

# An Experimental Comparison of Unsupervised Learning Techniques for Face Recognition

Dinesh Kumar, C.S. Rai, and Shakti Kumar

**Abstract**—Face Recognition has always been a fascinating research area. It has drawn the attention of many researchers because of its various potential applications such as security systems, entertainment, criminal identification etc. Many supervised and unsupervised learning techniques have been reported so far. Principal Component Analysis (PCA), Self Organizing Maps (SOM) and Independent Component Analysis (ICA) are the three techniques among many others as proposed by different researchers for Face Recognition, known as the unsupervised techniques. This paper proposes integration of the two techniques, SOM and PCA, for dimensionality reduction and feature selection. Simulation results show that, though, the individual techniques SOM and PCA itself give excellent performance but the combination of these two can also be utilized for face recognition. Experimental results also indicate that for the given face database and the classifier used, SOM performs better as compared to other unsupervised learning techniques. A comparison of two proposed methodologies of SOM, Local and Global processing, shows the superiority of the later but at the cost of more computational time.

**Keywords**—Face Recognition, Principal Component Analysis, Self Organizing Maps, Independent Component Analysis

## I. INTRODUCTION

**H**UMAN Face Recognition has always been an area of intensive research. It is a challenging biometric technique of identifying individuals by facial features. People in computer vision and pattern recognition have been working on automatic face recognition of human faces for the last two decades and it is attracting much more attention. The main reason for the attraction of many researchers towards this has been the variety of practical applications. It is an ideal application for security, such as, to limit employee access to sensitive data in private companies, to limit the physicians to have an access to their patient records in hospitals and the others like airport security, criminal identification, video surveillance etc. There are four basic methods for face recognition: Appearance based, Rule based, Feature based and Texture based methods. A face recognition system compares the current image with the images in the database. Neural Networks make use of new face image and the stored faces to

determine if there is a match. Various researchers have proposed a large number of supervised and unsupervised techniques for face recognition. Turk and Pentland [5] made significant contribution towards the machine recognition of faces. The method used for face recognition used Principal Component Analysis (PCA) for dimensionality reduction and was known as the Eigenface method. This approach was based on second order statistics. Linear Discriminant Analysis (LDA), also known as Fisherfaces, a supervised learning algorithm, was proposed by [1, 2] and it was claimed that this method is insensitive to large variations in lighting and facial expressions. It is generally believed that algorithms based on LDA are superior to those based upon PCA. However, from some recent work [4] it was concluded that when the training data set is small, PCA can outperform LDA and also PCA is less sensitive to different training data set.

The above two methods (PCA and LDA) aim to preserve the global structure where as in many real world applications, the local structure is more important. In order to preserve the intrinsic geometry of the data and the local structure, a new method was proposed by He et. al. [29] in which Locality Preserving Projections (LPP) were used for mapping the face images into the face subspace. The results showed that this algorithm is especially suitable for frontal face images. Several other methods such as Probabilistic Subspaces [12, 13, 16, 17], Feature Line Method [27], Evolutionary Pursuit [10], Support Vector Machines (SVM) [28] etc. have also been proposed by various researchers with their relative advantages and disadvantages. A large number of face recognition algorithms reported in literature used PCA that was based on second order statistics. Bartlett and Sejnowski [11, 14, 21] introduced method that considered the higher order statistics also. The method was based upon Independent Component Analysis (ICA).

Self-Organizing maps (SOMs) [30, 31] have also been successfully used as a way of dimensionality reduction and feature selection for face space representations [18, 19, 20]. This paper proposes integration of the two techniques, SOM and PCA, for dimensionality reduction and feature extraction. It also makes an attempt to compare two methodologies namely global processing and local processing of face image using self organizing maps and compare the performance of the two in terms of the recognition rate of face recognition system using the face database with subjects having variation in facial expressions and facial details. A comparison of SOM, an unsupervised learning algorithm with the popular and successful classical method PCA and the proposed method has been given in terms of recognition rate of face recognition

Dinesh Kumar is with the Department of Computer Science & Engineering, G. J. University of Science & Technology, Hisar, India. (Phone: +91-9896065440, Fax: +91-1662-276240 (off) email: dinesh\_chutani@yahoo.com).

Dr. C S Rai is Reader in University School of Information Technology, GGS Indraprastha University, Delhi, India. (e-mail: csrai\_ipu@yahoo.com).

Dr Shakti Kumar is Director of Institute of Science and Technology, Klawad, Yamuna Nagar, India (e-mail: shakti\_k@yahoo.com).

system. Finally it also compares the performance of the three unsupervised learning techniques SOM, PCA and ICA.

After a brief discussion of SOM, PCA and ICA in section 2, Section 3 describes the proposed method of combining SOM and PCA. The experiments are reported in section 4 and section 5 contains the conclusions.

## II. SOM, PCA AND ICA

### A. Self Organizing Maps

Self Organizing Map is a neural network model of the unsupervised class. It consists of two layers of neurons: an input layer and a so-called competition layer. The weights of the connections of the input neurons to a single neuron in the competition layer are interpreted as a reference vector in the input space. A self-organizing map is trained using competition learning. There is a competition among the neurons to be activated or fired. When an input pattern is presented to the network, the neuron in the competition layer is determined; the reference vector of which is closet to the input pattern. The neuron is called the winner-takes-all neuron. Its weights are changed. The changes are made in such a way that the reference vector represented by these weights is moved closer to the input pattern. The weights of the neighbouring neurons are also changed. The algorithm is summarized as follows [30].

Assume that the input vector  $\mathbf{X}^i = (x_1^i, x_2^i, \dots, x_n^i)$  drawn from the input space with a certain probability is presented to say  $r \times r$  field of neurons with weight vector  $\mathbf{W}_{jk}^i = (w_{1,jk}^i, w_{2,jk}^i, \dots, w_{n,jk}^i)$  at time instant  $i$ , applied to  $jk$ th neuron where  $j, k = 1, \dots, r$ . Initially we choose the random values for initial weight vector  $\mathbf{W}_{jk}^0$ , the value of neighbourhood around the winning neuron as  $h_{JK}^i$  and the learning rate as  $\eta^0$ . After initialization,

we pick a sample  $\mathbf{X}^i$ . Then similarity matching is done. In order to find the best matching (winning) neuron  $JK$ , we use minimum Euclidean distance criterion and the winning neuron is the one that minimizes the distance  $\|\mathbf{x}^i - \mathbf{w}_{JK}^i\| = \min_{jk} \{\|\mathbf{x}^i - \mathbf{w}_{jk}^i\|\}$  where  $\mathbf{W}_{JK}^i$  is the best matching weight vector. After this the synaptic vectors of only the winning cluster are updated using the update formula

$$w_{l,jk}^{i+1} = w_{l,jk}^i + \eta^i (x_l^i - w_{l,jk}^i) \quad jk \in h_{JK}^i$$

and  $\eta^i$  and  $h_{JK}^i$  are also updated. The above steps (after initialization) are repeated until no noticeable changes in the feature map are observed. At the start of the algorithm,  $h_{JK}^i$  usually includes all neurons in the vector field and gradually its value reduces. During the ordering phase  $h_{JK}^i$  shrinks linearly with  $i$  to finally include only a few neurons and during the convergence phase it may have only one or no

neighbours. In general the learning rate is close to unity in the beginning during the initial period (ordering phase) and then it is decreased either linearly, or exponentially or inversely with index  $i$  while maintaining it above 0.1 and during the convergence phase it has a very small value say 0.01 but it is never zero. Figure 1 shows the flow chart.

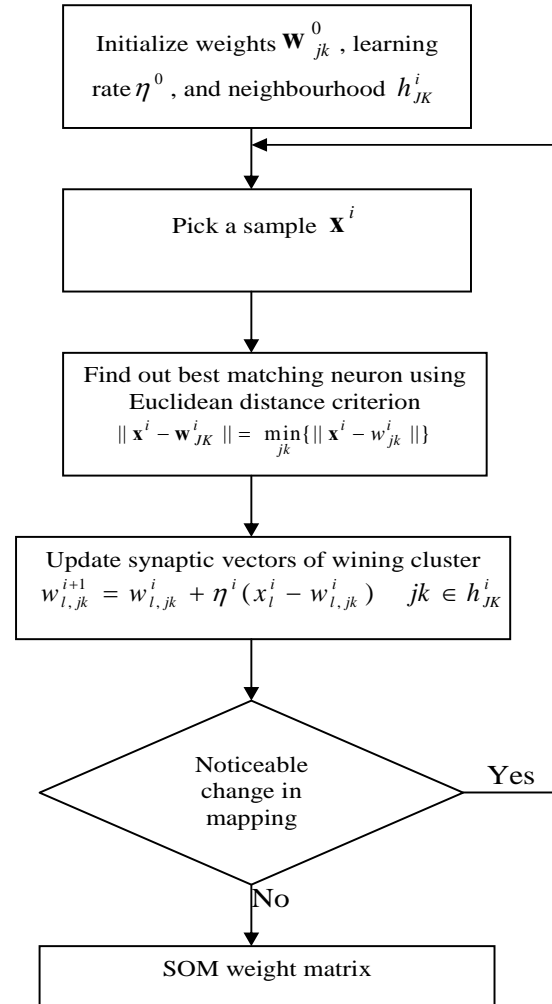


Fig. 1 Flowchart for SOM algorithm

### B. Global and Local Processing

The global processing is the one in which each and every pixel of the face image is fed into the self organizing map networks whereas in the local processing method, the face image is divided into blocks and these blocks of pixels are processed. Global processing requires substantially larger network as compared to that required for local processing technique. This is due to the fact that the usage of pixel blocks effectively results in reduction of dimensionality of data space that has to be topologically represented in the SOM space. The training images in the SOM method are mapped to lower dimension using SOM and the weight matrix of each training image is stored. At the time of recognition, the training images are reconstructed using the weight matrices and matching is

done with the test image using Euclidean norm (L2 norm) as the similarity measure.

In SOM method, the training images are mapped to lower dimension using Self Organizing Maps. It follows the steps as under.

- Obtain a single vector of pixel values by concatenating the face image for global processing. In case of local processing, obtain vectors by concatenating the pixels values contained in small blocks in which the face image was divided and finally a matrix with as many columns as the number of blocks and each column containing the number of pixels equal to the number contained by one small block.
- Choose a two dimensional self-organizing map of suitable size that represents the lower dimensions.
- Train the self-organizing map as per algorithm described in section 2.1. The input to the self-organizing map is the vector/ matrix as obtained in step 1 for global/local processing respectively.
- Retain the weight matrix as obtained after training corresponding to each training image.
- Reconstruct the images using the retained weight matrices at the time of matching.
- Match with the test images using Euclidean norm.

### C. Principal Component Analysis

Principal Component Analysis is a commonly used unsupervised statistical method for data analysis. It is widely used in signal processing and neural computing. The basic goal in PCA is to reduce the dimension of the data. It rotates the data such that maximum variabilities are projected onto the axes. It transforms a number of correlated variables into a smaller number of uncorrelated variables called the principal components, which are ordered by reducing variability. The uncorrelated variables are linear combinations of the original variables, and the last of these variables can be removed with minimum loss of real data. It has been successfully used for face recognition [4,5]. An image was treated as a vector in a very high dimensional space. Only the best eigenfaces (eigenvectors of the covariance matrix of a set of images) those that had the largest eigenvalues were used to approximate the face. Consider a set of  $N$  sample images  $\{\Gamma_1, \Gamma_2, \dots, \Gamma_N\}$  taking values in an  $n$ -dimensional image space. Let us also consider a linear transformation mapping the original  $n$ -dimensional image space to  $m$ -dimensional feature space where  $m < n$ . The new feature vectors  $Y_k$  are defined by the following linear transformation  $Y_k = \Phi^T \Gamma_k$  where  $\Phi$  is a matrix with orthonormal columns. There is a linear transformation mapping of original  $n$ -dimensional image space to  $m$ -dimensional feature space where  $m < n$ . The covariance matrix is defined as

$$C = \sum_{k=1}^N (\Gamma_k - \Psi)(\Gamma_k - \Psi)^T$$

where  $\Psi$  is the mean image of all the samples. Only  $m$  number of  $n$ -dimensional eigenvectors  $[V_1, V_2, \dots, V_m]$  of  $C$  is chosen that correspond to the  $m$  largest eigenvalues.

PCA method begins with the vertical concatenation of pixels of each row of the training images taken one at a time, vertically to form a single vector containing all the pixel values of an image thereby producing a matrix. It follows the following steps.

- Compute a matrix  $X$  containing the training images with mean subtraction from the original matrix, each column of which represents an image.
- Find the eigenvectors and eigenvalues from the covariance matrix  $X^T X$ .
- Sort the eigenvalues and eigenvectors by decreasing eigenvalues.
- Select the number of eigenvalues and hence the number of eigenvectors covering the maximum variance (Table 1).
- Retain the eigenvectors & KL coefficients.
- Compute KL coefficients for test images using retained eigenvectors.
- Match the KL coefficients for training & test images using Euclidean norm.

The Energy preservation factor, EPF, was computed by retaining only  $n$  number of eigenvalues for total of 200 images (200 eigenvalues).

$$EPF = \frac{\sum_{i=1}^n \lambda_i}{\sum_{i=1}^M \lambda_i} \times 100$$

$M$  is the total number of eigenvalues. The following table gives the Energy preservation factor for various values of  $n$

TABLE I  
 ENERGY PRESERVATION FACTOR FOR DIFFERENT N VALUES

No of Features Eigenvalues (n)	199	160	120	100	80	40	20
Energy Preservation Factor (EPF)	100	98.67	96.06	94.13	91.53	81.72	69.82

### D. Independent Component Analysis

Independent Component Analysis (ICA) was initially developed to provide solution to a problem known as Blind Source Separation (BSS). It is a method of separating out independent sources from linearly mixed data. Let  $U$  be the source vector and  $A$  is the mixing matrix, and then the observation vector  $X$  is given as

$$X = A U \quad (1)$$

where  $U$  and  $A$  both are unknown and  $X$ , an observation vector, is the only thing available to us. So, if  $X$  is known to us, we need to find the demixing matrix  $w$  such that the original source vector  $U$  can be recovered from the output vector  $Y$ , defined as

$$\begin{aligned} \mathbf{Y} &= \mathbf{W}\mathbf{X} \\ &= \mathbf{W}\mathbf{A}\mathbf{U} \end{aligned} \quad (2)$$

From the above relation it is clear that  $\mathbf{Y}$  becomes equal to  $\mathbf{U}$  when  $\mathbf{W} = \mathbf{A}^{-1}$ . A survey on Independent Component Analysis is given in [25]. We have vast variety of applications where ICA is being used. Some of the applications have been reviewed in [26]. Face Recognition is also one of the application areas where ICA is used. There are a large number of algorithms used for face recognition where the face representations found by unsupervised methods were based on the second order statistics only such as Principal Component Analysis. In an application like face recognition, it was emphasized that the higher order statistics also contain important information. ICA is one of the methods that deal with higher order statistics. Bartlett et. al. [11, 14, 15, 21] used a version of ICA derived from the principle of optimal information transfer through sigmoidal neurons. They proposed two architectures of images on which ICA was performed. The images were treated as random variables and the pixels as the outcomes in Architecture I whereas in Architecture II, the pixels were treated as random variables and the images as the outcomes. The Infomax algorithm as proposed by Bell and Sejnowski [24] was used for performing ICA. The entropy of the random vector  $\mathbf{Z}$  at the output of the nonlinearity  $\mathbf{G}$  is

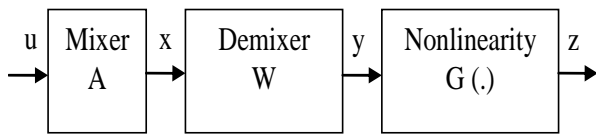


Fig. 2 Block diagram of maximum entropy method for ICA

$$\begin{aligned} h(\mathbf{Z}) &= -E[\log f_z(\mathbf{z})] \\ &= -E\left[\log\left(\frac{f_U(\mathbf{u})}{|\det(\mathbf{J}(\mathbf{u}))|}\right)\right] \end{aligned} \quad (3)$$

where  $\det(\mathbf{J}(\mathbf{u}))$  is the determinant of the Jacobian matrix  $\mathbf{J}(\mathbf{u})$  and using the chain rule of calculus, we may write

$$|\det(\mathbf{J})| = |\det(\mathbf{W}\mathbf{A})| \prod_{i=1}^m \frac{\partial z_i}{\partial y_i} \quad (4)$$

The maximization of entropy  $h(\mathbf{z})$  requires the maximization of expectation of the denominator term in Eq. (3) that is  $\log |\det(\mathbf{J}(\mathbf{u}))|$  with respect to the weight matrix  $\mathbf{W}$ . So we may consider the objective function as

$$\Phi = \log |\det(\mathbf{J})| \quad (5)$$

Putting Eq. (4) into (5) yields

$$\Phi = \log |\det(\mathbf{A})| + \log |\det(\mathbf{W})| + \sum_{i=1}^m \log \left( \frac{\partial z_i}{\partial y_i} \right) \quad (6)$$

Differentiating  $\Phi$  with respect to the weight matrix  $\mathbf{W}$  gives

$$\frac{\partial \Phi}{\partial \mathbf{W}} = \mathbf{W}^{-T} + \sum_{i=1}^m \frac{\partial}{\partial W} \log \left( \frac{\partial z_i}{\partial y_i} \right) \quad (7)$$

The nonlinearity used was the logistic function given as

$$\begin{aligned} z_i &= g(y_i) \\ z_i &= \frac{1}{1 + e^{-y_i}} \quad i = 1, 2, \dots, m \end{aligned} \quad (8)$$

Substituting Eq. (8) into (7), we get

$$\frac{\partial \Phi}{\partial \mathbf{W}} = \mathbf{W}^{-T} + (\mathbf{1} - \mathbf{2z})\mathbf{x}^T \quad (9)$$

The objective of learning algorithm is to maximize the entropy  $h(\mathbf{z})$ . Using the method of steepest ascent, the change applied to the weight matrix  $\mathbf{W}$  is [24]

$$\begin{aligned} \Delta \mathbf{W} &= \eta \frac{\partial \Phi}{\partial \mathbf{W}} \\ &= \eta (\mathbf{W}^{-T} + (\mathbf{1} - \mathbf{2z})\mathbf{x}^T) \end{aligned} \quad (10)$$

where  $\eta$  is the learning rate parameter. Using natural gradient we get

$$\begin{aligned} \Delta \mathbf{W} &= \eta (\mathbf{W}^{-T} + (\mathbf{1} - \mathbf{2z})\mathbf{x}^T) \mathbf{W}^T \mathbf{W} \\ &= \eta (\mathbf{I} + (\mathbf{1} - \mathbf{2z})(\mathbf{W}\mathbf{x})^T) \mathbf{W} \\ &= \eta (\mathbf{I} + (\mathbf{1} - \mathbf{2z})\mathbf{y}^T) \mathbf{W} \end{aligned} \quad (11)$$

Hence the weight update rule is

$$\Delta \mathbf{W} = \eta (\mathbf{I} + (\mathbf{1} - \mathbf{2z})\mathbf{y}^T) \mathbf{W} \quad (12)$$

ICA was performed on both the Architectures (I & II) as proposed by Bartlett et. al. [11, 14, 15, 21] and the matching of test images was done using Euclidean norm (L2 norm) as the similarity measure. Prior to performing ICA, the input data was sphered by passing  $\mathbf{X}$  through the whitening matrix

$$\mathbf{W}_z = 2 \times (\text{Cov}(\mathbf{X}))^{-(1/2)} \quad (13)$$

thus removing the first and second order statistics of data. The calculation of eigenvectors of the covariance matrix of a set of face images resulted in PC axes. The ICA was performed on the matrix containing the first forty percent of the Principal Component axes of total number of training images arranged in rows. The weights  $\mathbf{W}$  were updated according to Eq. (12) for 1600 iterations. The learning rate was initialized at 0.001 and annealed down to 0.0001. The Euclidean norm (L2 norm) was used as the similarity measure.

### III. COMBINING SOM AND PCA

SOM is an unsupervised learning process that has the property of topology preservation. It defines a mapping from an input space onto a set of nodes in a space that has dimension much lower than that of the input space. The set of nodes is topologically ordered. An image, divided into sub blocks, is mapped to a lower dimensional space with topologically ordered set of nodes thereby providing dimensionality reduction. Further feature extraction is provided with the method known as Karhunen – Loeve (KL)

transform via Principal Component Analysis (PCA). It is well known that PCA generates a set of orthogonal axes of projections known as principal components or the eigenvectors. PCA is applied to the weight matrix generated by mapping the image onto lower dimensional space using SOM. In order to further reduce the dimensionality, the eigenvectors with smaller eigenvalues are ignored and the eigenvectors corresponding to the largest eigenvalues are retained for image reconstruction. Figure 3 shows the flow chart for the proposed SOM – PCA combination algorithm.

Here two cases have been considered. Since the SOM weight matrix is of size (25x16), indicating that there are total of 25 neurons each with 16 weight elements. In the first case, the weight vectors of all the neurons were used for image reconstruction. After training the SOM, the weight matrix (25x16) of each image is retained and the eigenvectors of the weight matrix are found out. The eigenvectors corresponding to the eigenvalues retaining almost ninety nine percent of the energy (Table 1) are selected. The KL coefficients are also retained. At the time of recognition, the images are reconstructed and matching is done with the test images. Whereas in the second case, the weight vectors corresponding to first twenty of total twenty five neurons (two dimensional SOM size is 5x5) were considered for reconstruction of the image thereby reducing the memory space required to store the image.

#### IV. EXPERIMENTATION

The database (ORL face database) used for experimentation in this paper is composed of 400 images [22]. Each image is of resolution  $92 \times 112$ . The database contains 40 different persons. Each person has his/her 10 different images. These images vary in terms of facial expressions and facial details. These images have been taken at different times and lighting and the faces are in up-right position of frontal view with slight left right rotation.

The original image  $92 \times 112$  was resized to  $80 \times 80$  prior to further processing of the face image. Euclidean norm was used as the similarity measure to see which images are most alike. As many as five training images and the same number of test images were used for performing the experiments. There is no overlap between training and test sets. Each result is an average of three simulations. The experiments are as follows:

1. The first experiment was performed to see the effect of global and local processing on the recognition rate of the face recognition system. Two-dimensional self-organizing map of size 5-by-5 was chosen, for both global and local processing. The face image of size  $80 \times 80$  was concatenated to form a single vector of size  $1 \times 6400$ . This formed the input for the Self Organizing Map and it was trained. After training, a matrix of size  $25 \times 1$  was obtained and retained. The image was reconstructed (Figure 4(a)) with the help of this matrix at the time of recognition for matching purpose. As many as five training images and the same number of test images for the first ten classes of the image database were used for performing the experiments.

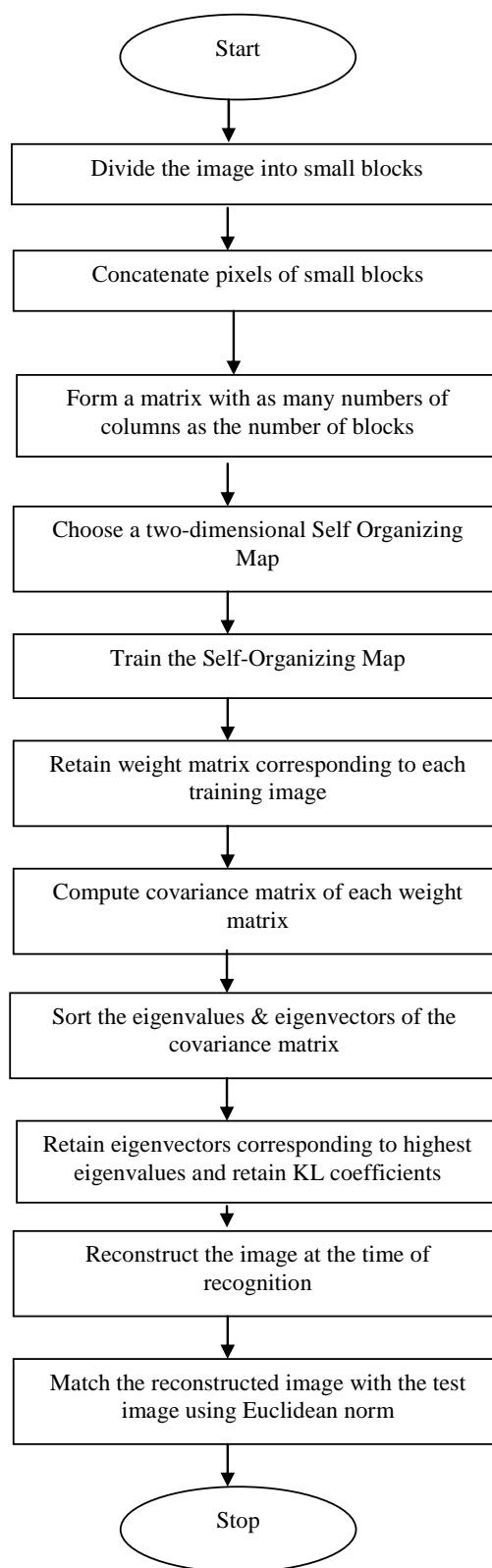


Fig. 3 Flowchart for SOM -PCA combination algorithm

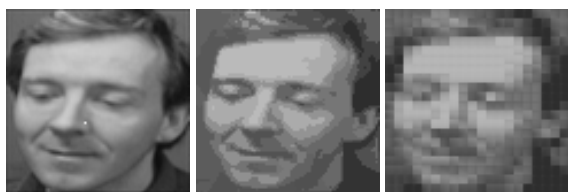


Fig. 4 Original image (a) Reconstruction using global processing (b) Reconstruction using local processing

In case of local processing, the image was first divided into small blocks. Block sizes of 4x4, 8x8 and 16x16 were chosen for the experimentation purpose. The pixels of each block were concatenated to form a single vector representing one small block to obtain the matrices of sizes 16x400 (400 columns with 16 pixels per column corresponding to one small block), 64x100 (100 columns with 64 pixels per column corresponding to one small block) and 256x25 (25 columns with 256 pixels per column corresponding to one small block) corresponding to 4x4, 8x8 and 16x16 size blocks respectively. These matrices formed the input for Self Organizing Map of size 5-by-5 and it was trained. The results of which were weight matrices of size 25x16, 25x64 and 25x256 respectively and these matrices were retained and used for reconstruction (Figure 4(b)) of images. So far as the local processing is concerned, Table 2 shows that there is no change in the recognition rates with respect to the change in the size of the block whereas the global processing system performs better in terms of recognition rate as compared to local processing system.

TABLE II  
 RECOGNITION RATE OF THE SYSTEM WITH CHANGING BLOCK SIZE

Recognition Rate (%)			
Method	Size of Block		
	4x4	8x8	16x16
LOCAL SOM (5 × 5)	96	96	96
GLOBAL SOM (5 × 5)	98		

2. The second experiment was performed to see the effect of changing the Self Organizing Map, SOM size on the performance of recognition system. For this purpose SOMs of two different sizes (3x3 and 5x5) were chosen and the experiment was performed for the first ten classes of the face database using five training images and the same number of test images and the matching was done using Euclidean norm. The block size for local processing was kept as 4x4. Table 3 and Figure 5 show that the recognition rate remains same for both, 3x3 and 5x5 sizes SOMs if the global processing method is used whereas local processing method results in the change in recognition rate as the size of the SOM is changed. There is increase in recognition rate of the system as the size of the SOM is increased.

TABLE III  
 RECOGNITION RATE OF THE SYSTEM WITH CHANGING SOM SIZE

Recognition Rate (%)		
Method	Size of SOM	
	SOM (3x3)	SOM (5x5)
LOCAL SOM	94	96
GLOBAL SOM	98	98

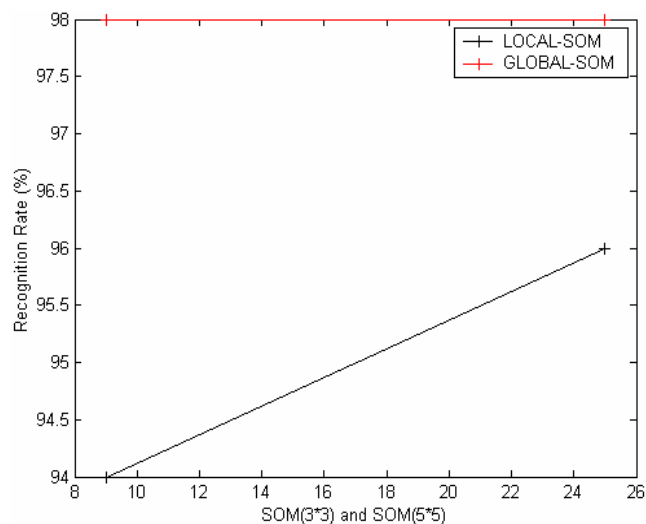


Fig. 5 Recognition rate as a function of changing size of the block

3. For this experiment, the sub blocks of size say 4x4 were chosen for the image of size 80x80. The pixels of each sub block were concatenated to form a single vector representing one sub block to result in a matrix of 16x400, having total of 400 columns, each of which representing 16 pixels corresponding to each sub block. This formed the input for the SOM. Two-dimensional SOM was chosen having say 5 nodes per dimension and it was trained resulting in a weight matrix of size 25x16. PCA was then applied to the transpose of the weight matrix and the eigenvectors corresponding to eigenvalues that retained almost ninety nine percent of the total energy (Table 1) were retained for reconstruction of the image. Euclidean norm (L2 norm) was used as the similarity measure for matching. As many as 5 training images and the same number of test images were used for performing the experiments. There was no overlap between training and test sets. This experiment was performed to see the effect of varying number of classes on the performance of the face recognition system. The 40 classes in the ORL face database were varied from 10 to 20 to 40 and recognition rate was found. In this experiment the ICA was also performed on the matrix containing the first forty percent of the Principal Component axes of total number of training images arranged in rows. Prior to performing ICA, the data was whitened by passing the input through the whitening matrix (13). The experiment was performed for both the architectures. Euclidean norm (L2 norm) was used as the similarity measure

for matching. Tables 4 & 6 and figures 6 & 8 show the recognition rate (%) as the number of classes is varied. Two-dimensional SOM was used having 5 as the number of nodes per dimension. As is clear from the graph, the recognition rate decreases as the number of classes is increased. This may be due to the fact that increase in number of classes increases the chances of wrong recognition because of more similar faces finally decreasing the performance of the system.

TABLE IV  
 RECOGNITION RATE OF THE FACE RECOGNITION SYSTEM WITH VARIING NUMBER OF CLASSES

Recognition Rate (%)			
Method	Number of Classes		
	10	20	40
SOM (5 × 5)	94.06	90.72	89.92
SOM (10 × 10)	94.06	91.86	90.82
PCA	93.39	90.25	89.51
SOM+PCA (1)	77.75	72.08	62.64

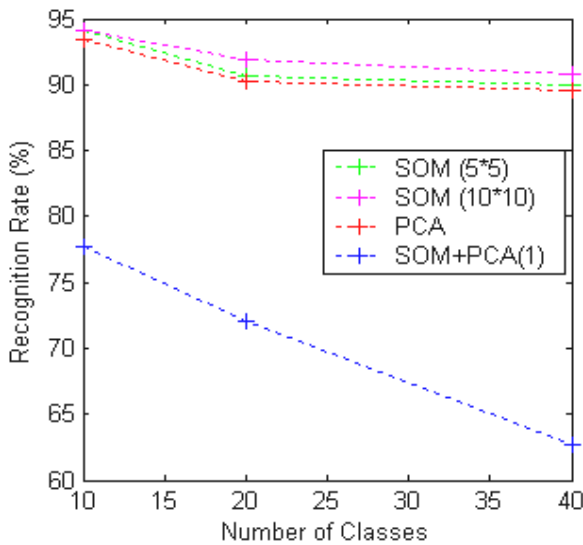


Fig. 6 Recognition rate as a function of number of classes

TABLE V  
 RECOGNITION RATE OF THE SYSTEM WITH CHANGING SUB BLOCK SIZE

Recognition Rate (%)			
Method	Size of sub block		
	(4 × 4)	(8 × 8)	(16 × 16)
SOM (5 × 5)	94.06	94.06	95.95
SOM+PCA (1)	77.75	77.17	72.83
SOM+PCA (2)	68.29	62.17	54.89

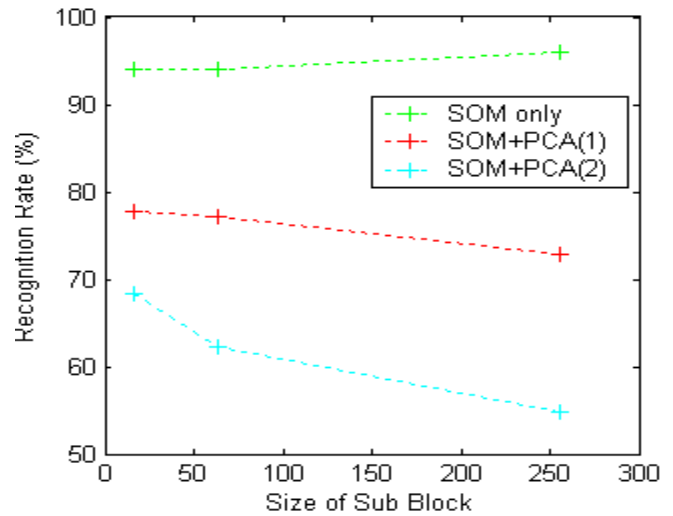


Fig.7 Recognition rate as a function of changing sub block size

TABLE VI  
 RECOGNITION RATE OF THE SYSTEM WITH VARYING NUMBER OF CLASSES FOR ICA

Recognition Rate (%)			
Method	Number of Classes		
	10	20	40
ICA (Arch-I)	91.20	90.60	88.70
ICA (Arch-II)	88.80	88.50	82.10

From the results it is clear that PCA is less sensitive to change in the number of classes as compared to ICA. The results corresponding to SOM are the best among all the techniques. From graph it is also clear that ICA-I is better than ICA-II. The table 5 and figure 7 show the effect of changing the sub block size on the performance of face recognition system. The experiment was performed on first 10 classes. As is clear from figure 7 that there is a little change in the recognition rate for SOM and SOM & PCA combined (1) techniques whereas for SOM & PCA combined (2), it is less as compared with the first two techniques and it reduces more rapidly as the size of sub block is increased.

4. In the last experiment the number of classes of the face database was varied from 10 to 20 to 40. This experiment was performed to see the effect of local and global processing methods on changing the number of classes of the face database. The block size was kept as 4×4 only for local processing. Table 7 and Figure 9 clearly show that the recognition rate decreases for both global and local processing systems as the number of classes is increased. The increase in the number of classes results in the increase in chances of similarity among the classes and hence results in the decrease in performance of the system. The results indicate that global processing still performs better than the Local processing method



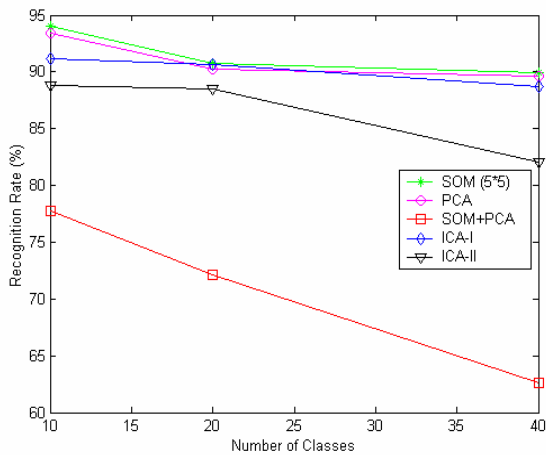


Fig. 8 Recognition rate as a function of number of classes

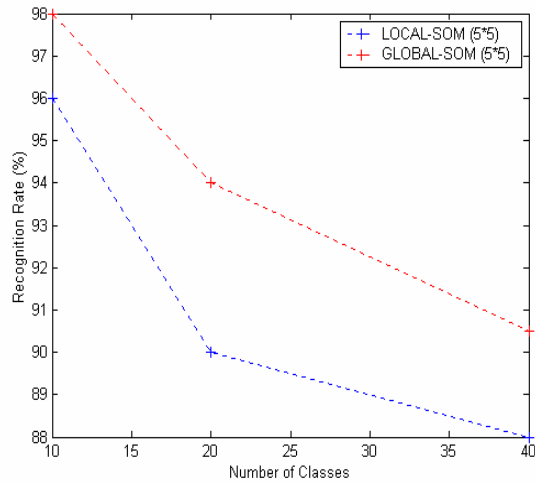


Fig. 9 Recognition rate as a function of changing number of classes

TABLE VII

RECOGNITION RATE OF THE SYSTEM WITH VARYING NUMBER OF CLASSES

Method	Recognition Rate (%)		
	Number of Classes		
	10	20	40
LOCAL SOM (5 × 5)	96	90	88
GLOBAL SOM (5 × 5)	98	94	90.50

### V. CONCLUSION

In this paper, a new idea was proposed in which PCA was integrated with SOM for feature selection and dimensionality reduction. Simulation results show that, though, the individual techniques SOM and PCA itself give excellent performance but the combination of these two can also be utilized for face recognition. The two different methodologies, global processing and local processing of face images using self-

organizing maps were explored and compared. From experimentation it was found that while training the self organizing map, the local processing approach took very less time as compared to global processing method for the very simple reason that pixel blocks were used that reduced the dimensionality of the data space that was to be represented topologically in the SOM space. The results show that there was no change in the performance of the recognition system as the size of the block is changed but the increase in the size of the self organizing map does result in the increase of recognition rate so far as local processing is concerned but still less than that using global processing. A comparison of the three unsupervised techniques was done along with a technique in which the two techniques SOM and PCA were combined together for dimensionality reduction and feature selection. The simulation results indicate that the performance of face recognition system decreases as the number of classes (subjects) is increased. This is true for all the three methods i.e. SOM, PCA, ICA (I & II), SOM & PCA combined and local and global processing as well. The decrease is more in case of SOM & PCA combined as compared to other methods. The reason for the decrease in performance of recognition system is that as the number of classes (subjects) increase, the chances of mismatch are more because of more similar faces.

### REFERENCES

- [1]. D.L.Swets and J.J.Weng, "Using Discriminant Eigenfeatures for Image Retrieval," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.18, No.8, pp.831-836, 1996.
- [2]. P.N.Belhumeur, J.P.Hespanha, D.J.Kriegman, "Eigenfaces Vs Fisherfaces: Recognition using Class Specific Linear Projection," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.19, No.7, pp.711-720, 1997.
- [3]. M.Kirby and L.Sirovich, "Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.12, No.1, pp.103-108, 1990.
- [4]. A.M.Martinez and A.C.Kak, "PCA versus LDA," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.23, No.2, pp.228-233, 2001.
- [5]. M.Turk and A.Pentland, "Eigenfaces for Recognition," Journal of Cognitive Neuroscience, Vol.3, No.1, pp.71-86, 1991.
- [6]. R. Chellapa, C.L.Wilson, S.Sirobey, "Human and Machine Recognition of Faces: A Survey," Proc. IEEE, Vol.83, No.5, pp. 705-740, 1995.
- [7]. Dinesh Kumar, C.S.Rai and Shakti Kumar, "Self Organizing maps for Face Recognition: Local vs Global Processing," Proceedings of International Conference on Systemics, Cybernetics and Informatics, ICSCI-2007, Vol.1, pp 758-761, Jan. 2007, Hyderabad, India.
- [8]. W.Zhao, R. Chellapa, A. Rosenfeld, P.J Phillips, "Face Recognition: A Literature Survey," ACM Computing Surveys, Vol.35, No.4, pp. 399-458, 2003.
- [9]. Dinesh Kumar, C.S.Rai and Shakti Kumar, "Face recognition using Self Organizing Map and Principal Component Analysis," Proceedings of IEEE International Conference on Neural Networks and Brain, ICNNB-2005, Vol. 3, pp 1469-1473, Oct. 2005, Beijing, China.
- [10]. C.Liu and H.Wechsler, "Evolutionary Pursuit and its Applications to Face Recognition," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.22, No.6, pp.570-582, 2000.
- [11]. M.S.Bartlett and T.J.Sejnowski, "Independent Components of Face Images: A Representation for Face Recognition," Proc. of the 4<sup>th</sup> Annual Joint Symposium on Neural Computation, Pasadena, 1997.
- [12]. B.Moghaddam, "Principal Manifolds and Bayesian Subspaces for Visual Recognition," International Conference on Computer Vision, Greece, pp.1131-1136, 1999



- [13]. B.Moghaddam and A.Pentland, "Probabilistic Visual Learning for Object Detection," Proc. International Conference on Computer Vision, pp. 786-793, 1995.
- [14]. M.S.Bartlett, H.M.Lades, T.J.Sejnowski, "Independent Component Representations for Face Recognition," Proc. of SPIE Symposium on Electronic Imaging: Science and Technology; Conference on Human Vision and Electronic Imaging III, California, 1998.
- [15]. M.S.Bartlett, "Face Image Analysis by Unsupervised Learning and Redundancy Reduction," Ph.D. Dissertation, University of California, San Diego, 1998.
- [16]. B.Moghaddam and A.Pentland, "Probabilistic Matching for Face Recognition," IEEE Southwest Symposium on Image Analysis and Interpretation, pp. 186-191, 1998.
- [17]. B.Moghaddam, T.Jebara, A.Pentland, "Efficient MAP/ML Similarity Matching for Visual Recognition," Proc. of Fourteenth International Conference on Pattern recognition, Vol. 1, pp.876-881, 1998.
- [18]. X.Tan, S.Chen, z.H.Zhou, F.Zhang, "Recognizing Partially Occluded, Expression Variant Faces from Single Training Image per Person with SOM and Soft k-NN Ensemble," IEEE Trans. on Neural Networks, Vol.16, No. 4, pp. 875-886, 2005.
- [19]. V.E.Neagoe, A.D.Ropot, "Concurrent Self Organizing Maps for Pattern Classification," Proc. of First International Conference on Cognitive Informatics, ICCI'02, pp. 304-312, 2002.
- [20]. S.Lawrence, C.L.Giles, A.C.Tsoi, "Convolutional Neural Networks for Face Recognition," Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.217-222, 1996.
- [21]. M.S.Bartlett, J.R.Movellan and T.J.Sejnowski, "Face Recognition by Independent Component Analysis," IEEE Trans. on Neural Networks, Vol.13, No.6, pp.1450-1464, 2002.
- [22]. <http://www.cam-orl.co.uk/facedatabase.html>.
- [23]. Simon Haykin, "Neural Networks – A Comprehensive Foundation," 2<sup>nd</sup> Edition, Pearson Education, 1999.
- [24]. A.J.Bell and T.J.Sejnowski, "An Information-maximization approach to blind separation and blind deconvolution," Neural computation, Vol.7, No.6, pp. 1129-1159, 1995.
- [25]. A.Hyvärinen, "Survey on Independent Component Analysis," Neural Computing Surveys, Vol.2, pp.94-128, 1999.
- [26]. A.Hyvärinen and E.Oja, "Independent Component Analysis: Algorithms and Applications," Neural Networks, Vol.3, No. 4-5, pp.411- 430, 2000.
- [27]. S.Z.Li and J.Lu, "Face Recognition using Nearest Feature Line Method," IEEE Trans. on Neural Networks, Vol.10, No.2, pp.439-443, 1999.
- [28]. P.J.Phillips, "Support Vector Machines Applied to Face Recognition," Technical Report, NISTIR 6241.
- [29]. X.He, S.Yan, Y.Hu, P.Niyogi, H.J.Zhang, "Face Recognition using Laplacianfaces," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.27, No.3, pp.328-340, 2005.
- [30]. T.Kohonen, "Self Organization and Associative Memory," 2<sup>nd</sup> Edition, Berlin, Germany: springer-Verlag, 1988.
- [31]. T.Kohonen, "Self Organizing Map," 2<sup>nd</sup> Edition Berlin, Germany: Springer-verlag, 1997.