

Dimensionality Reduction technique using Neural Networks – A Survey

Prof. Mrs. Shamla Mantri
Dept. of Computer Engg
MIT College of Engineering,
Pune University, India

Nikhil S. Tarale
Student, Dept. of Computer Engg
MIT College of Engineering,
Pune University, India

Sudip C. Mahajan
Student, Dept. of Computer Engg
MIT College of Engineering,
Pune University, India

Abstract— A self-organizing map (SOM) is a classical neural network method for dimensionality reduction. It comes under the unsupervised class. SOM is a neural network that is trained using unsupervised learning to produce a low-dimensional, discretized representation of the input space of the training samples, called a map. SOM uses a neighborhood function to preserve the topological properties of the input space. SOM operates in two modes: training and mapping. Using the input examples, training builds the map. It is also called as vector quantization.

In this paper, we first survey related dimension reduction methods and then examine their capabilities for face recognition. In this work, different dimensionality reduction techniques such as Principal component analysis [PCA], independent component analysis [ICA] and self-organizing map [SOM] are selected and applied in order to reduce the loss of classification performance due to changes in facial expression. The experiments were conducted on ORL face database and the results show that SOM is a better technique.

Keywords- Principal component analysis [PCA]; Independent component analysis [ICA]; self-organizing map [SOM]; Face recognition.

I. INTRODUCTION

Biometrics refers to the study of methods for uniquely recognizing human based upon one or more intrinsic or behavioral characteristics. Biometrics is used to identify the input sample when compared to a template used in cases to identify specific people by certain characteristics.

Face recognition is an important part of today's emerging biometrics and video surveillance markets. Face recognition can benefit areas of airport security, access control, driver's license, passports; homeland defense, customs and immigration etc. face recognition has been a research area for almost 30 years with significant increased research activity since 1990.

This has resulted in successful algorithms and the introduction of commercial products. The benefit of using a computer system for face recognition would be its capacity to handle large amount of data and the ability to do a job in a predefined repeated manner.

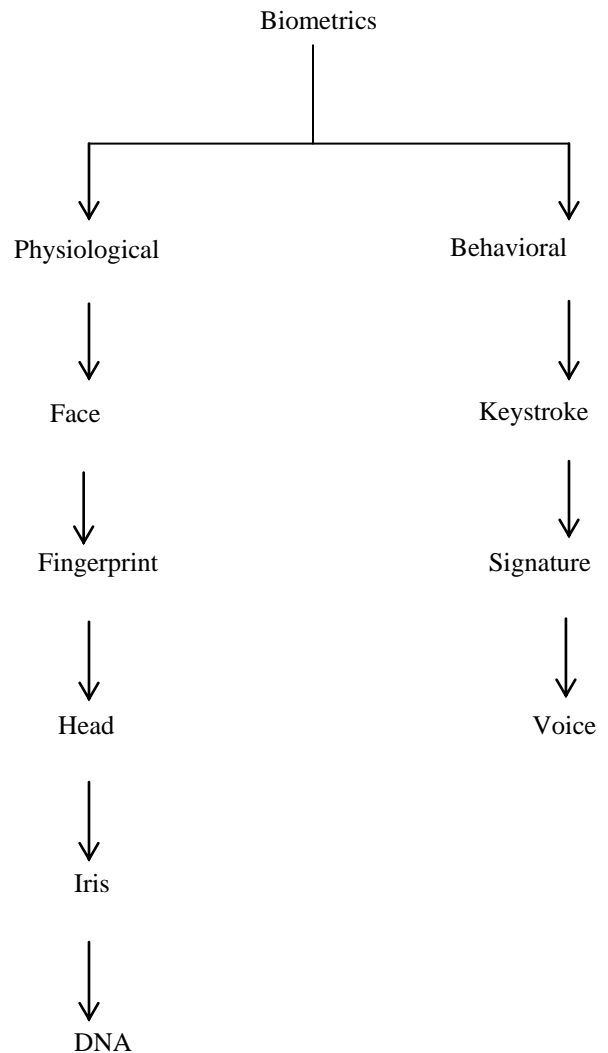


Fig 1. Biometrics

Suppose we have a collection of points of n-dimensional real vectors drawn from an unknown probability distribution but the situation in most of the cases is where dimensions are very large.

This leads one to the methods of dimensionality reduction that allows one to represent data in lower dimension space. The steps for face recognition in [3] are as follow:

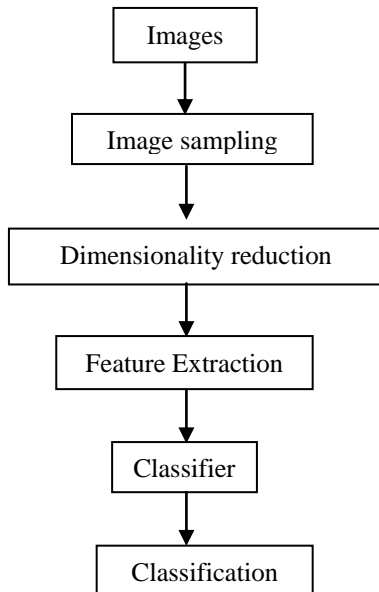


Figure 2. Face recognition

1. Selection and sampling of the image from the database;
2. Dimensionality reduction;
3. Feature extraction;
4. Classifier

A. Selection and Sampling:

Sampling is selection of those points which are required to represent the given image. It is mapping of the image from a continuum of points in space to a discrete set.

B. Dimensionality and reduction

In this phase, the dataset i.e. the images are reduced to minimum size by sampling. Thus, a reduced data set is obtained.

C. Feature Extraction

This process deals with extracting patterns from the data by using techniques such as classification, regression, segmentation or Deviation detection.

D. Classifier

Classification involves mapping data into one of several predefined or newly discovered classes.

In practical situation one is often forced to use linear or even sub-linear techniques. Principal component analysis [PCA], Independent component analysis [ICA] and Self-organizing mapping are the popular form of linear techniques. Using the SOM as a feature extraction method in face recognition applications is a promising approach, because the learning is unsupervised, no pre-classified image data are needed at all. When high compressed representations of face images or their parts are formed by SOM, the final

classification produced is fairly simple, needing only a moderate number of labeled training samples.

In this paper, we have introduced face recognition algorithms based on this consideration.

Technically, a principal component can be defined as a linear combination of optimally-weighted observed variables.

II. PRINCIPAL COMPONENT ANALYSIS

Principal Component Analysis is a variable reduction procedure.

It is useful for removing redundancy (i.e. some variables are correlated to each other) among data and to reduce the observed variables into smaller no of principal components that will account for most of the variance in the observed variables.

Principal Component Analysis provides mapping of n-dimensional data space into m-dimensional data space, a set of n sample images $\{I_1, I_2, \dots, I_N\}$ taking values of n-dimensional is considered and there mapping from original n-dimensional image spaced to m-dimensional feature space where $m < n$. the new feature vector Y_K is defined by the following linear transformation $Y_K = \varphi^T I_k$

where φ is a matrix with orthogonal columns $[e_1, e_2, \dots, e_m]$, the Eigen vectors covariance matrix

$$R = \sum_{k=1}^N (I_k - I_{mean})(I_k - I_{mean})^T$$

where I_{mean} is the mean image of the samples. Only m no of n-dimensional eigenvectors of R corresponding to m largest eigenvalues $\lambda_1, \lambda_2, \lambda_3$ are chosen.

III. INDEPENDENT COMPONENT ANALYSIS

ICA is a statistical and computational technique for revealing hidden factors that underlie sets or random variables, measurements, or signals. ICA is superficially related to principal component analysis and factors analysis.

The ICA algorithm aims at finding S_i component as independent as possible so that the set of observed signals can be expressed as a linear combination of statistically independent components.

We have a random vector

$X = (x_1, x_2, x_3, \dots, x_m)^T$, and an m-dimensional vector for observation and original source vector with n independent components.

$$S = (s_1, s_2, \dots, s_n)^T$$

The linear transformation can be modeled as

$$X = AS$$

Here, A is an $m \times n$ non-singular matrix.

The original source vector S recovered by following transformation:

$$Y = W X ;$$

where $Y = (y_1, y_2, \dots, y_n)^T$ is an n dimensional output vector, and W is an $n \times m$ weight matrix. In the information theoretic technique like ICA, 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 [5], the weight update rule for which is [6].

$$\Delta W = \eta (I + (1 - 2z)y^T) W$$

Where z is the output of nonlinearity (logistic function) used. ICA has been performed on both the Architectures (I & II) as proposed in [5].

IV. SELF-ORGANIZING MAP

T. Kohonen introduced self-organizing map[1]. It is unsupervised learning process, which learns the distribution of a set of patterns without any class information. It has the property of topological preservation.

SOMs have also been successfully used in dimensionality reduction and feature selection for face space representations.

Algorithm

PCA is applied as it generates a set of orthogonal axes of projections known as principal components or eigen vectors. PCA is applied to weight matrix generated by mapping the image onto lower dimensional space using SOM. Only the Eigen vectors for large values are considered and those for smaller values are ignored. The steps are as follows:-

Step 1. A face image of size $m \times m$ was divided into sub-blocks of size $b \times b$ resulting in total of $p = (m \times m) / (b \times b)$ blocks each of which gives $q = b \times b$ number of elements, concatenation of which gives a vector to represent one block resulting in a matrix $X = [X_1, X_2, \dots, X_p]$ of size $q \times p$. This gives a stream of training vectors $\{X_i\}_{i=1}^p$

Step 2. Consider 2-dimensional ($s \times s$) map of neurons each of which is identified as index $jk, j, k = 1, 2, \dots, s$.

The jk th neuron has an incoming weight

$W_{jk} = (w_{1,jk}, \dots, w_{q,jk})$ at instant i . The value of neighborhood function around the winning neuron as h_{jk} at instant i . Initialize weight W_{jk} , neighborhood h_{jk} and the learning rate η_0 .

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

Step 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.

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

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

Step 6. Update learning η_i and the neighborhood $h_{jk(i)}$

Step 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.

Step 8. Compute the Eigen vectors and Eigen values of the covariance matrix $M^T M$, sort the Eigen vectors and retain the Eigen vectors corresponding to highest values.

Step 9. Calculate the KL coefficients ($M^T * \text{Eigenvectors}$) and retain them.

Step 10. Repeat above steps for all training images.

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

V. TRAINING AND TEST DATA

For the face recognition experiment, we partitioned the ORL database into a training set and a testing set. The partition is done as follow:

First, k images are selected for test; the remaining images i.e. $(10 - k)$ are used for training set and for computing the projected matrix. All the ten images in the training and test sets were projected to a dimension reduced space.

VI. EXPERIMENTATION

Here the eigenvectors of the weight matrix were found. PCA was then applied to the transpose of the weight matrix and the Eigen vectors corresponding to the largest eigenvalues were retained for reconstruction of the image. The table shows the results for PCA, SOM and PCA+SOM

TABLE I. 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

The above table and the figure given below shows that SOM performs better than PCA and PCA + SOM.

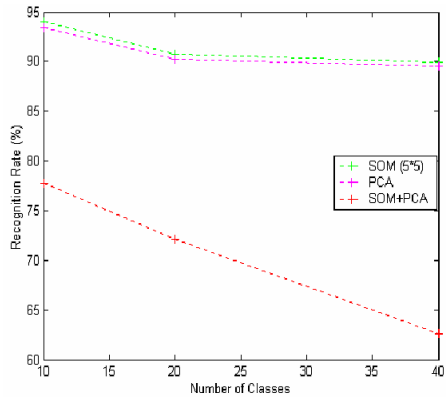


Figure 3. Recognition Rate as a function of number of classes

We used the AT&T database [10] for the face recognition experiments. Here experiments were done with nearly 400 images with variation of 40 persons. A preview image of the Database of Faces is as shown in Figure 4.



Figure 4. Example of faces in AT&T database.

The experimental observations of the experiments performed on datasets are shown in Table

TABLE II. EXPERIMENTAL OBSERVATIONS

Criteria	Values
Training images	30
Testing images	370
Learning coefficient	0.01
Iterations	50
Recognition Rate	85.5%

VII. CONCLUSION

The SOM algorithm is a typical dimensionality reduction technique which has good properties to preserve topological relationships even in lower dimensional space. The algorithm is very suitable for using K nearest neighbor classifier. The

experimental results show that our proposed algorithm performs better and faster in real data set.

An efficient system for face recognition using SOM has been proposed. Firstly, this system provides a general integration of multiple feature-sets using multiple self-organizing maps. Secondly, with the help of compressed feature vector, SOM is trained to organize all face images in database.

The highest average recognition rate of 85.5% is obtained for 40 persons' 400 images of AT&T database, where the training is done on 30 images only and tested on remaining images. Thus, the SOM method is an efficient face recognition process.

REFERENCES

- [1] T. Kohonen, "The Self -Organizing Map", IEEE, VOL. 78, NO. 9, SEPTEMBER 1990.
- [2] Santaji Ghorpade, Jayshree Ghorpade, Shamla Mantri, "Patter Recognition using Neural Networks", International Journal of Computer Science and Technology (IJCSST), Vol 2, No 6, December 2010
- [3] Santaji Ghorpade, Jayshree Ghorpade, Shamla Mantri, "Neural Networks for face recognition Using SOM", IJCSST Vol1, issue 2, December 2010
- [4] Jayshree Ghorpade, Siddhant Agarwal, "SOM and PCA Approach for Face Recognition – A Survey", International Journal of Computer Trends and Technology – March to April Issue – 2011
- [5] M. S. Barlett, "Face image analysis by unsupervised learning and redundancy reduction," *Ph.D. Dissertation*, University of California, San Diego, 1998.
- [6] 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.
- [7] Dinesh Kumar, C. S. Rai and Shakti Kumar, "Dimensionality Reduction using SOM based Technique for Face Recognition", *Journal of Multimedia*, VOL. 3, NO. 1, MAY 2008.
- [8] A. S. Raja and V. JosephRaj, "A New Multimodal Technique Without Subject's Cooperation Using Neural Network Based Organizing Maps", International Conference on Electronics and Information Engineering (ICEIE 2010).
- [9] Qiu Chen, Koji Kotani, Feifei Lee and Tadahiro Ohmi, "Face Recognition Using Self – Organizing Maps"
- [10] AT&T Laboratories Cambridge, The database of faces at <http://www.cl.cam.ac.uk/research/dtg/attarchive/facesataglab.html>.
- [11] Li chaoyang, lin fang, xie yinxiang, "Face recognition using Self-organizing features map and support vector machines",
- [12] M.A.kerin, T.J stomham. "face recognition using digital neural network with self-organising capabilities"
- [13] Firegoire Lefebvre and Christophe Garcia, " A probabilistic self-organizing map for Face Recognition".
- [14] Dey, A. K., & Saha, S. (2010). A Method of Genetic Algorithm (GA) for FIR Filter Construction : Design and Development with Newer Approaches in Neural Network Platform. International Journal of Advanced Computer Science and Applications - IJACSA, 1(6).
- [15] Trivedi, J. A. (2011). Framework for Automatic Development of Type 2 Fuzzy , Neuro and Neuro-Fuzzy Systems. International Journal of Advanced Computer Science and Applications - IJACSA, 2(1), 131-137.