Comparing the Performance of PCA and ICA in Face Recognition

Majid Mahrooghy

***Abstract*—PCA (Principal Component Analysis) can be used in face recognition .given an s-dimensional vector representation of each face in a training set of images ,PCA find a m-dimensional subspace(m<s) whose basis vectors correspond to the maximum variance direction in the original image space. Another technique that has been used in face recognition is ICA (Independent Component Analysis) .ICA attempts to find the basis along which the data projected are statistically independent. Bartlett and Sejnowski provided two architectures of ICA for face recognition[1].**

In this project, it is tried to use the architecture I in order to comparing PCA and ICA in two different using of data set.

***Index Terms*— Face recognition, Principal Component Analysis, Independent Component Analysis**

I. INTRODUCTION

Face recognition is used in many research field such as Human Computer Interaction(HCI),biometric and security. Face recognition is also very important because it is a typical Pattern Recognition problem whose solution can help other classification problem. In order to having good recognition, the feature selection is very important. In fact feature selection in pattern recognition obtains significant features from data to reduce the amount of data for classification and provide strong discriminatory [2].

One way to feature selection is Principal Component Analysis (PCA). This method is a classical feature extraction and data representation technique which is widely used in the area of pattern recognition and computer vision. Turk and Pentland presented the Eigenfaces method (PCA) in face recognition in 1991. Then PCA has been widely examined and has become one of effective methods on face recognition. In this technique the eigenvalue and eigenvectors of covariance matrix of original data set are calculated .the covariance matrix contains values that indicate how much the dimensions vary from the mean with respect to each other. The eigenvectors with highest eigenvalues contain the most information and are the principal components of the data set. By choosing only a few eigenvectors (correspond to highest eigenvalues) and multiplying these vectors with the original data set a new data set is obtained with less dimension, but still with the most important characteristics of the original data. In face recognition by using PCA , The dimensions of the Original feature vectors of the face images are reduced by calculating the eigenvectors of the covariance matrix and choosing the important eigenvectors.

The next step, the Similarities between a new image and the original images from data base are measured. .

One of the advantages of PCA-based algorithm for face recognition is reasonable performance levels and ease of implementing them. It provides an optimal signal representation technique in the mean square error sense. In other word, PCA produce projection axes based on the variation from all training samples.

Another feature selection and dimensionality reduction technique is Independent Component Analysis (ICA).that is one powerful solution to the problem of blind source separation. This technique also was used in face recognition by Batrtlett and Sejnowski.ICA searches for a linear transformation to express random variable as linear combinations of statistically independent source variable [1]. Because in face recognition, most of important information is in the high-order relationships among the image pixels, ICA is a technique that sensitive to high order relationship and works better than PCA in face recognition.PCA considers the 2nd order moment only and it has uncorrelated data, while ICA is suitable for higher order statistics and it identifies the independent source components from the observations. So ICA provides better representation than PCA [2].

`

II. METHODOLOGY

A. *PCA (Principal Component Analysis)*

Let matrix X be the column vectors of the training images, W represent the projection vectors. Then the samples projected are Y=WX .the projection is selected to maximize scatter S which is covariance multiplied by number of samples.

$$W_{opt} = argmax|W^T SW| \qquad (1)$$

W_{opt} is obtained by solution of eigenvector of SW=λW

In fact W_{opt} is a selecting of eigenvector corresponding to the largest eigenvalue of the scatter matrix [3].

B. ICA (Independent Component Analysis)

Some algorithms have been performed for ICA .in this project the informax algorithm which was proposed by Bell and Sejnowski[1,4] is used. Let X be an n-dimensional (n-D) random vector representing a Distribution of inputs. W is an $n \times n$ invertible matrix, U=WX and Y=f(U) an n-D random variable which represent the outputs of n-neurons .each component of f=(f1,…,fn) is an invertible squashing function, mapping real numbers into the [0,1] interval. Usually, the logistic function is used [1]:

$$f_i(u) = \frac{1}{1+e^{-u}} \qquad (2)$$

In Bell and Sejnowski algorithm, the mutual information between X and the output Y must be maximized .this is achieved by performing gradient ascent on the entropy of the output with respect to the weight matrix W. we have:

$$\triangle W \propto \nabla_w H(Y) = (W^T)^{-1} + E(Y'X^T) \qquad (3)$$

Where $Y_i' = {f_i''(U_i)}/{f_i'(U_i)}$ is the ratio between the second and first partial derivatives of the activation function.E is expected value and .H(Y) is the entropy of the random vector Y and $\nabla_w H(Y)$ is the gradient of entropy of Y.by multiplying $W^T W$ in above equation the matrix inverse can be avoided. So we have:

$$\triangle W \propto \nabla_w H(Y) W^T W = (I + Y'U^T)W \quad (4)$$

Where I is the identify matrix. From the logistic transfer function (2) gives $Y_i' = (1 - 2Y_i)$.

So maximizing the joint entropy of Y makes output component to be statistically independent. It can be shown that maximizing the joint entropy of the output Y can minimize the mutual information between the outputs in U. to implement an algorithm that covers above equation , first the row means of X must subtracted and then it must pass through the whitening matrix W_z where:

$$W_z = 2 * \left(Cov(X)\right)^{-(\frac{1}{2})} \qquad (5)$$

This removes the first and second order statistics of the data; both mean and co variances are set to zero and variances are equalized. After input data multiply to W_z then it should multiply to W which is learned by ICA (above equations).so the full transform matrix , W_I , is defined

$$W_I = W W_z \qquad (6)$$

MacKayand [6] showed that ICA algorithm converges to the maximum likelihood estimate of W^{-1}

For $X = W^{-1} S$

Where $S = (s_1, \ldots, s_n)'$ is a vector of independent random variable which is called the sources and has distribution function equal to f_i .that is ,using logistic activation function correspond to assuming logistic random sources. So W^{-1} can be clarified as the source mixing matrix and U=WX can be the maximum-likelihood estimates of the sources generating the data [1].

III. Performing ICA and PCA on data set

In this project, it is tried to use ICA algorithm architecture I from Bartlett & Sejnowski article [1]. In this approach the goal is to find a set of statistically independent basis images .so first, a data matrix X is produced which images are in row and the pixels are in column.

In This simulation, it is used 100 images (180×200) for training. So X has 100 rows and 36000 column .by performing the ICA algorithm W is calculated (according to above equations) we must find a W which the row of U=WX statistically independent as possible. When this is done ,we can use the rows of U as source images and basis images to represent test faces. The dimension can also be reduced by decreasing the number of these bases. If we don't reduce the dimension, the number of bases is 100 because we have 100 training images. This is the same as PCA reduction basis .in the PCA the eigenvectors of covariance matrix are bases .to reduce the dimension of PCA output data, the number of non-important eigenvectors should be decreased.

In this algorithm and also this project, ICA is performed on the data represented on PCA basis. It is done for two reasons.

1) To reduce the number of sources to a tractable number.

2) To provide a convenient method for calculating representation of test images.

By Assuming the P_m^T is the matrix containing the first m PCA axes in its row ,the ICA is performed on P_m^T to create a matrix of m independent source images in the row U. So we have $R_m = X P_m$.

If we want to obtain the original data from data transported, we can have $\hat{X} = R_m P_m^T$ which $\hat{X}$ is approximate of X.

As it was mentioned, ICA algorithm produced $W_I = W W_Z$ such that

$$W_I P_m^T = U$$
$$P_m^T = W_I^{-1} U$$
$$\hat{X} = R_m W_I^{-1} U$$

This means that a test image can be a linear combination of statistically independent source of U. So $R_m W_I^{-1}$ is a matrix that the row of it contains the coefficient for linear combination. We define B=$R_m W_I^{-1}$ as Coefficient matrix.

The coefficient vector for every image test must be found in face recognition.

Next step is to recognize the test images projected.The recognition algorithm performed in this project is the nearest neighbor algorithm (which uses cosines to identify similarity) . In this algorithm we have:

$$c = \frac{b_{test} \cdot b_{train}}{\|b_{test}\| \|b_{train}\|}$$

This equation shows how much similarity is between a test image and a train image. (by using the cosine of the angles between coefficient vectors)

In simulation, two tests have been performed as follow:

1- 100 Individual face images for training and 100 Individual faces images for test .(each case has only one image in training set (figure(1))
2- 100 face but 2 images for each case for training(50 individual face)and 50 different faces for test (figure(2,3))

The images were obtained from university of Sussex database [6].

Fig.1 samples of training data images in TEST1

Fig.2. samples of training data images inTEST2

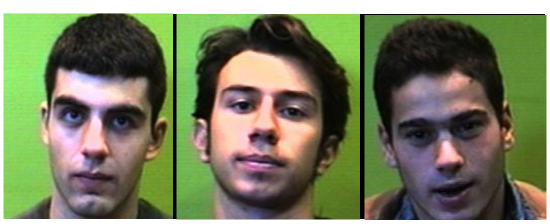

Fig.3. test sample images

Figures 3 and 4 show the result of performing ICA and PCA on TEST1 and TEST2 . As it is seen, ICA has better result than PCA.
Furthermore, the figures show that after 60 features the number of correct recognition doesn't change. It indicates that if the dimension of data projected is reduced to 60, the original data can be obtained. In fact this points out the power full dimension reduction in PCA and ICA technique.

In addition, the figures show that TEST2 has better result than TEST 1. It means if the number of images per individual faces (in training data set) increases the performance also increases.

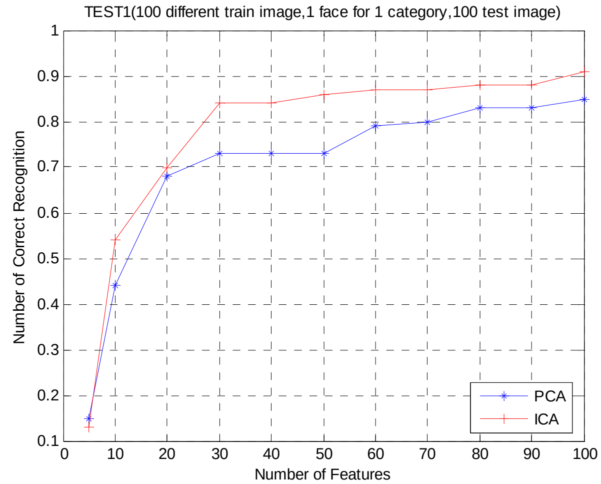

Fig.4 comparing PCA and ICA in TEST1

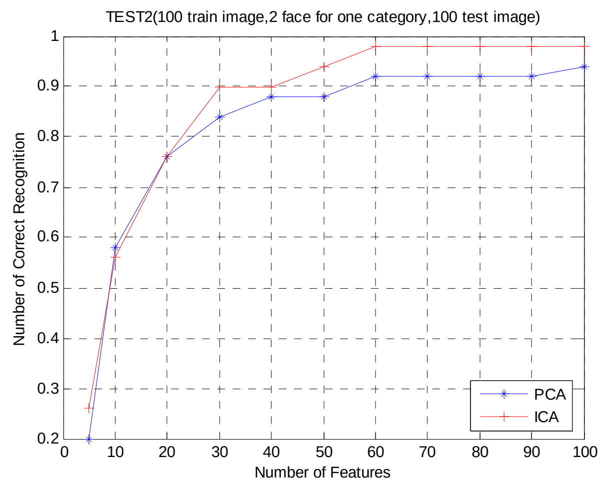

Fig.4 comparing PCA and ICA in TEST2

IV. CONCLUSION

As it was shown, the PCA and ICA can be used in face recognition. The goal of PCA is to find a subspace (with fewer dimensions) whose basis vectors correspond the maximum variance direction in original image space. ICA also tries to find the basis along which the data projected are statistically independent. it was shown that ICA has better performance than PCA in the our test(TEST1,TEST2) .and also it was seen that by reducing the number of representation basis from 100 to 60 ,the number of correct recognition doesn't change for PCA and ICA .It indicates that ICA and PCA have considerable ability to reduce feature.

REFERENCES

[1] .S. Bartlett, J.R. Movellan, *T.J. Sejnowski, Face Recognition by Independent Component Analysis, IEEE Trans. on Neural Netw*

[2] Liu, H. Wechsler, *Comparative Assessment of Independent Component Analysis (ICA) for Face Recognition, Proc. of the Second International Conference on Audio- and Video-based Biometric Person Authentication, AVBPA'99, 22-24 March 1999, Washington D.C., USA, pp. 211-216 orks, Vol. 13, No. 6, November 2002, pp. 1450-1464*

[3] P.H. Belhumeur, J.P. Hespanha, and D. J. Kriegmand, *"Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection", IEEE Transactions on Pattern Analysis and Machine Intelligence,vol. 19, pp. 711-720, July 1997*

[4] A. J. Bell and T. J. Sejnowski, *"An information-maximization approach to blind separation and blind deconvolution," Neural Comput., vol. 7,no. 6, pp. 1129–1159, 1995*

[5] D. J. C. MacKay, *Maximum Likelihood and Covariant Algorithms for Independent Component Analysis: , 1996.*

[6] *http://cswww.essex.ac.uk/mv/allfaces/faces96.html*