Proceedings of the 6th WSEAS Int. Conf. on NEURAL NETWORKS, Lisbon, Portugal, June 16-18, 2005 (pp216-223)

A comparison between linear and nonlinear principal component analysis using neural networks and a novel technique for face recognition

ARYAN SALMANPOUR, SAIED ALI SAIED SALEHI
Biomedical Engineering Department
Amirkabir University of Technology (Tehran Polytechnic)
424, Hafez Ave, Tehran-Iran, P.O.Box:15875-4413
IRAN

Abstract: - In this paper, two new methods of face recognition are presented. First method has two stages for face recognition including feature extraction and their classification. Three feature extraction methods, statistical principal component analysis, linear and non-linear principal component analysis using neural networks and for classification a perceptron neural network with two hidden layers has been used in the first method.
In the second method, many attempts were made for non-linear separation of person's data by their status data. The works were performed by neural networks through some innovative methods. In this model, 100 percentage of recognition rate orrectc for test sample of ORL face were achieved.

Keywords: Face recognition, non-linear principal component analysis, neural networks

1. Introduction

A wide variety of researches on machine face recognition methods have been published. These methods can be classified based on different standards. One of them is application of input images to the system like black and white, gray, color, 2-D infrared images, 3-D images or a combination of them. System can be designed for full, half, three face or a combination of them.
Systems can be classified based on time components. A system can be designed for fixed or time – variable images. They can also be classified based on computational tools , proceeding knowledge , statistical decision – making laws , neural networks , general algorithms and such like can be used in a system .
Many factors contribute in face recognition success. But, one of the most one which have led to some variations in methodology is a manual definition of the characteristics through statistical methods instead of automatic extraction.
Face recognition methods rely on man-defined geometric characteristics. Their values depend on face geometric characteristics including distance and angle of geometric points –e.g. eyes corners, mouth round, nose holes and chin projection. The characteristics defined for half – face include sets of characteristic points on face profile. For example Kaya and Kobayashi [1] used Echlidos distance between points selected on the face manually. Kanadeh [1] used the distance and angle between eye corners, mouth, nose holes and over chin. But the position of these characteristics was determined automatically. In the recent works, a local combination of distance, angle and illumination intensity on each location has been considered. Manual characteristics are recognizable directly, although they have a fundamental problem. Firstly, their automatic extraction is not reliable.
Secondly, the numbers of measurable characteristics are limited. Thirdly, determined values of two characteristics are hardly reliable. Therefore, such a classification method is optimum but leads to an unreliable system. Significant advancement is application of neural network enable to extract the characteristics automatically and indirectly.
Using neural methods, there is no need to define the characteristics for face to recognize, Kohonen [1] presented a method for self- organized mapping for face remember and recognition. Even in high – distributed images or partly – deleted images, the network was able to recognize the face among a few images. Perceptron multi-layer neural and radial function based networks have been employed in face recognition. Post –

propagated error algorithm may be used in training a perceptron multi-layer network in small and limited – variety images. For instance, Rowely [1] used low resolution images to find a face in an image and able to conduct successful tests through a forward multi-layer network and training algorithm after propagation the error with momentum. In large images and high varied effective classes, training methods are more effective and more complex (like statistical methods).

In fact, statistical methods are another form of image illustration. Sirovich and Kirby [1] believe that an image is a vector with many dimensions and each pixel of image are their components. They used Karhanen – loeve method for determining the image of vector in corresponding space to extract the image specifications. Although they did not use of this idea in face recognition, but their idea included presenting the illumination of an image through a linear superposition of principal component vectors. Sentland and Turk [1] used this technique in face recognition. This statistical technique was used in 3-D object image recognition. Through this vector representing, linear distinguishing analysis could be employed in face recognition independently.

It has been proposed that this method can be named as “appearance based technique to distinguish with the others. To separate this method from other neural networks illumination intensity is used directly. It is named “statistical – appearance technique”. Statistical appearance based techniques extract the characteristics from image illumination intensity directly. It is noteworthy that neural networks and statistical techniques do not conflict each other. In fact, a large amount of recent research on neural networks uses a combination of statistical techniques and a network as computational structure. For example, Song and Pujo [1] used a combination of statistical techniques and a perceptron multi-layer network for face recognition. In this paper, new methods of non-linear signal processing in face recognition through artificial neural network structures are presented. In fact, we are looking after simulation of a neural network model which is able to recognize the face appearance variations and also variations due to face view. It requires firstly, extraction an appropriate feature through neural network. Secondly, the employed neural networks should be comparable. In this research, analysis methods of linear and non-linear principal components are evaluated to extract the characteristics of face. Neural networks are used in implementing each of above mentioned methods. The efficiency and effectiveness of aforesaid methods are studied and compared.

Next sections will concern with the structure of first proposed method for face recognition (second section), test results of the first proposed method (third section), second proposed method (forth section) and finally evaluation and conclusion (fifth section).

2- The structure of first proposed method of face recognition

In Fig.1 block diagram of first proposed method of face recognition is presented:

Fig.1 - The structure of first proposed method of face recognition

As it can be seen, firstly, a preprocessing is performed on image data. Then, feature extraction is conducted. This stage is performed through statistical PCA, linear and non- linear PCA neural network. The structures of aforesaid methods are illustrated on the following. Finally, after feature extraction from image data, the patterns are classified. In this stage, extracted features are classified in two different class: training data and test data. Training data are used in training the recognition network which is a forward network with two hidden layer. Then, test data are used to measure the training capability of network. In the following, firstly, a brief summary on face base is presented and different block diagram is described in more details.

2-1- ORL face database

In this paper, ORL [2] face database is used whose face images have been collected in Olivetti research base of Cambridge University from April

1992 to April 1994. There are 400 images from 40 persons. These images include face status (opposite look, left/right look and up/down look) face physiognomy (open/closed eyes, smiling and unsmiling and for some people image with glass and without it). All of the images have a dark background. These images have resolution of 92 * 112 and 256 gray level.

2-2- Pre-processing of face data

Let consider a face image Z_i as a 2-D array M*N of illumination values. You can consider it as a vector MN*1. Value of each pixel of Z_i image is divided to the largest illumination level to normalize the pixel values in 0,1 range. As stated in ORL database, image resolution in this database is 92*112 which is changed to 46*56 to reduce the computational costs.

Now, consider a set of N_T face images $\{Z_K\}_{k=1,2,\ldots,N_T}$. Mean vector of these images can be expressed as:

$$\bar{Z} = \frac{1}{N_T}\sum_{k=1}^{N_T} Z_k \qquad (1)$$

Then, MN*N_R matrix is:

$$Z = \left(Z_1 - \bar{Z}, Z_2 - \bar{Z}, \ldots, Z_{N_T} - \bar{Z}\right) \subset \Re^{MN \times N_T} \qquad (2)$$

In the following, Matrix Z of 2576*400 is used in feature extraction.

2-3- Feature extraction from image data through Statistical PCA method

The first method used in feature extraction from face data is statistical PCA method. The main aim of principal component analysis is converting the mapping of principal components of high dimension space of face to a less dimension one. The most important property of PCA is optimum signal reconstructing based on minimizing the mean square error when just some of the principal components are used for expressing the signal. PCA converting can be considered as follows:

$Y=P^TZ$ that $Z \subset \Re^{MN \times N_T}$ where is face data matrix, $P = [P_1\, P_2 \ldots\, P_n] \subset \Re^{MN \times n}$ including n specific vector of face data covariance matrix and $Y \subset \Re^{n \times N_T}$ includes feature matrix in new space. Matrix Y with less dimension, stores the most important features of principal data. Then, PCA converts face data from MN to n which in n<< MN. We will use of this concise in face data classification.

2-4- Feature extraction from image data through Linear PCA neural network

Linear PCA is the first neural network which is used in feature extracting from face data. Its structure has been shown in Fig.2.

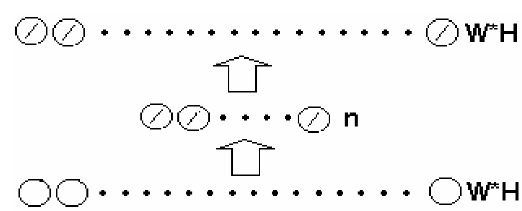

Fig.2. Structure of linear PCA neural network

This network has three layers. Input and output layer have the same number of neurons. As it has been shown in network structure, neuron numbers is W*H which in W and H are width and height of image. Network output has neurons with linear function. The middle layer of this network have n neuron with linear function which in n << W*H. This network acts like a linear filter and compacts the face images in middle layer. Each neuron in hidden layer corresponds a principal component of image. These components are not perpendicular in contrary to principal components extracted through statistical PCA. Values of neurons in hidden middle layer are used in face classification as feature. Back propagation error law is used in training the network.

2-5- Feature extracting from image data through non-linear PCA neural network

Before we describe second network used in this work to extract the feature from image data, we mention the formulation of nonlinear PCA. We define the nonlinear extraction function from data Z onto the feature vector Y as

$$Y = \phi(Z), \qquad \phi \in S_c \qquad (3)$$

And a nonlinear reconstruction function from Y onto the reconstructed vector $\hat{Z}$ as

$$\hat{Z} = \Psi(Y), \qquad \psi \in S_r, \qquad (4)$$

Where S_c and S_r are the sets of nonlinear functions. The data nonlinearly correspond to the principal components through the nonlinear extraction function and the nonlinear reconstruction function.
Our problem is to minimize the mean square reconstruction error:

$$E = E\left[\left\|Z - \hat{Z}\right\|^2\right] = E\left[\left\|Z - \psi(\phi(Z))\right\|^2\right] \qquad (5)$$

When we find the optimal ψ ,and ϕ, we can efficiently describe the data with fewer principal component than that of PCA. In order to obtain the extraction functions $\{\phi_i\}_{i=1,\ldots,m}$ and the reconstruction functions $\{\psi_i\}_{i=1,\ldots,m}$, we use a hierarchical nonlinear principal component network composed of MLPs that are hierarchically arranged.
Second network used in this work to extract the feature from image data was nonlinear PCA neural network with three hidden layers that above we describe the formulation about it. Its structure is shown in Fig.3 This network have three hidden layer with the same number of neurons in input and output layers which is equal to numbers of image pixels. As resolution of images has been changed to 46*56 after change in image size, numbers of neurons in input and output will be 2576. Output layer of this network have neurons with linear function. First and third hidden layer in this network have 256 neurons with sigmoid non-linear function. First hidden layer is responsible for data coding from input to feature space and third hidden layer is responsible for data decoding from feature to output space. The middle hidden layer in this network has n neuron with sigmoid nonlinear function. Face data which have 2576 dimensions in input space, are decreased to n dimensions in this layer which in n<2576.

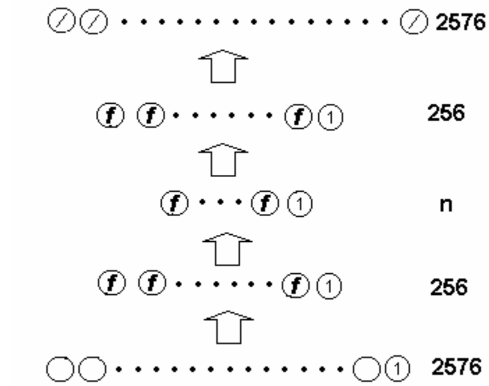

Fig.3. Structure of nonlinear PCA neural network with three hidden layer

Third network used in feature extraction from image data is nonlinear PCA neural network with 5 hidden layers. Its structure is shown in Fig.4. The main difference with Fig.3 structure is adding 2 hidden layers to nonlinear PCA network. While, in Fig. 3 each layer was responsible for coding and decoding of data, in new structure two layers are considered for data coding and decoding. It will be shown that these two hidden layers increase the processing depth and better features are extracted by network.

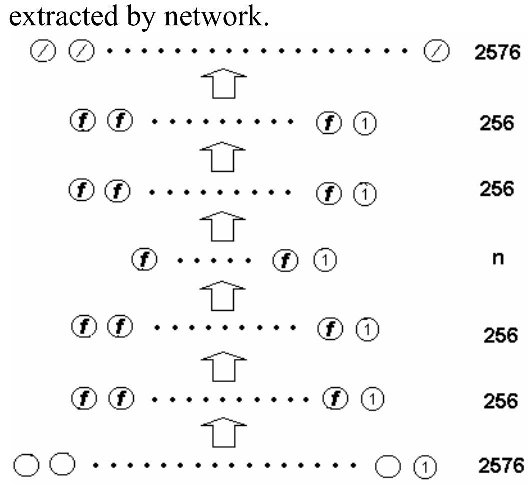

Fig.4. Structure of nonlinear PCA neural network with five hidden layers

2-6- Patterns classification through forward neural network

After feature extracting from face data, a forward neural network with two hidden layer is used for face classification. As it can be seen from Fig. 5, the neural network has n neuron in input layer, where n is dimension of features extracted by one of the aforesaid methods. First and second hidden layer of classifier network have 256 and 128 neurons with sigmoid nonlinear function. Output layer have 40 neurons with sigmoid nonlinear function. As here are forty persons image in database, there is a neuron in output layer corresponding to each person. Corresponding to each person, a neuron in output is active and the others are passive.

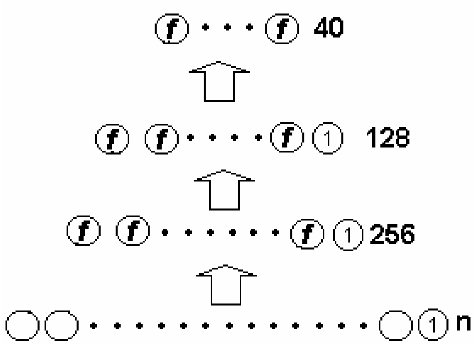

Fig.5. Structure of neural network for classification of patterns

3- Experimental results

Firstly, experimental methods are presented. ORL face database have 400 images from 40 persons. Ten images of each person are available. Firstly, an initial preprocessing is performed on overall image. Then, five images are selected randomly for each person. Five reminder images are considered for test set. Totally, 200 images are specified for training and 200 images are specified for test set. All of 200 images for training set are mounted on one of available feature extraction method. After training the network, input images with 2576 dimensions are compacted in middle layer with dimension of n. Further stage is pattern classification. After training stage, we used testing set for evaluate our methods. Table 1 shows the results of tests. As it can be seen, the most proper recognition is related to nonlinear PCA neural network with five hidden layers and 1024 neuron in middle hidden layer. Typical diagrams of errors in nonlinear PCA neural network with five hidden layers and classifier neural network are presented.

Table 1. Experimental results of first proposed method

Type of feature extracting	Properness of recognition
Statistical PCA and n=80	91.5%
Linear PCA neural network and n=80	82.5%
Nonlinear PCA neural network with 3 hidden layers and n=128	84%
Nonlinear PCA neural network with 3 hidden layers and n=256	86%
Nonlinear PCA neural network with 5 hidden layers and n=32	88%
Nonlinear PCA neural network with 5 hidden layers and n=256	89.5%
Nonlinear PCA neural network with 5 hidden layers and n=512	91.5%
Nonlinear PCA neural network with 5 hidden layers and n=1024	92.5%

Fig.6 illustrates the curves of nonlinear PCA neural network with five hidden layers and 1024 neurons in middle hidden layer against iteration frequency. Fig.7 shows error curve of classifier neural network after training against iteration frequency.

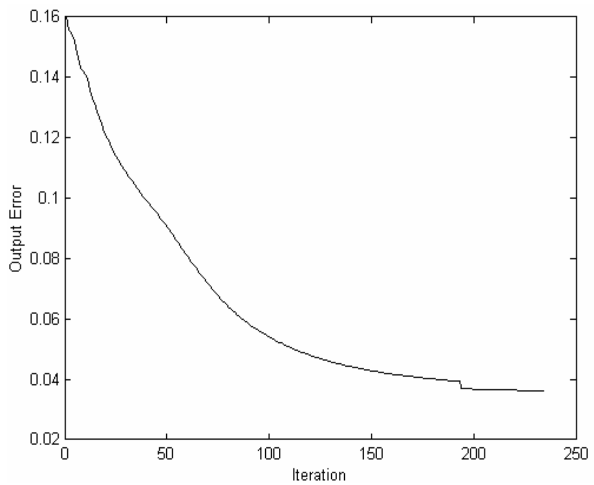

Fig.6. Error curve of nonlinear PCA network with 5 hidden layers and 1024 neurons in middle hidden layer

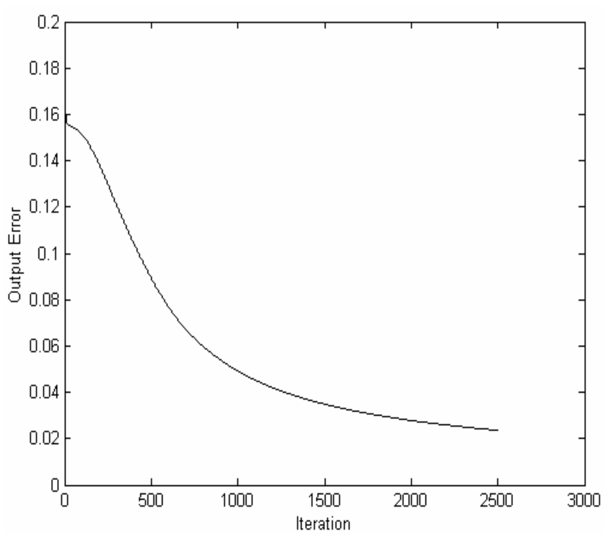

Fig.7. Error curve of classifier neural network after training

4- Second proposed method

In this section, second method applied in this work is presented which in nonlinear separating of person's data from status data is performed. It was conducted through innovative methods by neural networks. Structure of this model has been presented in Fig.8. As it can be seen from Fig.8, new neural network is the same nonlinear PCA neural network with five layers but the middle hidden layer has been divided into two divisions.

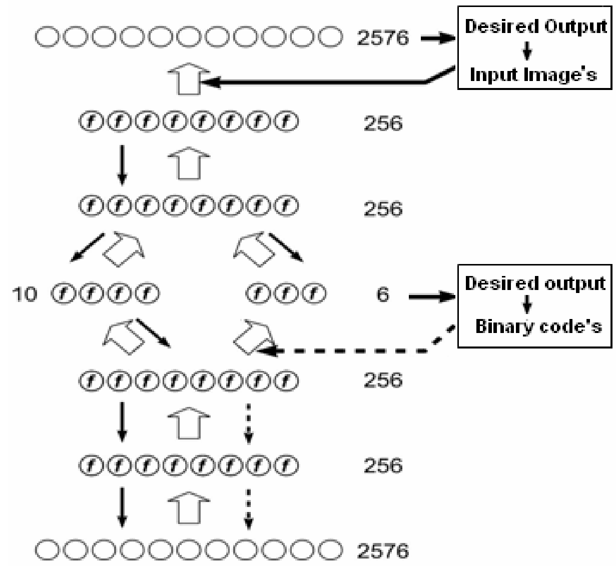

Fig.8 – Structure of new model for face recognition based on nonlinear PCA neural network

First section has six neuron which is responsible for person's separation. Each person have a binary code as there is 40 persons images, at least six neurons are required. After code allocating, relevant codes are used as desired output for training. Second neuron group which are 10 items, it is expected that status information be mounted on nonlinear PCA network. Numbers of network, input and output neurons equals to the image pixel and desired outputs are input images. In order to train the system, back propagation law is used. Training is performed in a way that an error is back propagated from network output to input in order to training the weights. In the middle hidden layer a label- related error is back – propagated to input which adds to error come from output and used for modification of first and second hidden layer. Five images are selected from each person and five remind images are used for test. Totally, 200 images are allocated for training and 200 images for test. Firstly a preprocessing is performed on all of the images. Then 200 training images are used. Fie.9 shows error curve of status recognition neural network. In Fig.10 training and reconstructed images related to second person by this network have been shown.

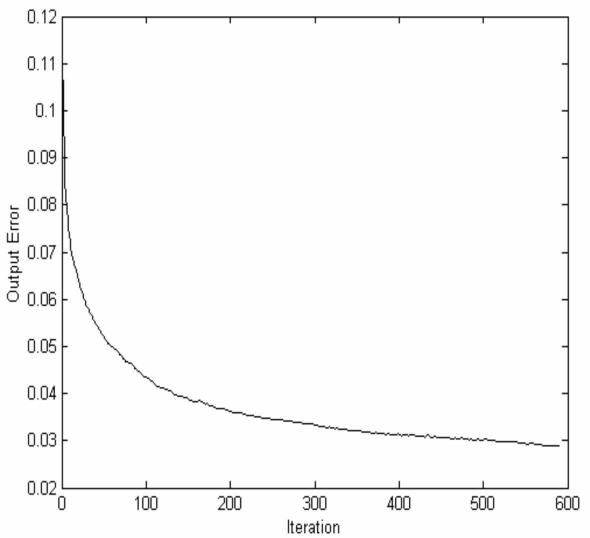

Fig.9 – Error curve of neural network shown in Fig. 8 after complementing the training for the first experiment

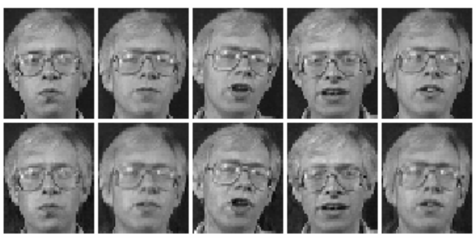

Fig.10- Training (upper) and reconstructed images by network (lower) for the second person in ORL face image database

After training complement 200 image tests were performed on status recognition neural network. Since, test images were not available in training; the network expresses images based on training images. We repeat this experiment for 4 times. Table 2 presents the test results.

Table 2- Result of fitting the test images on Fig.8 neural network.

Nr. of experiment	Dimension of input data	Nr. of test images	Properness of recognition for test samples
1	2576	200	100%
2	2576	200	99%
3	2576	200	99.5%
4	2576	200	100%

Fig.11 presents the test images related to second person along with reconstructing through network. As it can be seen, with this idea in the middle layer persons separating is performed further to compacting and recognition properness has been fully achieved. In Table 3 some face recognition methods available on ORL database have been suggested. As it can be seen in Table 3, our proposed method has the highest recognition percentage among the aforesaid methods.

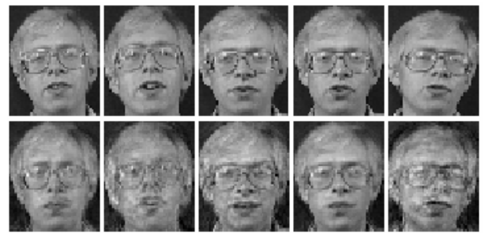

Fig.11. Test images (upper) and reconstructed images by network (bottom) for second person.

Table 3- A comparison between our proposed and the other methods for face recognition properness

Nr.	Methods	ecognition rate orrectCr
1	MLP/BP NN+DWT [7]	93/25%
2	MLP/BP NN+DCT [7]	97/50%
3	MLP/BP NN+PCA [7]	90/25%
4	RBF NN+DWT [7]	93/75%
5	RBF NN+DCT [7]	91%
6	RBF NN+PCA [7]	81%
7	KPCA+GA [8]	98/37%
8	RBF NN+PCA+FLD [9]	98/08%
9	2DPCA[10]	96%
10	ICA-FX[11]	99%
11	FHLA+RBF [12]	98/5%
12	IGF [3]	100%
13	Our proposed method	100%

5- Conclusion

In this paper, analyzing methods of linear and non-linear principal components are used for feature extraction. Neural networks were used in implementing the above mentioned methods. It was proved that using nonlinear principal components, better features can be extracted from face which can lead to higher recognition properness. Finally 100 percent properness in face recognition for test samples of ORL data base was achieved through presenting a model for nonlinear separating of data related to the persons from status information of each person and changing the structure of middle hidden layer of nonlinear PCA in such a way that further to compacting, face classification is performed.

References:

[1] Rama Chellappa, C. l. Wilson, Saad Sirohey ," Human and machine recognition of faces" IEEE Transaction on Neural Network, vol 83, NO 5, MAY 1995
[2] http://www.cam-orl.co.uk
[3] Chengjun Liu, Harry Wechsler, "Independent Component Analysis of Gabor Features for Face Recognition " , IEEE Trans. Neural Network, vol. 14, no.4, July 2003
[4] Marian Stewart Bartlett, Javier R. Movellan , Terrence J. Sejnowski," Face Recognition by Independent Component Analysis " , IEEE Trans. Neural Network, vol. 13, no. 6, November 2002
[5] Pong C. Yuen, J.H. Lai, " Face representation using independent component analysis " , Pattern Recognition 35 (2002) 1247–1257
[6] Ryo Saegusa,Hitoshi Sakno,Shuji Hashimoto "Nonlinear principal component analysis to preserve the order of principal components "Neuro computing , ELSEVIER 2004
[7] M.R.M Rizk , A.Taha "Analysis of Neural Networks for face Recognition systems with feature Extraction to Develop An Eye localization Based Method " IEEE 2002
[8] Zhang Yankun , Liu changing "Face Recognition using Kernel Principal Component Analysis and Genetic Algorithms " IEEE 2002
[9] Meng Joo Er, Shiqian Wu, Juwei Lu, Hock Lye Toh "Face Recognition With Radial Basis Function (RBF) Neural Networks",IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 13, NO. 3, MAY 2002
[10] Jian Yang, David Zhang, "Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition " , IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 26, NO. 1, JANUARY 2004
[11] Nojun kwak ,Chong-Ho choi , Narendra Ahuja "Face Recognition using feature extraction based on Independent Component analysis" , IEEE ICIP 2002
[12] Javad Haddadnia , Karim Faez , Majid Ahmadi "A fuzzy hybrid learning algorithm for radial basis function neural network with application in human face recognition " , PATTERN RECOGNITION ,2003
[13] Juha Karhunen, Erkki Oja, Liuyue Wang, Ricardo Vig´ario, Jyrki Joutsensalo, " A Class of Neural Networks for Independent Component Analysis ", IEEE Trans. Neural Network, vol. 8, no. 3, May 1997