to efficiently represent a collection of sample points, reducing the dimensionality of the description by projecting the points onto the principal axes, where an orthonormal set of axes points in the direction of maximum covariance in the data [3] [10].these vectors best account for the distribution of face images within the entire image space (fig4) [9] [10] [[15]].

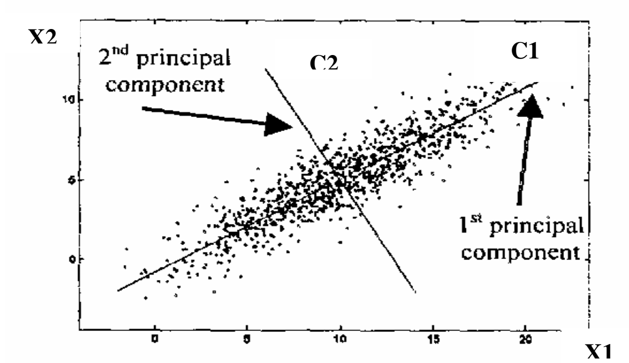


Figure 4. Projection of the original data onto the 02 principal axes, (Turk et pentland, 1991)

We'll describe the PCA algorithm as proposed by M.Turk et A.Pentland from Mit Media laboratory in 1991.

Training phase: initialization of the system

Let the training set of M face images be $\Gamma_1 ; \Gamma_2 ; \Gamma_3 ....\Gamma_M$, each vector Γ_i represents an image of size (N*M).

The average face of the training is defined by:

$$\Psi = \frac{1}{M} \sum_{n=1}^{M} \Gamma_n \tag{1}$$

Each face differs from the average face by the vector

$$\Phi_i = \Gamma_i - \Psi \tag{2}$$

Where i =1 to M.

We rearrange then these vectors in a matrix $\phi = [\phi_i ....\phi_M]$ of dimension N*M.

Matrix ϕ has zero-mean (mean value subtracted) vectors of each training face image in its columns; this is known as a translation of the origin to the mean face.

The second step is to find a set of the M-1 orthogonal vectors ui; which best describes the distribution of the input data in a least squars sence [9].We find the covariance matrix C :

$$C = \frac{1}{M} . \sum_{n=1}^{M} \Phi^T{}_n . \Phi_n \tag{3}$$

The eigenvalues (λ_i) and the eigenvectors (ui) are obtained from covariance matrix. C is real and symmetric. λ is a diagonal matrix with eigenvalues on its main diagonal [3] [9] [[15]][14].

The eigenvectors ui are sorted according to their corresponding eigenvalues. Larger value means that associated eigenvector captures more of the data variance.

In PCA, we keep only the best k Eigen vectors (with the highest k eigenvalues) [3] [9].

The directions with the largest variance are the most principal (Component C1 for example in figure (4)).The eigen faces are essentially the basis vectors of the eigen face decomposition [14].

We can then reconstruct faces in the new space with the best eigenvectors (respectively eigenvalues) that verify the equation:

$$[(\Gamma_i - \psi)]_{rc} = \sum_{k=1}^{L} [\Gamma_i - \psi]_{re} . u_{ik} \tag{4}$$

$[(\Gamma_i - \psi)]_{rc}$ means the reconstructed (projected)images

$[(\Gamma_i - \psi)]_{re}$ means the real projection of initial images

For each class or person we compute the matrix of prototyps by calculating the mean image in eigen face space, the corresponding matrix is called ΩK [14].

$$\Omega_K = [\Gamma_{moy1} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \Gamma_{moyK}] \tag{5}$$

Recognition phase:

After creating the eigenspace, we can proceed to recognition using eigenfaces. Given a new image of an individual (Γx), the pixels are concatenated as the same way as the training image, where the mean image ψ is subtracted $\Gamma_x - \psi$ and the result is projected onto the face space:

$$\Gamma_{x\ projected} = \sum_{k=1}^{L} (\Gamma_x - \psi).u_{ik} = \Omega_i \tag{6}$$

this is the projection of an unknown face into the eigen face space.

Ωi is then used to establish the predefined face classes best describes the new face. However, we will find the face class k that minimizes the euclidian distance:

$$\varepsilon_k = \sqrt{\|\Omega_i - \Omega_K\|^2} \tag{7}$$

Where ΩK is a vector describing the kth face class.A face is classified as belonging to a certain class when the minimum ε_k is below some certain threshold [3] [14].

B. Linear Discriminant Analysis (Fisherface approach)

In this section, we briefly review the conventional LDA face recognition approach"Fisherface". Fisher faces method [10] derives from Fisher's linear disriminant analyis (FLD ou LDA); it works on the same principle as the eigenfaces method.

For appearance-based face recognition, a 2Dface image is viewed as a vector with length N in the high dimensional image space. The training set contains M samples $\{x_i\}_{i=1}^{M}$ belonging to C individual classes $\{xj\}_{j=1}^{C}$ [13].

LDA tries to find a set of projecting vectors w best discriminating different classes. According to the Fisher criteria, it can be achieved by maximizing the ratio of determinant of the between-class scatter matrix Sb and the determinant of the within-class scatter matrix [13].

The objective of LDA (fig5)is to perform dimensionality reduction while preserving as much of the class discriminatory information as possible by finding direction along which the classes are best separated (fig 6 and fig7).In the Fisherface method [9], the face data is first projected to a PCA subspace spanned by M-C largest eigenfaces.

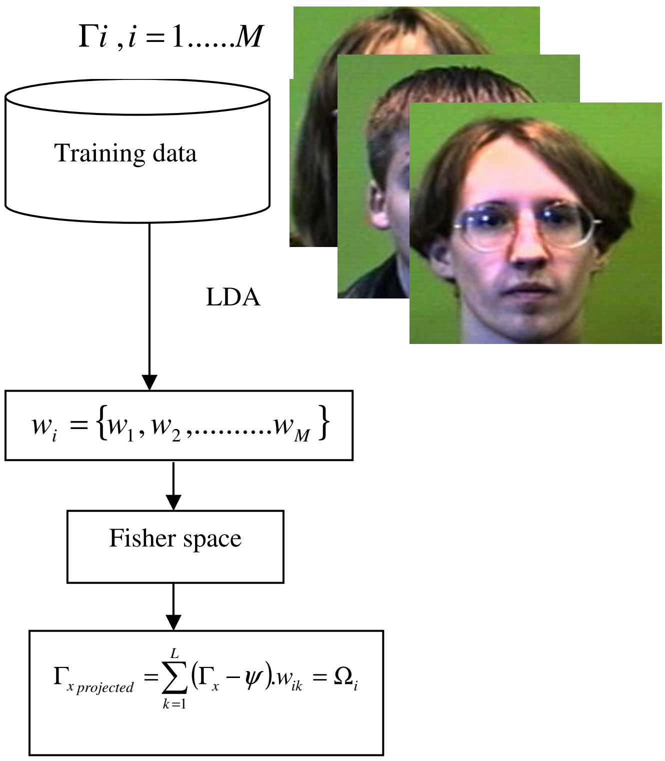


Figure 5. Principle of LDA approach

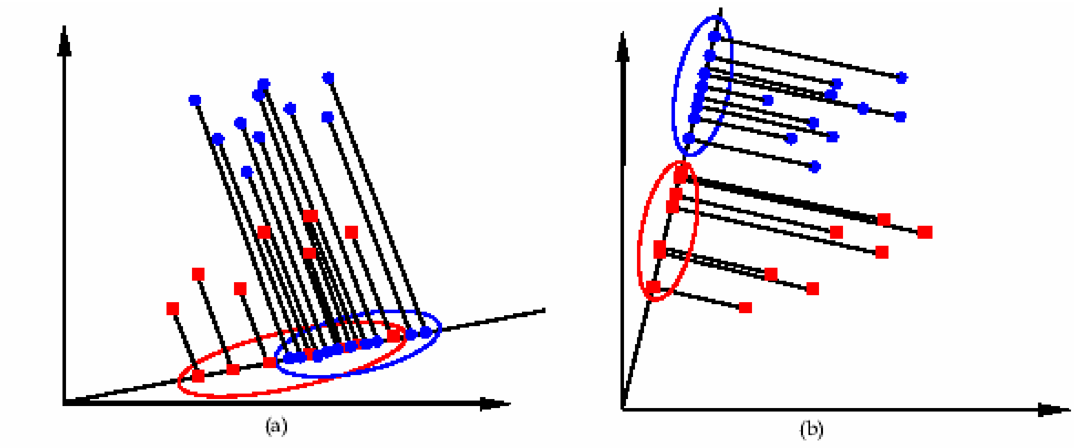
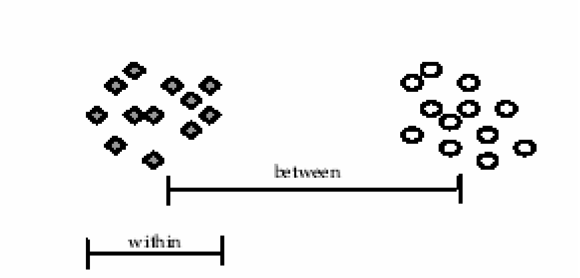


Figure 6 . (a) Points mixed when projected onto a line. (b) Points separated when projected onto another line [10]



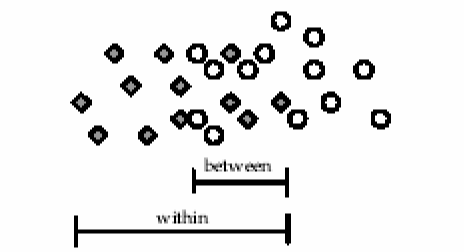


Figure 7 . (a) Good class separation [10] (b) Bad class separation[10]

Training phase:

LDA finds the vectors in the underlying space that best discriminate among classes [9][10].

For all samples of all classes, the between class scatter matrix Sb and the within class scatter Sw are defined by:

$$S_b = \sum_{i=1}^{C} q_i (\Psi_{ci} - \Psi)^T.(\Psi_{ci} - \Psi) \tag{8}$$

$$S_w = \sum_{i=1}^{C} \sum_{\Gamma_k \in C_i} (\Gamma_k - \Psi_{ci})^T * (\Gamma_k - \Psi_{ci}) \tag{9}$$

q_i is the number of training samples in class i, C the number of distinct class.

Ψ_{ci} is the mean vector of samples belonging to class i defined by the equation 10:

$$\Psi_{ci} = \frac{1}{q_i} \sum_{k=1}^{q_i} \Gamma_k \tag{10}$$

$\Psi = \frac{1}{M} \sum_{n=1}^{M} \Gamma_n$ is the mean of the set of training images

We used matrix of dimension N*M in equation 8 and 9 to calculate the within class scatter of dimension (M*M) that deals with covariance between individuals.

The within class scatter Sw represents how face images are distributed closely within classes and between class scatter matrix Sb how classes are separated from each other [12].

The goal of LDA is to maximize Sb while minimizing Sw; the images in the training set are divided into the corresponding classes. LDA finds a set of vectors wLDA such that the fisher discriminant criterion is maximized.

$$w = \arg\max_T (J(T))$$

$$\Rightarrow \max(J(T)) = \frac{|T^T S_b T|}{|T^t S_w T|} \Bigg| T = w \tag{11}$$

w can be constructed by calculating the eigenvectors of the matrix $S_w^{-1}.S_b$:

$$w = eig(S_w^{-1}.S_b) \tag{12}$$

When face images are projected into the discriminant vectors w, face images should be distributed closely within classes and should be separated between classes as much as possible [12].

These eigenvectors are called the fisher faces [12].

Fisherface approach is similar to eigenface approach, which makes use of projection of training images into a subspace.

Recognition phase:

Given a test image (Γx), where the mean image ψ is subtracted $\Gamma x - \psi$ and the result ϕ_t is projected onto the face space and identified using the euclidean distance as a similarity measure.

$$g(\phi_t) = .\phi_t.w^T$$

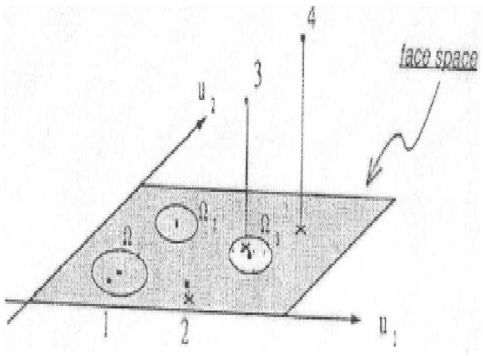


Figure 8 . Face space for projection (Turk and pentland)

The face which has the minimum distance with the projected face images is labelled with the identity of that image.

The same procedure is established as in section III.A for calculating the minimum distance to find the corresponding face class k that minimizes the Euclidian distance in fisher space.

Face recognition systems using LDA/FLD have also been very successful (Belhumeur et al [9]; Swets and Weng [1]; Zhao et al [7] [11].

Zhao et al [7] [11] describes the LDA approach for face recognition using the class probability: the face image is projected from the original vector space to a face subspace via Principal Component Analysis where the subspace dimension is carefully chosen, the LDA is used to obtain a linear classifier in the subspace. In addition, a weigted Euclidean distance metric is employed to improve the performance of the subspace LDA method.

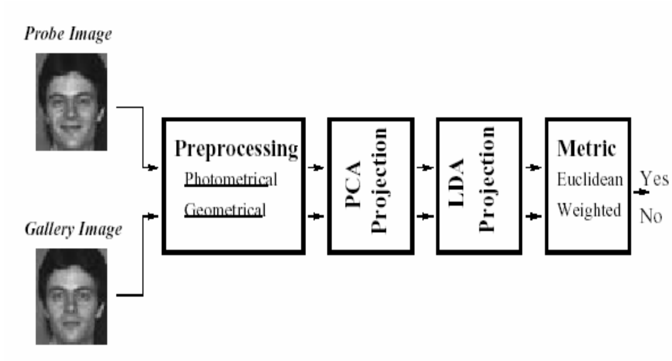


Figure 9 . The subspace LDA face recognition system [7]

Two or four training samples per person are available; LDA training is carried out via scatter matrix analysis [7]. For M class problem, the within and between-class matrices S_w and the S_b are computed as follows:

$$S_w = \sum_{i=1}^{M} \Pr(C_i) \sum_i$$

$$S_w = \sum_{i=1}^{M} \Pr(C_i)(m_i - m_O).(m_i - m_O)^T$$

Where Pr (C_i) is the ***prior class probability*** and usually replaced by (1/M) in practice with the assumption of equal priors.

Here S_W is the within-class matrix showing the average Scatter $\sum_i$ of the sample vector x of different classes C_i around their respective means m_i

$$\sum_i = E\left[(x - m_i)(x - m_i)^T \;\middle|\; C = C_i\right]$$

Zhao et al [7] claim that both weighted and regular LDA perform better than PCA, while weighted LDA is better than regular LDA.

IV. RESULT

A. Data

We used the standard computer vision data set, it contains frontal images of 395 individuals, and each person has 20 frontal images [16]. This data set contains images of people of various racial origins, mainly of first year undergraduate students, so the majority of indivuals are between 18-20 years old but some older individuals are also present. Some individuals are wearing glasses and beards. The total number of images is 7900. In our experiments, three face images are selected for training and reference, and the 17 remaining for testing.

B. Training

To train LDA Algorithm, we used M=30 images of C=10 classes (different persons).each class contains 3 frontal images.

For the ten classes; the images were taken at different times, varying the lighting, facial expressions (open/closed eyes, smiling/not smiling) and facial details (glasses/no glasses). Some exampls are shown in fig 10.

Figure 10. Some examples of training images

In our case there are 30 face images in the face training set of dimension (200*180) pixels.the average face obtained is shown in figure 11.

Figure 11. The total mean image

For the ten classes; the mean image for each class defined by the equation (10) is presented in figure 12.

Figure 12. The ten mean image of each class.

We calculate then the between class scatter matrix Sb and the within class scatter Sw, the eigenvalues of of the matrix ($S_w^{-1}.S_b$) and their corresponding eigenvectors. We choose the k eigenvectors with the highest associated eigenvalues.figure 13 shows the eigenvalues calculated.

The eigenvalues of the $S_w^{-1}.S_b$ shows that we keep only four eigenvalues (respectively eigenvectors) for the LDA subspace.

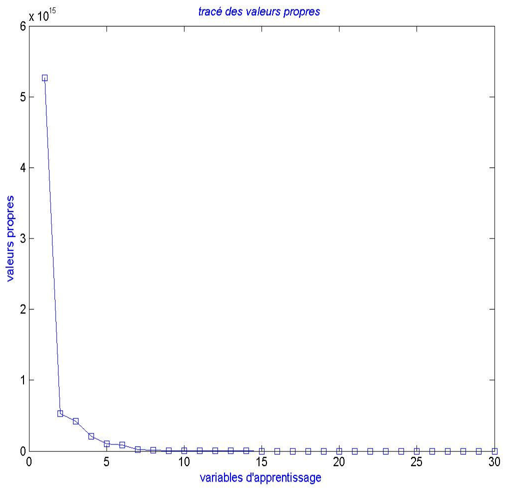


Figure 13. The 30 eigenvalues calculated of the matrix ($S_w^{-1}.S_b$)

We keep all the fisher faces with non-zero eigenvalues in our experiment; we can ignore then those fiherfaces with relatively small eigenvalues. The fisher faces are originally the eigenvectors of the matrix ($S_w^{-1}.S_b$).

After all the subspaces have been derived, all images from training set were projected onto each subspace (fisherface space), recognition using nearest neighbor classification with euclidean distance measures was conducted.

For each individual, we calculate the class vector ΩK by averaging the eigenspace pattern vector Ω calculated from the original image of the individual.

For each new face to be identified, we calculate it's pattern vector Ωi, the distance εk to each known class.

The following figure shows the minimum distance calculated with the 20 test images of each person with all projected traning images.

We noticed that we have a good separation (the minimum distance is lower than a threshold) for the individual L6, L8, L10 and bad separation for the individual L5, L7 and L9 (fig 13).

The discrimination in the fisher space is very clear. We conclude that our Fisherface space is efficient for recognition. From experiments, a recognition rate of 90% was obtained.

Recognition with different head tilts

The robustness of fisherfaces recognition algorithm to head tilt is studied by testing 200 images of our laboratory with different head tilts either left-oriented or right-oriented.

From experiments performed, a poor recognition rate (60%) is obtained with different head tilts and varying illumination.

Various groups have reported various performed results [17] for this algorithm over the years: Zhao et al [7] propose the possibility to combine different subspace features such us a wavelet based decompostion with LDA.

V. CONCLUSION

In this paper, we have presented the face recognition system using LDA. The LDA approach is similar to the eigenface method, which makes use of projection of training image into a subspace. The test images are projected into the same subspace and identified using the euclidean distance.

This method was tested on a number of face images with variations in illumination and orientation. When the number of samples is large and representative for each class, LDA provides better recognition rates than PCA under illumination variance in our experiments.

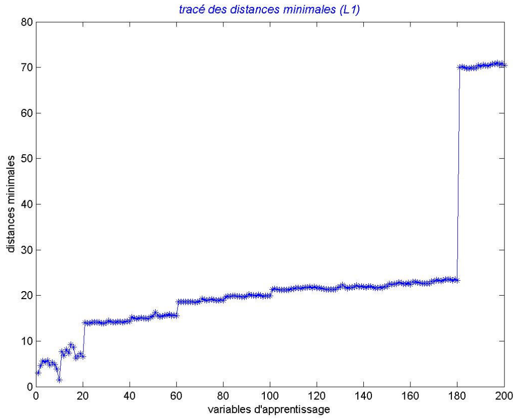
For future work, we propose other appearance based statistical methods like Independent Component Analysis (ICA). Linear Discriminant Analysis (LDA) finds an efficient way to represent the face vector space by exploiting the class information. However, ICA captures both second and higher order statistics and projects the input data onto the basis vectors that are statistically independent as possible.We also proposed a comparative study between appearance methods (PCA-LDA-ICA) in order to improve recognition rate.

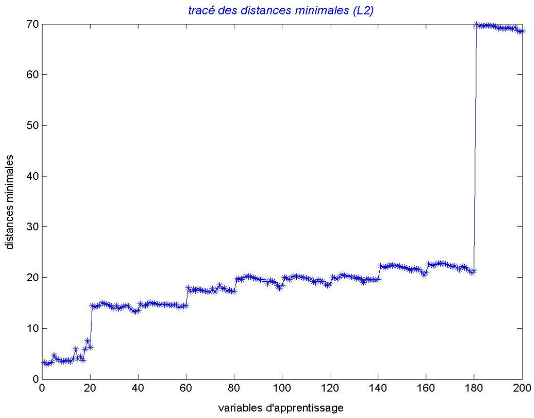
LDA is supervised method, while it is performed to solve a linear solution, we need also to propose a non linear solution like kernel method, the local linear embedding, the Isomap , the laplacianfaces that aims to preserve the local information while the unwanted variation can be eliminated or reduced. Support vector machine using kernel function aims to project the input data onto a high dimension feature space and then constructs an optimal separating hyperplane in that space.

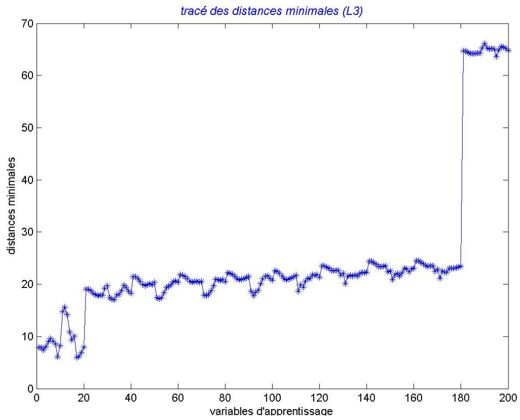
While the human perception system uses both local features and the whole face region to recognize a face, we need to use then hybrid method like the local eigenfeatures (Eigen eyes, eigenmouths ...). We evaluate the performance of systems on different database.

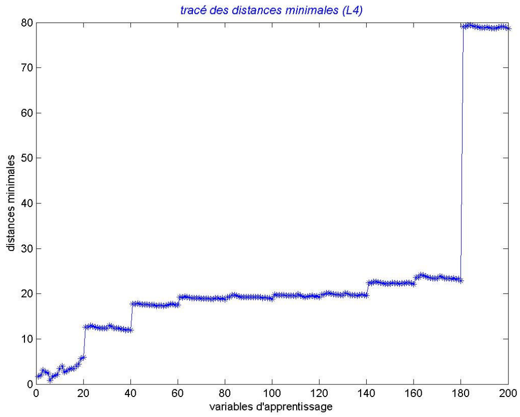
ACKNOWLEDGMENT

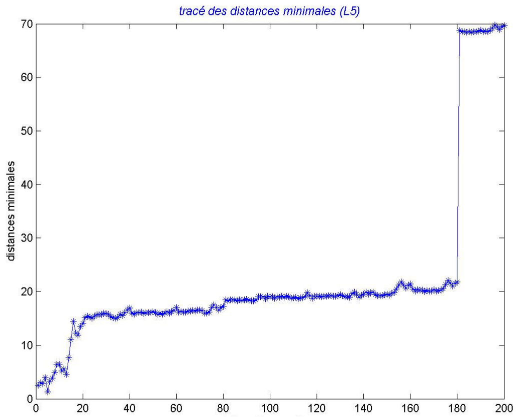
We would like to thank Dr Libor Spacek from computer science department; university of Essex; UK, for providing us the dataset.This data set encouraged comparative research in pattern recognition.

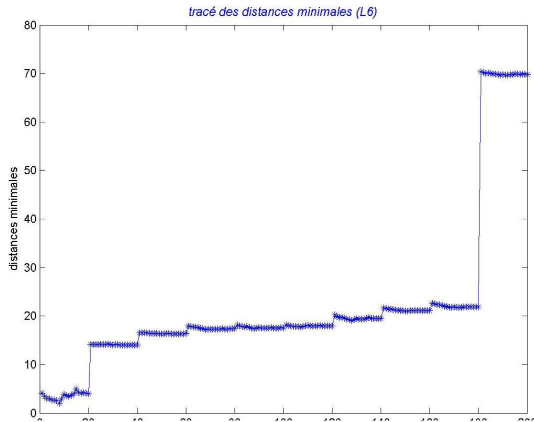


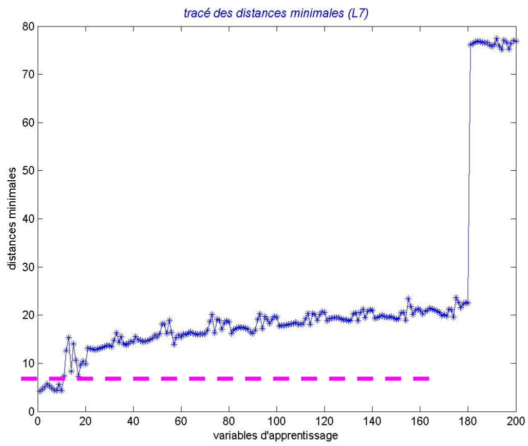


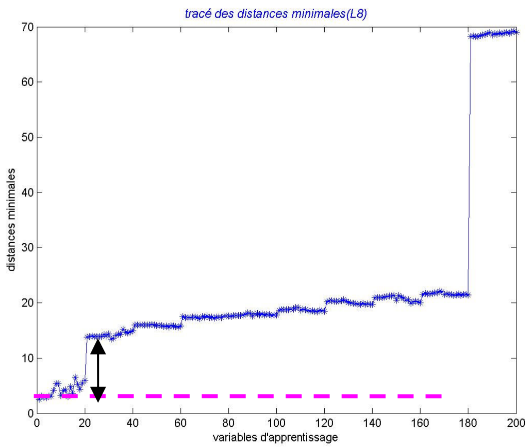












978-1-4244-3757-3/09/$25.00 ©2009 IEEE

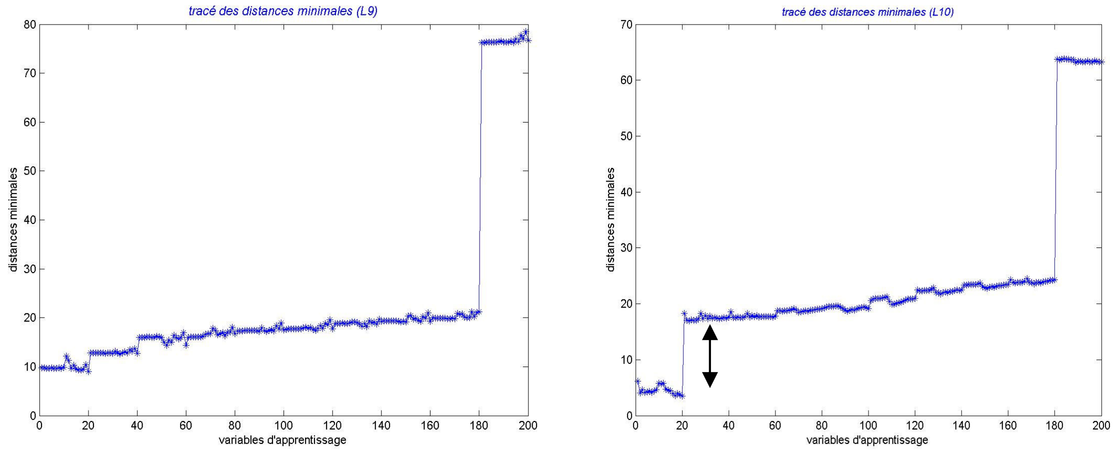


Figure 14. The minimum distance calculated with the 20 test images for the ten classes

REFERENCES

[1] John J. Weng and Daniel L. Swets, “Face Recognition”, Book Biometrics, page 65-86, editeur springer US, ISBN 978-0-387-28539-9, 1999.

[2] Borut Batagelj in Franc Solina, «*Face recognition in different subspaces - A comparative study”,* 6th International Workshop on Pattern Recognition in Information Systems, PRIS 2006 in conjunction with ICEIS 2006, May 23-24, 2006, Paphos, Cyprus.

[3] Turk and A.pentland, “Eigen faces for recognition”,journal of cognitive neuroscience, vol3,N°1,1991.

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

[5] Kyungnam Kim, “face recognition using principal Component Analysis”, courseproject,2001.Available: http://www.cae.wisc.edu/~ece533/project/f03/luopanitchob.doc

[6] Önsen TOYGAR and Adnan ACAN,”Face recognition using PCA, LDA and ICA approaches on colored images”,Journal of electrical and electronics engineering, volume 3, N°1,page 735-743,2003.

[7] W. Zhao,R.Chellappa and P.J.Phillips , “Subspace Linear Discriminant Analysis for Face Recognition”, April 1999. Tech. rep. CAR-TR-914, Center for Automation Research, University of Maryland, College Park, MD.

[8] Kamran Etemad and Rama Chellappa, “Discriminant analysis for recognition of human face images”, journal optical society for America, A/vol .14, N°8. August 1997.

[9] Belhumeur,.P, Hespanha. J, & Kriegman.D, “Eigenfaces vs. Fisherfaces: Recognition using Class Specific Linear Projection”. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 7, (July 1997) 711-720, 0162-8828.

[10] Kresimir delac, Mislav Grgic, Sonja Grgic, “independant comparative Study of PCA, ICA and LDA on the FERET data Set”, 2006 wiley periodicals,Inc. vol15,p252-260.

[11] W.Zhao, R.Chellappa, J.Phillips and A.Rosenfeld , “Face recognition in still and video images:A litterature survey”, ACM comput surv 35(2003),399-458.

[12] Eleyan, A. & Demirel, H. “Face Recognition System based on PCA and Feedforward Neural Networks”, Proceedings of Computational Intelligence and Bioinspired Systems, pp. 935-942, 978-3-540-26208-4, Spain, June 2005, Springer-Verlag, Barcelona.

[13] Xiaogang wang and Xiaoou Tang , “Random Sampling for Subspace Face Recognition”, International Journal of Computer Vision 70(1), 91–104, 2006.

[14] Fatma-Zohra Chelali, Amar Djeradi: “Real time face recognition from video stream using eigen faces”, International Conference on Artificial Intelligence and Pattern Recognition, AIPR-08, Orlando, Florida, USA, July 7-10, 2008. ISRST 2008, ISBN 978-1-60651-000-1,pp 75-80.

[15] Min luo, “Eigen face for recognition” , ECE533 Final project Report, pall’03. Department of biomedical Engineering. Available: http://www.cae.wisc.edu/~ece533/project/f03/luopanitchob.doc

[16] Computer Vision Science Research Projects data set, Available:http://cswww.essex.ac.uk/mv/allfaces/index.html

[17] Kresimir Delac, Mislav Grgic, panos liatsis , “*Appearance based statistical methods for face recognition*” ,47th international symposium EL-MAR 2005, Croatia.

978-1-4244-3757-3/09/$25.00 ©2009 IEEE