Dimensionality Reduction using SOM based Technique for Face Recognition

Dinesh Kumar
Dept of C.S.E., G. J.University of Science and Technology, Hisar, Haryana, INDIA
Email: dinesh_chutani@yahoo.com

C.S. Rai* and Shakti Kumar**
*University School of Information Technology, GGS I.P. University, Delhi, INDIA
**Director, Institute of Science & Technology Klawad, Y.Nagar, Haryana, INDIA
Email: * csrai_ipu@yahoo.com,** shaktik@gmail.com

***Abstract*— Unsupervised or Self-Organized learning algorithms have become very popular for discovery of significant patterns or features in the input data. The three prominent algorithms namely Principal Component Analysis (PCA), Self Organizing Maps (SOM), and Independent Component Analysis (ICA) have widely and successfully been used for face recognition. In this paper a SOM based technique for dimensionality reduction has been proposed. This technique has also been successfully used for face recognition. A comparative study of PCA, SOM and ICA along with the proposed technique for face recognition has also been given. Simulation results indicate that SOM is better than the other techniques for the given face database and the classifier used. The results also show that the performance of the system decreases as the number of classes increase.**

***Index Terms*— Face Recognition, Principal Component Analysis, Self-Organizing Map, Independent Component Analysis**

I. INTRODUCTION

In face recognition system, we have database of images stored in the system. Whenever we get a new image, it is compared with the database of images already stored in the system. Neural Networks make use of new face image and the stored face images to determine if there is a match. It is a very interesting and challenging biometric technique of identifying individuals by facial features. There are many researchers working on pattern recognition, computer vision and image analysis for the last two decades. Face recognition has variety of potential practical applications such as for security, 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. Appearance based, rule based, feature based and texture based methods are the basic methods for face recognition. Eigenface method [1] was successfully and efficiently used for face recognition. This method is based on Principal Component Analysis that considers the statistics up to second order only. Another method [4,5] for face recognition was proposed that was insensitive to large variations in lighting and facial expressions. This method was known as Fisherfaces. The authors [2] in their work concluded that when the training data set is small, PCA can outperform Fisherfaces method. It was also shown that PCA is less sensitive to different training data set.

The above two methods give global representations. In some real world applications, it becomes necessary to preserve the local structure. Keeping in view the preservation of intrinsic geometry and the local structure, the authors [16] proposed a new method in which Locality Preserving Projections (LPP) were used for mapping the face images into the face subspace and this method was found most suitable for frontal face images. There are many methods proposed by different researchers that were based on second order statistics only whereas in some applications like Face Recognition we may need to deal with higher order statistics also because these higher order statistics may contain important information. The authors [6, 7, 8, 12] introduced methods that considered the higher order statistics also. These methods were based upon Independent Component Analysis (ICA).

Self-Organizing maps (SOMs) [17, 18] have also been successfully used as a way of dimensionality reduction and feature selection for face space representations [9, 10, 11]. All the three techniques PCA, SOM and ICA have individually been used for face recognition. This paper proposes a SOM based technique for feature extraction and dimensionality reduction. The performance of proposed technique has been defined in terms of the recognition rate of face recognition system using the face database with subjects having variation in facial expressions and facial details. After a brief discussion of SOM, PCA and ICA in section 2, Section 3 proposes the SOM based technique. The experiments have been reported in section 4 and section 5 contained the conclusions.

II. PCA, SOM AND ICA

A. Principal Component Analysis

In applications like face recognition, we deal with high dimensional data. The high dimensional data is projected onto lower dimension while retaining most of the features required for reconstruction of data with acceptable level of error thereby leading to less storage requirement. Principal Component Analysis, also known as Karhunen-Loeve (KL) transform, an unsupervised technique, provides mapping of n-dimensional data space onto m-dimensional data space where $m < n$. In context of face recognition, a set of N sample images $\{\mathbf{I}_1, \mathbf{I}_2, \cdots, \mathbf{I}_\mathbf{N}\}$ taking values in an n-dimensional image space is considered and there is a mapping from the original n-dimensional image space to m-dimensional feature space where $m < n$. The new feature vectors $\mathbf{Y_k}$ are defined by the following linear transformation $\mathbf{Y}_K = \varphi^T \mathbf{I}_K$ where φ is a matrix with orthonormal columns $[\mathbf{e_1}, \mathbf{e_2}, \cdots, \mathbf{e_m}]$, the eigenvectors of covariance matrix).

$$R = \sum_{k=1}^{N} \left(\mathbf{I_k} - \mathbf{I}_{mean}\right)\left(\mathbf{I_k} - \mathbf{I}_{mean}\right)^\mathbf{T} \qquad (1)$$

where $\mathbf{I}_{mean}$ is the mean image of all the samples. Only m numbers of n-dimensional eigenvectors of R corresponding to m largest eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_m$ are chosen.

B. Self-Organizing Maps

Self Organizing Feature Map is a neural network model of the unsupervised class. It is based on the fact that topological information that is there in high dimensional input data can be transformed onto one or two-dimensional layer of neurons. There is a competition among the neurons to be activated or fired and only one neuron that wins the competition is fired and is called winner-takes-all neuron. In a two dimensional neuronal field for the mapping we need input neurons to be exposed to a sufficient number of different inputs. For a given input, only the winning neuron and its neighbours adapt there connections. A similar weight update procedure is employed on many adjacent neurons which comprise topologically related subsets. Finally the resulting adjustments are such that it enhances the responses to the same or to a similar input that occurs subsequently.

C. Independent Component Analysis

Independent Component Analysis (ICA), an extension of linear transform, Principal Component Analysis (PCA) 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. Consider a random vector $\mathbf{x} = (x_1, x_2, \cdots, x_m)^T$, an m-dimensional observation vector and $\mathbf{s} = (s_1, s_2, \cdots, s_n)^T$ is the original source vector having n independent components. The observed data vector is modeled by the linear transformation as given below

$$\mathbf{x} = \mathbf{As} \qquad (2)$$

where $\mathbf{A}$ is an $m \times n$ nonsingular mixing matrix. The aim of ICA algorithm is to find the components s_i as independent as possible so that the set of observed signals can be expressed as linear combination of statistically independent components. The original source vector $\mathbf{s}$ is recovered with the help of the following linear transformation

$$\mathbf{y} = \mathbf{Wx} \qquad (3)$$

where $\mathbf{y} = (y_1, y_2, \cdots, y_n)^T$ is an n dimensional output vector and $\mathbf{W}$ is $n \times m$ weight matrix. ICA is an information theoretic technique. Various objective functions based on information theoretic concepts such as negentropy, minimization of mutual information, maximum entropy, maximum likelihood have been used for source separation problem. This paper follows maximum entropy based ICA method for face recognition [8], the weight update rule for which is [15].

$$\Delta\mathbf{W} = \eta(\mathbf{I} + (1 - 2\mathbf{z})\mathbf{y}^T)\mathbf{W} \qquad (4)$$

where $\mathbf{z}$ is the output of nonlinearity (logistic function) used. ICA has been performed on both the Architectures (I & II) as proposed in [8].

In the proposed method, an image, divided into sub blocks, is mapped to a lower dimensional space with topologically ordered set of nodes. Further Principal Component Analysis (PCA) is applied. 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. The eigenvectors with smaller eigenvalues are ignored and the eigenvectors corresponding to the largest eigenvalues are retained for image reconstruction

III. ALGORITHM

In the proposed technique, the dimensionality of the face image is reduced using two dimensional self organizing feature maps and further using principal component analysis technique thereby reducing the storage requirement. The steps to the proposed algorithm are as follows

1. A face image $\mathbf{I}$ of size $n \times n$ was divided into sub blocks of size say $a \times a$ resulting in total of $p = (n * n)/(a * a)$ blocks each of which contains $q = a * a$ number of elements, concatenation of which produces a vector to represent one block resulting in a

matrix $\mathbf{X} = [X_1, X_2, \ldots, X_p]$ of size $q \times p$. This gives a stream of training vectors $\{X_i\}_{i=1}^{p}$

2. Consider two dimensional $(r \times r)$ map of neurons each of which is identified as index $jk,\ j, k = 1, 2, \ldots, r$. The $jkth$ neuron has an incoming weight vector $W_{jk} = (w_{1,jk}, \ldots, w_{q,jk})$ at instant i. The value of neighbourhood around the winning neuron as h_{JK} at instant i. Initialize weights W_{jk}, neighbourhood h_{JK} and the learning rate η_0

3. Pick a sample vector X_i at random and present it to a two dimensional $(r \times r)$ map of neurons with a total of $z = r * r$ neurons.

4. Find out best matching (winning neuron) using following distance criterion

$$\| X_i - W_{JK(i)} \| = \min_{jk} \{ \| X_i - W_{jk(i)} \| \}$$

where W_{JK} is the best matching weight vector.

5. Update the synaptic weight vectors of only the winning cluster

$$W_{jk(i+1)} = W_{jk(i)} + \eta_i (X_{(i)} - W_{jk(i)}) \qquad jk \in h_{JK(i)}$$

6. Update learning rate η_i and the neighbourhood $h_{JK(i)}$

7. Continue with step 3 until no noticeable changes in the feature map are observed. Finally a matrix M of size $z \times q$ is obtained.

8. Compute the eigenvectors and eigenvalues of the covariance matrix $M^T M$, sort the eigenvectors and retain the eigenvectors corresponding to highest eigenvalues.

9. Calculate the KL coefficients $(M^T * Eigenvectors)$ and retain them.

10. Repeat the above steps for all training images.

11. Reconstruct the images at the time of recognition and match with the test image using nearest neighbour classifier.

At the start of the algorithm the neighbourhood $h_{JK(i)}$ usually includes all neurons in the vector field and its value reduces gradually. During the initial period of adaptation called the ordering phase, the learning rate η_i is kept close to unity and then decreases either linearly or exponentially or inversely with index i. During the tuning phase which occurs after ordering phase, it has a very small value but never zero. For the experimentation purpose hextop topology has been chosen together with linkdist as the distance function. The ordering phase learning rate was kept 0.9 while maintaining the tuning phase rate as 0.02.

For principal component analysis, the energy preservation factor, EPF, was computed by retaining only n number of eigenvalues for total of 200 images (200 eigenvalues).

$$EPF = \frac{\sum_{i=1}^{n} \lambda_i}{\sum_{i=1}^{L} \lambda_i} \times 100 \qquad (5)$$

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

TABLE I.
ENERGY PRESERVATION FOR DIFFERENT VALUES OF N

No.of feature Eigenvectors (n)	199	160	120	100	80	40	20
Eigen Preservation Factor	100	98.67	96.06	94.13	91.53	81.72	69.82

IV. EXPERIMENTATION

ORL face database has been used for experimentation in this paper [13]. The database is composed of 400 images with each image having a resolution 92×112. As many as 40 different persons are contained in the database and that too 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 92×112 was resized to 80×80 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 three different sets of five training images and the same number of test images were used for performing all the experiments. There is no overlap between training and test sets. The experiments are as follows:

1. In the first experiment the image was divided into sub blocks. Sub blocks of size say 4×4 were chosen. The pixels of each sub block were concatenated to form a single vector representing one sub block. We had a matrix of 16×400, 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 25×16. The same experiment was repeated for two-dimensional SOM having 10 nodes per dimension. Figure 1 and table 2 show that the recognition rate for SOM (10×10) is slightly better than that obtained for SOM (5×5). This is due to the fact that more number of neurons in SOM (10×10) has been considered to represent the image.

TABLE II.
RECOGNITION RATE OF THE FACE RECOGNITION SYSTEM FOR DIFFERENT VALUES OF TWO-DIMENSIONAL SOM

Recognition Rate (%)			
Method	Number of Classes		
	10	20	40
SOM (5×5)	94.06	90.72	89.92
SOM (10×10)	94.06	90.72	89.92

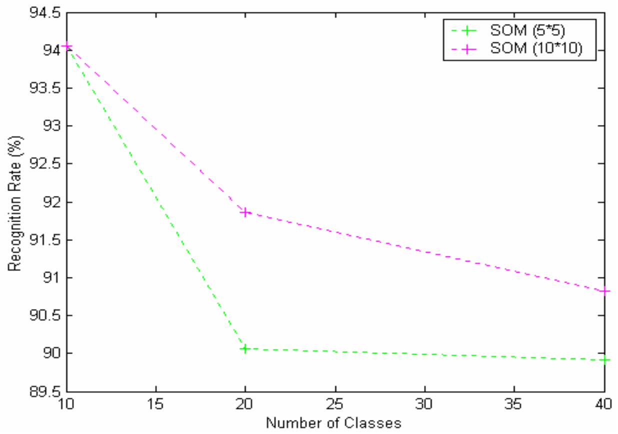

Fig. 1 Recognition rate as a function of SOM size.

2. The second experiment was performed for the proposed method. Here the eigenvectors of the weight matrix (size 25×16) were found. PCA was then applied to the transpose of the weight matrix and the eigenvectors corresponding to largest eigenvalues (Table I) were retained for reconstruction of the image. The KL coefficients are retained and at the time of recognition, the images are reconstructed and matching is done with the test images. The table III shows the results for PCA, SOM and the proposed technique. Figure 2 and table III show that SOM performs better than PCA and the proposed one. The recognition rate decreases as the number of classes is increased.

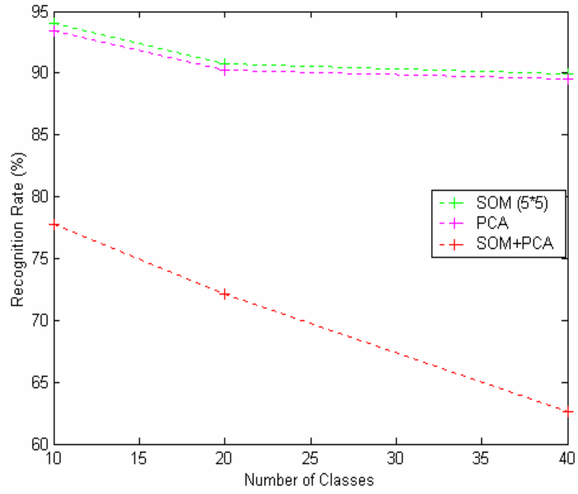

Fig. 2 Recognition rate as a function of number of classes

3. This experiment was performed to see the effect of changing the sub block size on the performance of face recognition system. The experiment was performed on first 10 classes of the database. Figure 3 and table IV shows that there is a little effect of the change in size of the sub block on the recognition rate for SOM and the proposed technique.

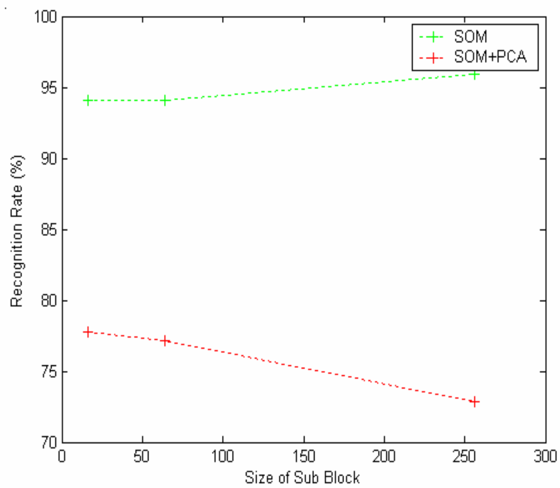

Fig. 3 Recognition rate as a function of sub block size

TABLE III.
RECOGNITION RATE OF THE FACE RECOGNITION SYSTEM FOR VARYING NUMBER OF CLASSES

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

TABLE IV.
RECOGNITION RATE OF THE FACE RECOGNITION SYSTEM FOR VARYING 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	77.75	77.17	72.83

4. ICA was performed on both the Architectures (I & II) and the matching of test images was done using Euclidean norm (L2 norm) as the similarity measure. The calculations 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 X of total number of training images arranged in rows. Prior to performing ICA, the input data was whitened by passing X through the whitening matrix $W_Z = 2\times(Cov(X))^{-(1/2)}$ thus removing the first and second order statistics of data. The weights W were updated as per weight update rule $\Delta\mathbf{W} = \eta(\mathbf{I} + (1-2\mathbf{z})\mathbf{y}^T)\mathbf{W}$ 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. Figure 4 and table 5 show that ICA architecture I perform better than architecture II.

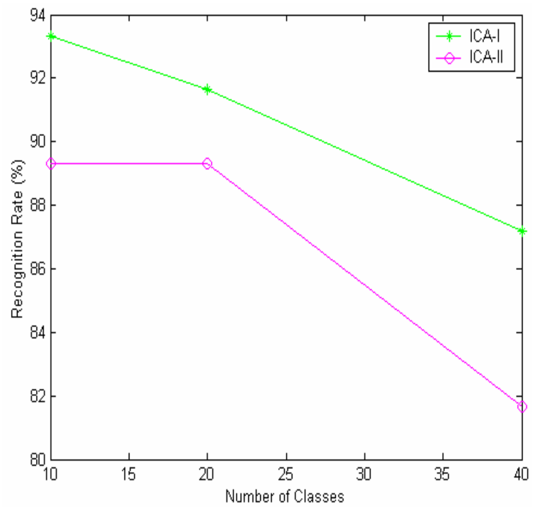

Fig. 4 Recognition rate as a function of number of classes

TABLE VI.
RECOGNITION RATE OF THE FACE RECOGNITION SYSTEM FOR VARYING NUMBER OF CLASSES

Method	Recognition Rate (%)		
	Number of Classes		
	10	20	40
ICA (Arch-I)	93.33	91.66	87.17
ICA (Arch-II)	89.33	89.33	81.67

5. Finally Table VI and figure 5 show the results for individual techniques i.e. PCA, SOM and ICA (Architecture I and II) and the proposed technique. In order to get the more exact values of the results the number of simulations were increased. Each result is an average of six different sets of nonoverlapping training and test images.

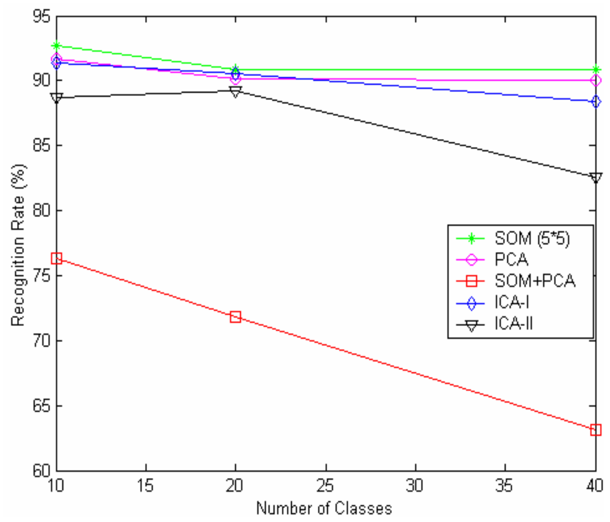

Fig. 5 Recognition rate as a function of number of classes

TABLE V.
RECOGNITION RATE OF THE FACE RECOGNITION SYSTEM FOR VARYING NUMBER OF CLASSES FOR PCA, SOM, ICA AND PROPOSED TECHNIQUE

Method	Recognition Rate (%)		
	Number of Classes		
	10	20	40
SOM (5 × 5)	92.67	90.83	90.83
PCA	91.67	90.17	90.00
SOM+PCA	76.33	71.83	63.08
ICA-I	91.33	90.50	88.42
ICA-II	88.67	89.17	82.58

IV. CONCLUSIONS

In this paper a new idea was proposed where dimensionality reduction was done using SOM and PCA. It was shown that this combination can also be used for face recognition. The performance of the proposed method is fairly good. The simulation results indicate that the performance of SOM (10×10) is better than SOM (5×5). This is due to the fact that more number of neurons in SOM (10×10) has been considered to represent the image. The simulation results also 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) and the proposed one. The reason for this is that as the number of classes (subjects) increase, the chances of mismatch are more because of more similar faces. The results also show that SOM performs better among all the techniques for the face database and the number of training and test images used as stated above.

REFERENCES

[1] M.Turk and A.Pentland, “Eigenfaces for recognition,” *Journal of Cognitive Neuroscience*, Vol.3, No.1, pp.71-86, 1991.

[2] 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

[3] Dinesh Kumar, C.S.Rai and Shakti Kumar, “Face recognition using self organizing map and principal component analysis,” *Proc. of IEEE International Conference on Neural Networks and Brain*, ICNNB-2005, Vol.3, pp 1469-1473, Beijing, China, 2005.

[4] 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.

[5] 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.

[6] M.S.Bartlett and T.J.Sejnowski, “Independent Components of face images: A Representation for Face Recognition,” *Proc. of the 4th Annual Joint Symposium on Neural Computation*, Pasadena, 1997.

[7] 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.

[8] M.S.Bartlett, “Face image analysis by unsupervised learning and redundancy reduction,” *Ph.D. Dissertation*, University of California, San Diego, 1998.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] http://www.cam-orl.co.uk/facedatabase.html.

[14] Simon Haykin, “Neural networks – A comprehensive foundation,” 2nd Edition, Pearson Education, 1999.

[15] 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.

[16] 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.

[17] T.Kohonen, “Self organization and associative Memory,” 2nd Edition, Berlin, Germany: springer-Verlag, 1988.

[18] T.Kohonen, “Self organizing map,” 2nd Edition Berlin, Germany: Springer-verlag, 1997.