Face Recognition using Two-dimensional Subspace Analysis and PNN

BENOUIS MOHAMED
Departement of informatique
&mathematique
University of Oran
Algeria

tlmesani Redwan,Ph.D
INTTIC,Oran

Algeria

Senouci Mohamed,Ph.D
Departement of informatique
&mathematique
University of Oran

ABSTRACT

In this paper, we present an new approach to face recognition based on the combination of feature extraction methods, such as two-dimensional DWT-2DPCA and DWT-2DLDA, with a probabilistic neural networks. This later is used to classify the features matrix extracts for space data created by Two-dimensional Subspace Analysis .The technique 2D-DWT is used to eliminate the illumination ,noise and redundancy of face in order to reduce calculations of the probabilistic neural network operations ,and improve a face recognition system in accuracy and computation time. The proposed approach is tested on ORL and FEI face databases. Experimental results on this databases demonstrated the effectiveness of the proposed approach for face recognition with high accuracy compared with previous methods..

Keywords

biometric, face recognition, 2DPCA, 2DLDA, DWT, PNN,DCT

1. INTRODUCTION

The security of persons is one the major concerns of today's society. Face recognition is one the most commonly used solutions to perform automatic identification of persons. However, automatic face recognition should consider several factors that contribute to the complexity of this task i.e. occultation, changes in lighting, pose, expression and structural components (hair, beard, glasses, etc.).

Several techniques have been proposed in the past to solve the problem of face recognition. Each of them has their strengths and weaknesses, which in most cases depend on the conditions of acquiring information. Recently several efforts and research in this area attributed to increases in the performance in face recognition systems by methods known in the field of pattern recognition such as support vector matching (SVM), Markov hidden model (HMM), methods of probabilities (Bayesian networks) and artificial neural networks[1]. This latter has attracted researchers because its effectiveness in detection and classification shapes which have recently been adopted in face recognition system.

2. The face recognition system

A face recognition system is a system for the identification and verification of individuals, which checks if a person belongs to the database, and identifies whether this is the case.

The methods used in face recognition based on 2D faces are divided into three categories[2]: global methods, local methods and hybrid methods.

- Approaches to facial features, local features, or analytical. This type is to apply transformations in specific locations of the image, most frequently around feature points (corners of the eyes, mouth, nose ...). They therefore require a priori knowledge of the images.

- Global approaches use the entire surface of the face as a source of information without considering the local characteristics such as eyes, mouth, etc.

- Hybrid methods are used to combine the advantages of global and local methods by combining the detection of geometrical characteristics (or structural) with the extraction of characteristics of local appearance.

In this paper we are interested in methods of facial feature extractions in two dimensional images which preserves the original shape of the face in a 2D matrix compared with methods of one dimensional images which represent a face by a vector. This passage matrix to vector losses some geometric and temporal information related to the pixels of the image.

2.1 Principal approach of two-dimensional component analyses (2DPCA)

Proposed by Yang in 2004 [3], 2DPCA is the same method of feature extraction and dimensionality reduction based on Principal Component Analysis (PCA), but deals directly with facial images as matrices without having to turn them into vectors such as the traditional global approach. The authors [4] refer that 2DPCA has high performance in terms of recognition, computation time and reconstruction of the original images compared to Fisher faces, ICA, kernel PCA and eigenfaces. However, this superiority has not been theoretically justified in their article.

**A. The steps of face recognition using 2DPCA*:*

Step training: a training set S of N face images, the idea of this technique is to project a matrix X (n, m) via a linear transformation such that:

$$Y_i = X.R_i \qquad (1)$$

where Y_i vector principal component (n, 1) and R_i is the vector base projection in size (m×1). The optimal vector R_i the projection is obtained by maximizing the total variance criterion generalized

$$J(R) = R^T.G_t.R \qquad (2)$$

where G_t is the covariance matrix (mxm) given by :

$$G_t = \frac{1}{M}\sum_{j=1}^{M}(X_j - \overline{X})^T (X_j - \overline{X}) \qquad (3)$$

with X_j : l the j$^{\text{éme}}$ image of the training set

$\overline{X}$: The average face matrix of the set is:

$$\overline{X} = \frac{1}{M}\sum_{j=1}^{M} \qquad (4)$$

In general one optimal projection axis is not enough. We must select a set of projection axes as:

$$\{R_1, R_2, \ldots, R_d\} = arg\ max\ J(R) \qquad (5)$$

$$R_i^T . R_j = 0, i \neq j, i, j = 1, \ldots, d$$

These axes are the eigenvectors of the covariance matrix corresponding to the 'd' largest eigenvalues. The extraction of characteristics of an image using 2DPCA is:

$$Y_k = X . R_k \quad ; k=1,\ldots d \qquad (6)$$

where $[R_1, R_2, \ldots, R_d]$ is the projection matrix and $[Y_1, Y_2, \ldots, Y_d]$ is the features matrix of the image X.

B) Choice of the number of eigenvectors

The number of eigenvectors associated with the largest eigenvalues to be retained is a big drawback of this technique. To remedy this, several researchers have adopted different solutions to select optimal eigenvalues to remedy this problem. Nguyen recently proposed a new method named random subspace two-dimensional PCA (RS-2DPCA), which combines the 2DPCA method with the technical random subspace (RS) [4]

C) The reconstitution face image using 2DPCA

In the "eigenfaces" approach, the principal components and eigenvalues can be combined to reconstruct the face image in the same manner as in the 2DPCA[5].

In method 2DPCA, a reconstruction of the images from the features is possible. An approximation of the original image with the retained information determined by d is obtained.

$$Y_k = AX_k \quad (k=1,2,\ldots d), \qquad (8)$$

Where V= $[Y_1, Y_2, \ldots, Y_d]$ et U= $[X_1, X_2, \ldots, X_d]$

$$V=AU \qquad (9)$$

Where $X_1, X_2, \ldots, X_d$ are orthonormal. Thus, it is easy to obtain the reconstructed image of A as follows:

$$\hat{A} = VU^T = \sum_{k=1}^{d} Y_k X_k^T \qquad (10)$$

With $\hat{A}_k = Y_k X_k^T$ (k=1,2,…d) represents the reconstructed image of A. According to Yang et al. [4], the energy of the constructed image is focused on a small number of the first component vectors corresponding to the most important eigenvalues .

D) Step classification using Euclidean distance

Classification is performed by comparing the feature matrix of the face library members with the feature matrix of the

input face image. This comparison is based on the K-nearest neighbor classifier (KNN), the distance between two characteristic matrices, $B_i = [Y_1^{(i)}, Y_2^{(i)}, \ldots, Y_d^{(i)}]$ and $B_j = [Y_1^{(j)}, Y_2^{(j)}, \ldots, Y_d^{(j)}]$ is define by :

$$d(B_i, B_j) = \sum_{k=1}^{d} \left\| Y_k^{(i)} - Y_k^{(j)} \right\|_2 \qquad (7)$$

We assume that the training images are $B_1, B_2, \ldots B_N$ (where N is the total number of training images), and each image is assigned to a class P_k .

We take a test image B, Si $d(B, B_l) = min_j\ d(B, B_j)$ and $B_l \in w_k$, compared with a threshold to decide the appropriate class of the test image $(B \in w_k)$

2.2 Approach 2DLDA

In 2004, Li and Yuan [6] proposed a new approach to two-dimensional LDA. The main difference between 2DLDA and the classic LDA is model of data representation. LDA classic is based on the analysis of vectors, while the algorithm 2DLDA is based on the analysis of matrix.

a)Face recognition using 2D LDA

Let X be a vector of the columns n-dimensional unitary. The main idea of this approach is to project the image as a random matrix, $m \times n$, on X by a linear transformation follows:

$$Y_i = A_j X \qquad (8)$$

A_j :Matrix (m x n)

Y: the m-dimensional feature vector of the image A projected.

on suppose L:number of classes

M: the total number of training images

The training image is represented by a matrix $m \times n$ $A_j (j = 1, \ldots M)$

$\overline{A}_i$ (i=1,……L) : the mean of all classes

N_i : number of samples in each class

The optimal vector projection is selected as a matrix with orthonormal columns that maximize the ratio of the determinant of the dispersion matrix of the projected images to the inter-class determinant of the dispersion matrix of the projected images intra-class:

$$J_{FLD}(W_{opt}) = arg\ max_W \frac{|W^T S_b W|}{|W^T S_w W|} \qquad (9)$$

P_b =trace (S_b)

P_W =trace (S_w)

where, S_b: the dispersion matrix inter-class,

S_w : the dispersion matrix intra-class

$$S_b = \sum_{i=1}^{N} N_i\ (\overline{Y}_i - \overline{Y})((\overline{Y}_i - \overline{Y}))^T$$

$$S_b = \sum_{i=1}^{L} N_i\ [(\overline{A}_i - \overline{A})X][(\overline{A}_i - \overline{A})X]^T$$

$$S_W = \sum_{i=1}^{L} \sum_{y_k \in P_i} (\overline{Y}_k - \overline{Y})((\overline{Y}_k - \overline{Y}))^T$$

$$S_W = \sum_{i=1}^{L} \sum_{y_k \in P_i} [(\overline{A}_k - \overline{A})X][(\overline{A}_k - \overline{A})X]^T$$

that means:

$$trace(S_b) = X^T \left[\sum_{i=1}^{L} N_i(\overline{A}_i - \overline{A})^T\ (\overline{A}_i - \overline{A})\right]X$$

$$= X^T S_b X \qquad (10)$$

$$trace(S_W) = X^T \left[\sum_{i=1}^{L} \sum_{y_k \in P_i} (\bar{A}_i - \bar{A})^T (\bar{A}_i - \bar{A}) \right]$$

$$= X^T S_W X \qquad (11)$$

The criterion can be expressed by:

$$J(X) = \frac{X^T S_W X}{X^T S_b X} \qquad (12)$$

where X : column vector unitary.

The unitary vector X maximizing J (X) is called the projection axis. The optimal projection is chosen when X_{OPT} maximizes the criterion, as the following equation:

$$X_{OPT} = argmax_X J(X) \qquad (13)$$

if S_W is invertible, the solution of optimization is to solve the generalized eigenvalue problem.

$$S_b X_{opt} = \lambda S_W X_{opt} \qquad (14)$$

λ is the maximum eigenvalue of $S_W^{-1} S_b$

In general, it is not enough to have only one optimal projection axis. We need to select a set of projection axes,$x_1, x_2, \dots x_d$ with under constraints are the following:

$$\{x_1, x_2, \dots x_d\} = argmax_X J(X)$$

$$X_i^T X_j = 0, i \neq j, i, j = 1, 2, \dots \dots d$$

Indeed, the optimal projection axes, $x_1, x_2, \dots x_d$ are orthonormal eigenvectors of $S_W^{-1} S_b$ corresponding to the largest eigenvalue of the scatter matrix for creating a new projection matrix X, which is a matrix of $n \times d$:

$$X = [x_1, x_2, \dots x_d]$$

A) Feature extraction

We will use the optimal projection vectors 2DLDA $x_1, x_2, \dots x_d$ to extract the image feature, use the following equation:

$$y_k = AX_k, \ k = 1, 2, \dots d \qquad (15)$$

Then we have a family of feature vectors $y_1, y_2, \dots y_d$ which represent by $y = [y_1, y_2, \dots y_d]$ of size $M \times d$ called matrix image characteristics A.

B) Step classification using distance Euclidean

Given two face images A1, A2, represented by the feature matrix 2DLDA, is the same principle as method 2DPCA

3. Discrete wavelet transform (DWT)

Discrete wavelet transform (DWT) is a well-known tool in the field of signal processing. It is widely used in feature extraction, compression and denoising applications. The discrete wavelet transform has been used in various studies of face recognition. The main advantage of the wavelet transform over the Fourier transform is the location in time-scale. Mallat [7] shows that the discrete wavelet transform (DWT-"Discrete Wavelet Transform") may be implemented using a bank of filters including a low-pass filter (PB) and a high-pass filter (PH).

Figure 1. Wavelet decomposition at different levels (a) original image, (b) the wavelet decomposition to a single level (c) 2-level

4. Face classification using PNN

Several studies have shown improved face recognition systems using a neural classification compared to classification based on Euclidean distance measure [8]..

4.1 The neural network

An artificial neural network is a computational model whose design is inspired by the schematic of biological neurons.
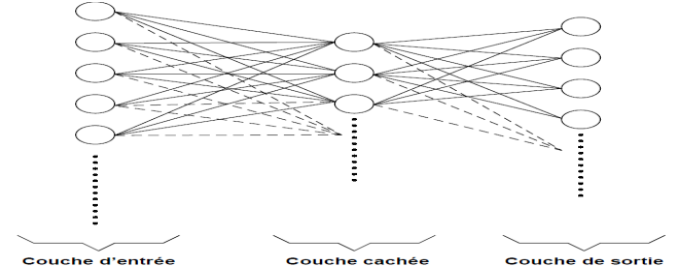

Figure 2 . Schematic representation of a neural network with a single layer standard Hidden.

4.2 probabilistic neural networks

The probability neural network is proposed by D. F. Specht for solving the problem of classification in 1988 [11]. The theoretical foundation is developed based on Bayes decision theory, and implemented in a feed-forward network architecture. The Bayesian decision theory is the principle concept of classification that uses statistical inference, mainly to find the maximum expectation of the smallest risk.

The raw data are divided into *C* classes, which represent the *C* situations to be explained. Each class has *m*-dimensional observations, i.e. $X = [X_1, X_2, \cdots X_m]$. The following Bayes decision rule is used to classify the raw data into *C* classes:

$$h_i c_i P_i(X) > h_j c_j P_j(X) \quad \forall \quad j \neq i,$$

where h_C is the prior probability of class *C*; M_C is the representative center that belongs to the class *C*, but the loss function has been a miscarriage of justice; and P_C is the probability density function of class *C*

The probability density function for each class can be expressed as follow:

$$P_i(x) = \frac{1}{(2\pi)^{\frac{d}{2}} \sigma^d} \frac{1}{N_i} \sum_{j=1}^{N_i} \exp \left[\frac{-(x - x_{ij})^t (x - x_{ij})}{2\sigma^2} \right] \qquad (16)$$

where $P_i(x)$ denotes the dimension of training vector, σ is the smoothing parameter and d is the dimension of training vector. N_i denotes the total number of training vector in category i, and x_{ij} is the neuron vector and x is test vector.

PNN represent mathematically by the flowing expression :

$$a = radbas(\|IW - x\|b) \qquad (17)$$
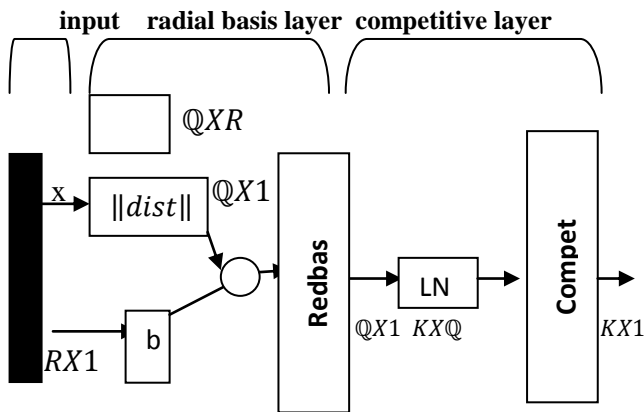
$$y = compet(LW\alpha) \qquad (18)$$

fig 3. Architecture of a probabilistic neural networks[13]

Q:number of input target/pairs :number input layers
R:number of classes of input data: number output layers
IW: input weight
LW: layer weight

   a) The structure   PNN **:** The PNN architecture consists of two layers[12]:
the first layer computes distances from input vector to the input weights (IW) and produces a vector whose elements indicate how close the input is to the IW .
The second layer sums these contributions for each class of inputs to produce as its net output a vector of probabilities .Finally a compete transfer function on the output of the second layer picks the maximum of these probabilities and produces a 1 for that class and a 0 for the other classes. The architecture for this system is shown above.

The probability of neural network with back propagation networks in each hidden unit can approximate any continuous nonlinear function. In this paper, we use the Gaussian function as the activation function:

$$radbas = \exp\left[(-n^2)\right] \qquad (19)$$

Finally, one or many larger values are chosen as the output unit that indicates these data points are in the same class via a competition transfer function from the output of summation unit [11], i.e.

$$compet(n) = e_i = [000\,0_1 0\,.....0_i], n(I) = MAX(n)$$

   In order to implement a face recognition system by our approach, we follow this methodology:

- stage pretreatment using technique DWT
- feature extraction using 2DPCA/2DLDA
- classification using PNN network

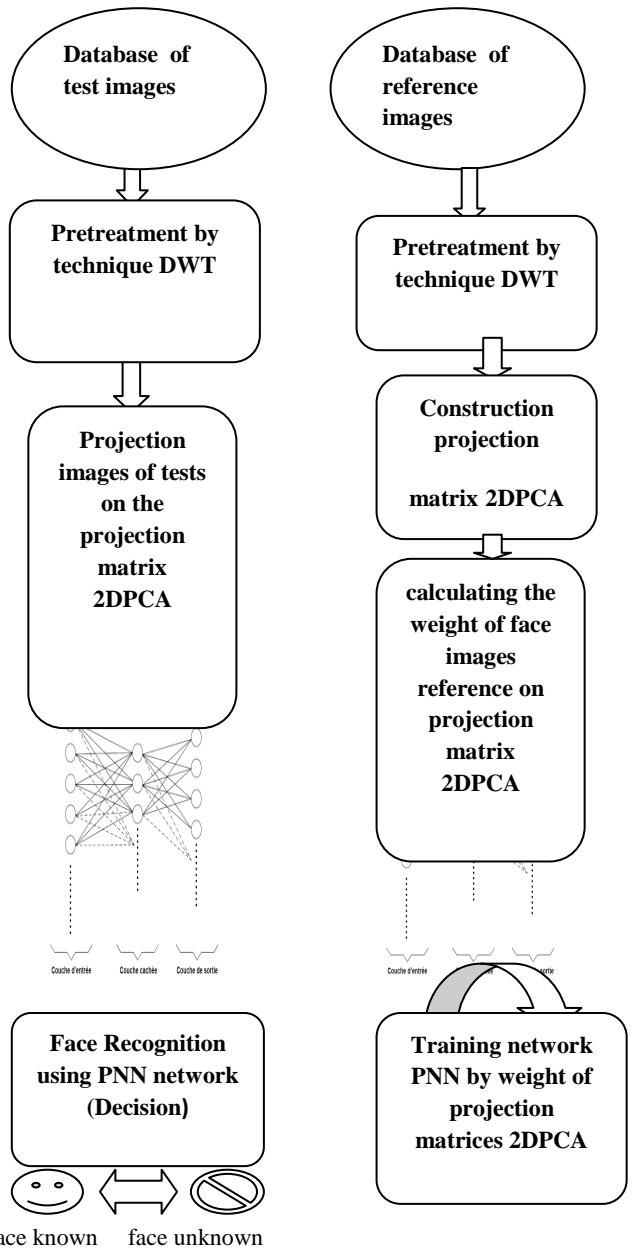Figure 1 shows the block diagram of this approach.

Figure 3. diagram of our approach applied to face recognition

For face recognition task, we applied Two-dimensional Subspace Analysis (2DPCA,2DLDA) which capable of creating the face subspace, defined as orthogonal basis of vectors that contain the most relevant information about a face. These vectors are the eigenvectors of the covariance matrix of the distribution. The recognition process is achieved by forwarding feature matrix which must be transformed into vectors before providing them to the PNN classifier to performed classification and decision.

4. Results and Discussion

In order to evaluate and test our approach described for a face recognition system, we chose three databases: ORL database, FEI and our database constructed by our laboratory. All experiments were performed on MATLAB installed on a laptop with a dual processor T5870 at 2.03 GHz and 2 GB RAM.

A) ORL database

In this database, the faces are made in an environment where the light variation is homogeneous. Different facial expressions may appear without being defined by the constructors of the base. It contains 10 different images for each of 40 people photographed. Inclinations of 20 degrees more, head to the left such as the right, may occur. These images are grayscale with a dimension of 92 * 112 pixels.

Figure 4. our images of the same person from the ORL database

B) FEI database:

Constructed between June 2005 and March 2006, the Artificial Intelligence Laboratory of the FEI in São Bernardo do Campo, São Paulo, Brazil, contains a set of face images, consisting of 14 different images for each of the 200 people photographed, with no head tilt. These images are grayscale with a dimension of 200 * 180 pixels.
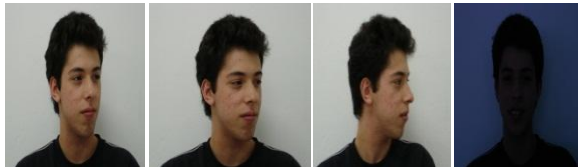

Figure 5. four images of the same person from the FEI database

C) Our database*:* We have face images collected at different times using a capture device (webcam) to form our own database. The database includes 100 face images in a JPG taken on 10 different subjects (N = 10) each registered under 10 different views.

Figure 6. five images from our database

All experiments were performed using the ORL database and FEI with 5 images for training and 5 test images per person for a total of 200 images for each phase (training and test).

A. Experience one

In this experiment we compare the performance of two approaches 2DPCA and PCA in terms of recognition rate and execution time. We use ORL database. For best results, illustrate the approach PCA and 2DPCA, fix the number of dimensions of PCA and 2DPCA, which gives better recognition rates.

After a series of experiments we choose the best values of parameters to define the choice of eigenvectors which gives a better recognition rate.

Table 1. comparison between PCA and 2D-PCA based on ORL

	2DPCA	**PCA**
recognition rate %	96.00	94.00
execution time (s)	13.718743	35.442794

Discussion:

Table (1) shows comparison between PCA and 2DPCA approaches in terms of recognition rate and execution time. We find that the recognition rate obtained by 2DPCA is better compared to that obtained with PCA.
We see that the execution time 2DPCA is less long than PCA. After these figures below, we noticed that the image reconstruction using approach 2D-PCA is better than the PCA whose projection error of 2D-PCA (5.05%) and fable relative to the projection error of PCA (16, 34%).
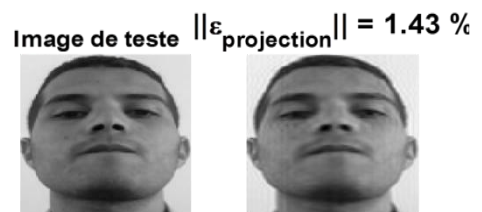

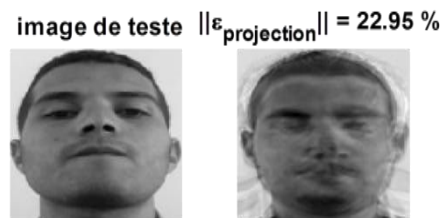Figure 7.The reconstruction of a test image by using 2DPCA

Figure 8. The reconstruction of a test image by using PCA

B. Experience two

In this experiment, we used the ORL database. The classification stage is the phase in which the system assigns a face recognition test face to one class among those of a learning base according to a certain criterion. In our experiment we use the Euclidean distance.
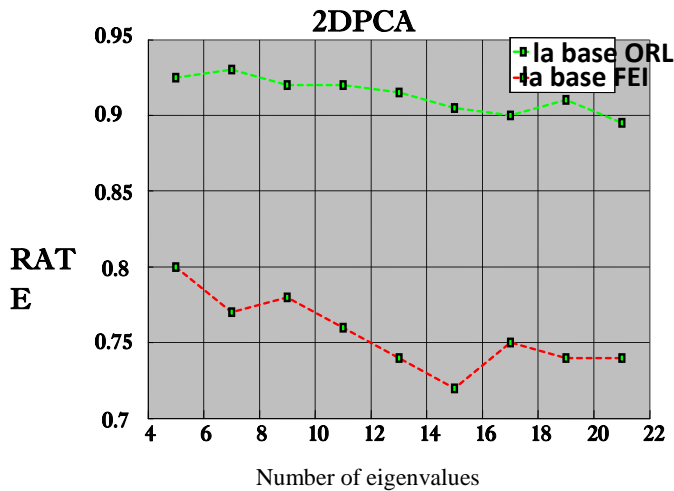
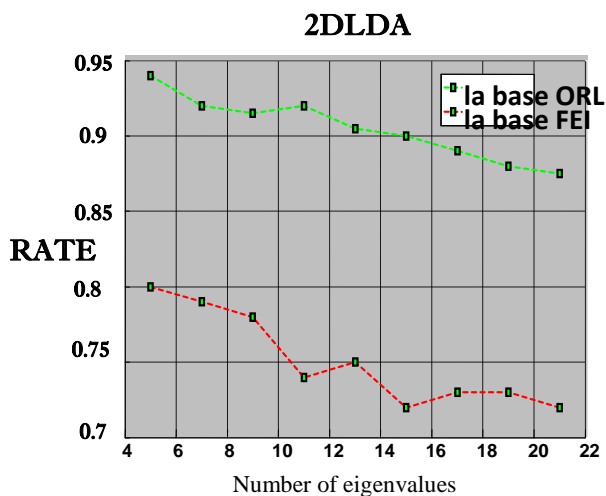

Figure 9. the recognition rate obtained by using 2D-PCA

Figure 10. the recognition rate obtained by using 2D-LDA

Table 2.: Execution time

	Database FEI	Data base ORL
2DPCA(s)	21.685602	13.718743
2DLDA(s)	27.229849	20.558097

Discussion:

Figures (9) and (10) show the evaluation of the recognition rates based on clean faces. After a series of experiments, we observe a better recognition rate based on the seventh clean face (K = 7), which reached 94.4%. for 94.6% and 2DPCA.

✓ From the 11th eigenvector we see a decrease in the recognition rate up to 88.4%.
✓ The decrease in recognition rates for the two approaches is related to the phenomenon of overfitting.

C. Experience three

To evaluate the performance of our proposed approach we have tested using two databases: ORL and FEI.

a) Database: The global performance of algorithms tested on face database FEI are not as good as database ORL. There are two main reasons:

➢ The quality of the image in the database ORL is better than the database FEI.

➢ Database FEI is more difficult due to variations in the details of the face and head orientations.

b) The stage preprocessing: We propose to add a preprocessing stage in order to improve our system in speed by reduced size and eliminating redundant information from the face images by using DWT technique.

c) Feature extraction using 2DPCA/ADLDA: After the reduction of the size of face images using DWT, we used feature extraction approaches 2DPCA and 2DLDA in order to extract the weight images, (features images in new space) which must be converted into vectors before implementing the PNN network.

d) Choice of the number of eigenvalues: Two dimensional methods do not escape this problem and the choice of the appropriate number depends on the method and the database faces used. In our experiments we have selected the best eigenvalues corresponding to the best values of variance (eigenvectors).
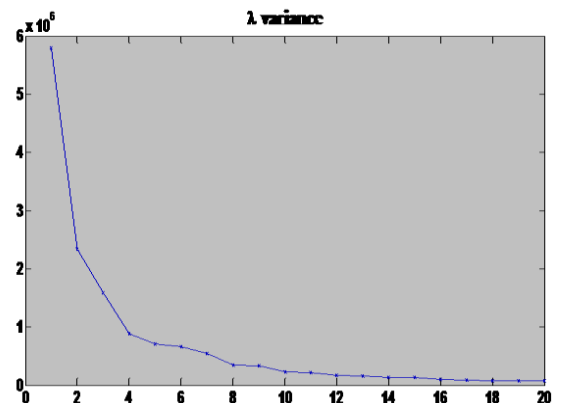

Figure 11. The variance of eigenvalues

e) Selection parameters and architecture PNN : our neural network training algorithm used in system face recognition is not require many parameters compared other neural networks (MLP,BP,....ex),that only parameter that is needed for performance of the network is the smoothing parameter σ .Usually, the researchers need to try different σ in a certain range to obtain one that can reach the optimum accuracy[12].

To get a higher recognition rate, we have made a series of experiments to choose the best smoothing parameter σ used in PNN .

The probabilistic Neural Network used in our system is composed of two layers:

Input layer: The first layer is the input layer and the number of hidden unit is the number of independent variables and receives the input data (number of feature extraction for each approach used in this paper)

output layer: gives the number of faces used in the Database training(ex: ORL 200 person's)

In this experiment we test the performance of face recognition using PNN on three databases: ORL, FEI, and our database.

Table 4. The rate of face recognition by using PNN

	Our data base	**ORL**	**FEI**
recognition rate	90%	%88	%79

Discussion:

Table 4 shows the recognition rates obtained by using PNN on three databases. The performance of this technique decreased up to 79% when tested on FEI and long execution times.

In the last experiment, we compared the different methods described in this paper and evaluated on two databases, ORL and FEI. We followed the same protocol testing previous experiences.
we propose a stage preprocessing based technique 2DDWT to reduce affect illumination ,noise and size the databases and other hand reduce the memory of our neural network training algorithm (PNN)

In this work, it is wanted to test our system with and without added noisy in the two data base in order to evaluate robustness of these approaches namely 2DPCA, 2DLDA, DWT-2DLA, DWT-2DPCA combined by using two classifier :KNN (Euclidean distance)and PNN

Two types of noise are used in this simulation: the Saltand Pepper type noise with a noise density a=0.06 and Gaussian noise with mean m=0, variance v=0.04 and mean m=0, variance v=0.06 .

Figure 12 illustrates these effects which are obtained as follows

(**a**)Salt&pepper Noise (**b**) Gaussian Noise (**c**) Gaussian Noise m=0,v=0.01 m=0, v=0.04

(**a**)Salt&pepper Noise (**b**) Gaussian Noise (**c**) Gaussian Noise m=0,v=0.01 m=0, v=0.04

Fig 12:Adding Noise

Table 5. The recognition rate obtained by different methods on the database ORL

Type of classifier	2DPCA	2DLDA	DWT-2DPCA	DWT-2DLDA
Distance Euclidean	94%	94 .8%	94%	95%
Network PNN	95.8	96%	97%	98%

Table 6. The recognition rate obtained by different methods on the database ORL with added noisy

Type classifier	**2DPCA**	**2DLDA**	**DWT-2DPCA**	**DWT-2DLDA**
Distance Euclidean	90%	91%	90%	92%
Network PNN	92%	92%	94%	95%

Table 7. The recognition rate obtained by different methods on the database FEI

Type of classifier	2DPCA	2DLDA	DWT-2DPCA	DWT-2DLDA
Distance Euclidean	80%	82 .8%	90%	94%
Network PNN	88%	90%	93%	96%

Table 8. The recognition rate obtained by different methods on the database FEI with added noisy

Type of classifier	2DPCA	2DLDA	DWT-2DPCA	DWT-2DLDA
Distance Euclidean	80%	82 .8%	85%	88%
Network PNN	85.8	88%	90%	92%

Discussion:

After these series of experiments we clearly see the superiority of the two-dimensional methods combined with a probabilistic neural classifier combining those of a Euclidean distance classifier.

We also note that the choice of optimal component and the choice smoothing parameter which represents a better recognition rate for both methods, 2DPCA and 2DLDA and accuracy of classification PNN.

The results obtained after using our approach shows that it reduces the computation time of training and improves the recognition rate .

5.CONCLUSION

In this paper, we propose an approach for face recognition based on the combination two approaches, one used for the reduction of space and feature extractions in two dimensions and the other for classification and decision.

After our experience we have found clear improvement of our system, based on DWT-2DPCA and 2DLDA DWT-combined by a probabilistic neural classifier, achieved a better recognition rate compared to other techniques described in this article.

Our choice of using DWT techniques as a preprocessing stage is demonstrated by improved performance of our system in terms of recognition rate ,speed of calculation and reduce memory computation of PNN .

As a perspective, we propose to use this approach in an uncontrolled environment (video surveillance) based on video sequences (dynamic images) in order to make the task of face recognition more robust.

6. REFERENCES

[1] Handbook of Biometrics edited by Anil K. Jain Michigan State University, USA Patrick Flynn University of Notre Dame, USA Arun A. Ross West Virginia University, USA © 2008 Springer Science+Business Media, LLC.

[2] biometric recognition challenges and opportunities Joseph N. Pato and Lynette I. Millett, Editors Whither Biometrics Committee Computer Science and Telecommunications Board Division on Engineering and Physical Sciences Copyright 2010 by the National Academy of Sciences.

[3] D. Zhang and Z. -H. Zhou. (2D) 2PCA: Two-directional two-dimensional PCA for efficient face representation and recognition. Neuro computing, Vol. 69, pp. 224-23 1, 2005.

[4] N. Nguyen, W. Liu and S. Venkatesh. Random Subspace Two-Dimensional PCA for Face Recognition. Department of Computing, Curtin University of Technology, WA 6845,

[5] Two-Dimensional PCA:A New Approach to Appearance-Based Face Representation and Recognition Jian Yang, David Zhang,. 26, NO. 1, JANUARY 2004

[6]. S. Noushath, G.H. Kumar, and P. Shivakumara. (2D)2LDA : An efficient approach for face recognition. Pattern Recognition, 39(7) :1396–1400, 2006.

.[7] B-L. Zhang, H. Zhang, and S.S. Ge. Face recognition by applying wavelet subband representation and kernel associative memory. IEEE Transactions on Neural Networks, 15(1) :166–177, 2004

[8] S. Mallat. A theory of multiresolution signal decomposition : the wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 11(7) :674–693, 1989.

[9]   J.Sima. Introduction to neural networks. Technical Report. N. V-755, August 10, 1998.

[10]. Nazish, 2001. Face recognition using neural networks. Proc.IEEE INMIC 2001, pp: 277-281 (2007)

[11] M. Rizon, M.H. Firdus, P. Saad, S. Yaacob, M.R. Mamat, A.Y.M. Shakkaf, A.R. Saad, Hdesa and M. Karthigayan. Face recognition using Eigenfaces and Neural Networks. American Journal of Applied Sciences, Vol. 2, pp. 1872-1875, 2006..

[12] D. F. Specht, 1990. "Probabilistic neural network and the polynomial adaline as complementary techniques for classification" *IEEE Trans. Neural Networks,* 1(1): 111-121.

[13] neural network toolbox matlab User's Guide COPYRIGHT 1992 - 2002 by The MathWorks, Inc.