Face Recognition using Tchebichef Moments

D. Sridhar*, Dr I. V. Murali Krishna**
* Balaji Institute of Technology and Science, Warangal (AP), India.
** Chief Consultant Offshore Programs in Asia, Jackson State University, USA

Article Info	ABSTRACT
	In this paper, a new Face Recognition method based on Discrete orthogonal Tchebichef moments with Linear Discriminant Analysis and Probabilistic Neural Network is proposed. Tchebichef moments are having good energy compaction property that made them useful in image compression and dimensionality reduction operations. Moreover the translation and scale invariant properties of Tchebichef moments are very much useful in almost all pattern recognition applications. The proposed face recognition method consists of three steps, i) Dimensionality reduction using Tchebichef moments ii) Feature extraction using Linear Discriminant Analysis and iii) classification using Probabilistic Neural Network. Linear Discriminant Analysis selects features that are most effective for class seperability in addition to dimensionality reduction. Combination of Tchebichef moments and Linear Discriminant Analysis is used for improving the capability of Linear Discriminant Analysis when few samples of images are available. Probabilistic Neural network (PNN) is a promising tool and gives fast and accurate classification of face images. Evaluation was performed on two face data bases. First database of 400 face images from Olivetty Research Laboratories (ORL) face database, and the second database of thirteen students are taken. The proposed method gives fast and better recognition rate when compared to other classifiers. The main advantage of this method is its high speed processing capability and low computational requirements.

Corresponding Author:

First Author,
Balaji Institute of Technology and Science, Warangal (AP), India.
Email: sreedhardsr@yahoo.co.in

1.   INTRODUCTION

Face recognition [1] one of the primary biometric technologies became more important owing to rapid advances in technologies such as digital cameras, Internet and mobile devices and increased demands on security. Face recognition has several advantages over other biometric technologies, it is natural, non intrusive and easy to use. But face recognition is one of the challenging problems in research, till now there is no unique solution for all face recognition applications. The wide range of variations in human face due to view point, pose, illumination and expression deteriorate the recognition performance of the Face recognition systems. But everyone accept that the face recognition system is good, if it has less computational complexity, good recognition performance and occupies less memory.

In any Face Recognition system Dimensionality Reduction and Feature Extraction are very important aspects. Face images though small in size are having large dimensionality this leads to very large computational time, complexity and memory occupation. The performance of any classifier mainly depends on high discriminatory features of the face images [2] – [3]. In the proposed method we used Discrete orthogonal Tchebichef moments and LDA for dimensionality reduction and feature extraction.

Tchebichef moments preserve almost all the information of the image in few coefficients. Keeping these coefficients and ignoring the rest, we can reduce the dimensionality of the face image features. When dimensionality of the face images is high, LDA is not applicable. To resolve this problem we combine the Tchebichef moments and LDA methods [4]. By applying Tchebichef moments we get discriminatory features of the images of dimensionality 81, which are applied to LDA. LDA searches the directions for maximum discrimination of classes in addition to dimensionality reduction and finally produces image features of

Journal homepage: *http://iaesjournal.com/online/index.php/ IJINS*

dimensionality of 39. Probabilistic Neural Network classifies the images based on their LDA features. The flow chart of the proposed method is shown in figure 1.
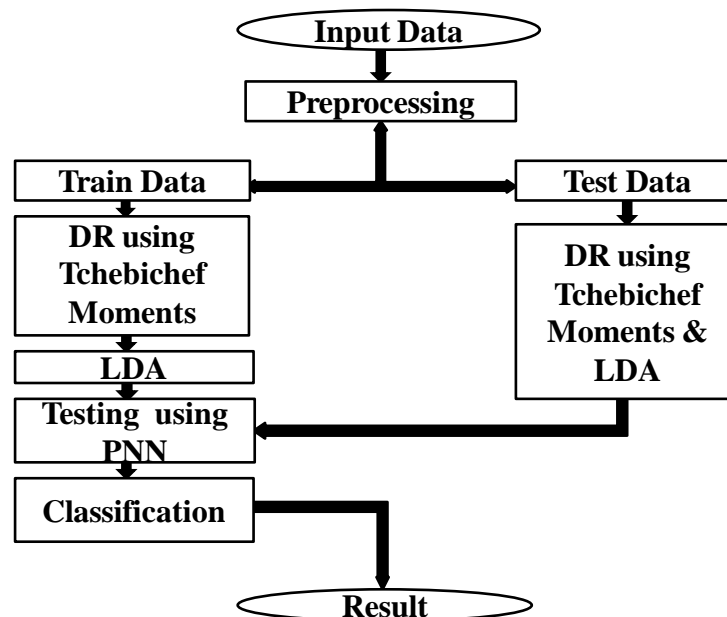

Figure 1. Flow chart for the proposed method

The rest of this paper is organized as follows: Section II discusses computation of Tchebishef moments for face images. Section III discusses the feature extraction with LDA. Section IV describes the PNN classifier. Section V shows experimental results, and discusses possible modifications and improvements to the system. Section VI presents concluding remarks.

2.   TCHEBISHEF MOMENTS

Hu [5] introduced the concept of moment, since than invariant moments and moment functions have been widely used in the fields of image analysis and pattern recognition. Hu's moment descriptors are invariant with respect to scale, translation and rotation of the image. However, the kernel function of geometric moments of order $(p + q)$, $\Psi_{pq} (x, y) = x^p y^q$, is not orthogonal, thus the geometric moments suffer from the high degree of information redundancy, and they are sensitive to noise for higher-order moments.

Zernike and Legendre moments were introduced by Teague who used the corresponding orthogonal polynomials as kernel functions. These orthogonal moments have been proved to be less sensitive to image noise as compared to geometric moments, and possess better feature representation ability that can be used in image classification. The translation and scale invariants of Zernike and Legendre moments were achieved by using image normalization method.

Both the Zernike and Legendre moments are defined as continuous integrals over a domain of normalized coordinates. The computation of these moments requires a coordinate transformation and suitable approximation of the continuous moment's integrals, leading to further computational complexity and discretization error. To overcome the shortcoming of the continuous orthogonal moments, Mukundan *et al*. proposed a set of discrete orthogonal moment functions based on the discrete Tchebichef polynomials [6]. The use of discrete orthogonal polynomials as basis functions for image moments eliminates the need for numerical approximations and satisfies perfectly the orthogonality   property in the discrete domain of image coordinate space. This property makes the discrete orthogonal moments superior to the conventional continuous orthogonal moments in terms of image representation capability.

The two-dimensional Tchebichef moment of order (p + q) of an image intensity function, $f(x, y)$, is defined as

$$T_{pq} = \frac{1}{\rho(p,N)\rho(q,N)}\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} t_p(x)t_q(y)f(x,y)$$

(1)

Where p, q = 1, 2, 3, …, N-1 and $\rho(p,N) = \dfrac{N\left(1-\dfrac{1^2}{N^2}\right)\left(1-\dfrac{2^2}{N^2}\right)...\left(1-\dfrac{p^2}{N^2}\right)}{2p+1}$

(2)

Is the squared norm and $t_p(x)$ is the scaled Tchebichef polynomial of order p. The orthogonal Tchebichef polynomials are defined by the recursive relation

$$\hat{t}_p(x) = \alpha(2x+1-N)\hat{t}_{p-1}(x) + \beta\,\hat{t}_{p-2}$$

(3)

Where p = 0, 1, 2, … , N − 2 ;   x = 0, 1, 2, …, N − 1 and

$$\alpha = \frac{\sqrt{4p^2-1}}{p\sqrt{N^2-p^2}}$$

(4)

$$\beta = \frac{(p-1)\sqrt{2p+1}\sqrt{N^2-(p-1)^2}}{p\sqrt{2p-3}\sqrt{N^2-p^2}}$$

(5)

The initial conditions of the above recurrence relation are

$$\hat{t}_0(x) = N^{-\frac{1}{2}}$$

$$\hat{t}_1(x) = \frac{\sqrt{3}(2x+1-N)}{\sqrt{N(N^2-1)}}$$

(6)

The discrete orthogonal polynomials defined as above satisfy the following condition for all p

$$\rho(p,N) = \sum_{i=0}^{N-1}\left\{\hat{t}_p(i)\right\}^2 = 1.0$$

(7)

The Tchebichef polynomials can be renormalized to minimize propagation of any numerical errors through the recurrence relation

$$\hat{t}_p(x) = \frac{\hat{t}_p(x)}{\sqrt{\sum_{x=0}^{N-1}\left(\hat{t}_p(x)\right)^2}}$$

(8)

It is observed that the reconstruction accuracy improves considerably by the renormalization of the orthogonal Tchebichef moments.

3.   LINEAR DISCRIMINANT ANALYSIS (LDA)

Linear discriminant analysis (LDA) is a supervised method for classification and recognition applications [7] − [13]. This is achieved by maximizing the ratio of the magnitude of between − class scattering matrix to the magnitude of within − class scattering matrix. Within − class scattering matrix is

definedas $S_W = \sum_{i=1}^{C}\sum_{x\in\{c_i\}}(x-m_i)(x-m_i)^T$

(9)

Where **C** is the number of classes, $\mathbf{C_i}$ is a set of data within the ith class, and m_i is the mean of the ith class. The within class scatter matrix represents the degree of scatter within classes as a summation of covariance matrices of each class. A total scatter matrix $\mathbf{S_T}$ and a total mean **m** are defined as

$$S_T = \sum_x (x - m)(x - m)^T \tag{10}$$

and

$$m = \frac{1}{n} \sum_x x = \frac{1}{n} \sum_{i=1}^{C} n_i m_i \tag{11}$$

Where n is the number of total samples and n_i is the number of samples within the ith class. Then we get

$$S_T = S_W + \sum_{i=1}^{C} n_i (m_i - m)(m_i - m)^T \tag{12}$$

The second term in the above equation is defined as a between – class scatter matrix S_B, so that the total scattering matrix is the sum of the within – class scatter matrix and the between – class scatter matrix.

$$S_B = \sum_{i=1}^{C} n_i (m_i - m)(m_i - m)^T \tag{13}$$

and

$$S_T = S_W + S_B \tag{14}$$

The between – class scatter matrix represents the degree of scatter between classes as a covariance matrix of means of each class. The projection of d – dimensional input samples onto r – dimensional space (r $<<$ d) is done by $y = W^T x$ Transformation matrix W is obtained in such a way that it maximizes the criterion function J(W) given as

$$J(W) = \arg \max_W \left| \frac{W^T S_B W}{W^T S_W W} \right| \tag{15}$$

The columns of optimal W are the generalized eigen vectors w_i that correspond to the largest eigen values in $S_B W_i = \lambda_i S_W W_i$

4. PROBABILISTIC NEURAL NETWORK

Probabilistic Neural Network is a type of Radial Basis Function (RBF) network, which is suitable for pattern classification. The basic structure of a probabilistic neural network is shown in figures 2 and 3. The fundamental architecture has three layers, an input layer, a pattern layer, and an output layer. The pattern layer constitutes a neural implementation of a Bayes classifier, where the class dependent Probability Density Functions (PDF) are approximated using a Parzen estimator[14]-[18]. Parzen estimator determines the PDF by minimizing the expected risk in classifying the training set incorrectly. Using the Parzen estimator, the classification gets closer to the true underlying class density functions as the number of training samples increases.

The pattern layer consists of a processing element corresponding to each input vector in the training set. Each output class should consist of equal number of processing elements. The pattern layer classifies the input vectors based on competition, where only the highest match to an input vector wins and generates an output. Hence only one classification category is generated for any given input vector. If there is no relation between input patterns and the patterns programmed into the pattern layer, then no output is generated. In order to obtain more generalization a smoothing factor is included while training the network.
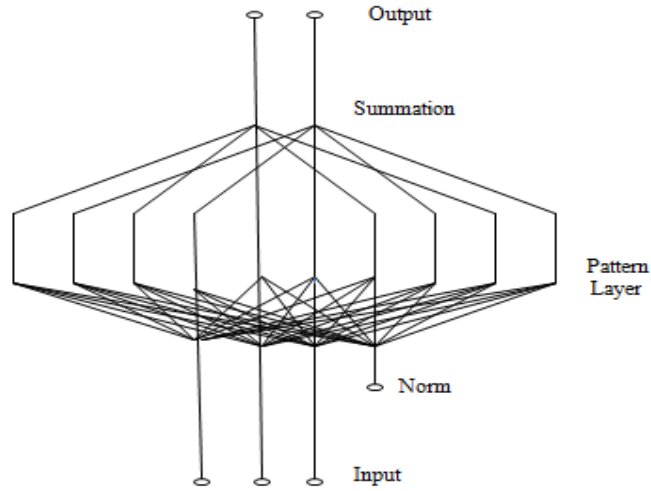
Figure 2. Probabilistic Neural Network

Compared to the feed forward back propagation network, training of the probabilistic neural network is much simpler. Since the probabilistic networks classify on the basis of Bayesian theory, it is essential to classify the input vectors into one of the two classes in a Bayesian optimal manner. The Bayes rule classifies an input vector belonging to class A as,

$$P_A C_A f_A(x) > P_B C_B f_B(x) \qquad (14)$$

Where,

P_A - Priori probability of occurrence of patterns in class A

C_A - Cost associated with classifying vectors

$f_A(x)$ - Probability density function of class A

The PDF estimated using the Bayesian theory should be positive and integratable over all x and the result must be 1. The probabilistic neural net uses the following equation to estimate the probability density function given by,

$$f_A(x) = \frac{1}{(2\Pi)^{n/2} \sigma^n} \frac{1}{m_n} \sum_{i=1}^{m_A} \exp\left[-2 \frac{(x - x_A)^r (x - x_{Ai})}{\sigma^2} \right] \qquad (15)$$

Where

x_{Ai} - ith training pattern from class A

n - Dimension of the input vectors

σ - Smoothing parameter (corresponds to standard deviations   of Guassian distribution)

The function $f_A(x)$ acts as an estimator as long as the parent density is smooth and continuous. $f_A(x)$ approaches the parent density function as the number of data points used for the estimation increases. The function $f_A(x)$ is a sum of Guassian distributions.
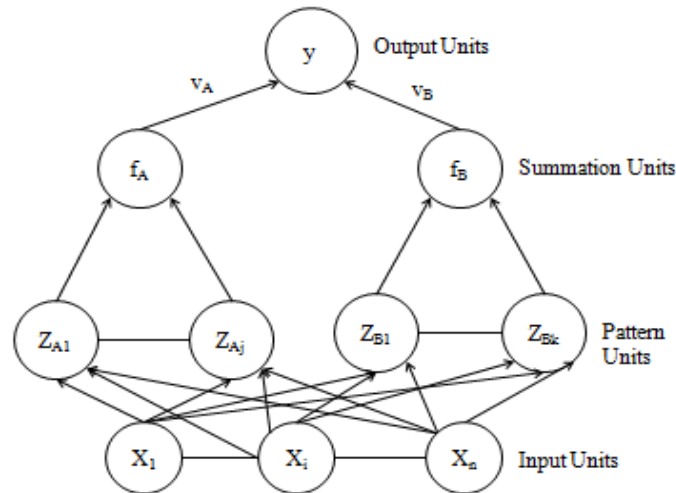
Fig – 3 Architecture of Probabilistic Neural Networks

5.   NUMERICAL RESULTS

5.1.   Simulation Results – 1

The Proposed new face recognition method was applied on two different databases. The first database is AT&T (ORL) face database, containing 40 subjects and each subject having 10 images of size 112×92 and the second is student's database consists of thirteen students face images each with 5 images of different facial expressions, poses and background conditions. For simulations and proposed method evaluation Matlab is used on a PC with Intel(R) core (TM) 2 Duo CPU and 2 GB RAM. The obtained results for tow databases are given separately as follows.

5.1.1.   AT&T (ORL) Face DATABASE

ORL Face database is shown in figure 4 it consists of 40 subjects and each subject having 10 images of size 112×92 with different facial expressions, poses and background conditions. Before doing simulation size of the face images are reduced to 9×9 using matlab. These reduced size images are used as inputs to the Tchebichef polynomials and it gives 81 features per image as output. For this database the graph drawn between Image size vs recognition rate is shown in figure 5 and the graph drawn between number of features per sample (Sample dimension) vs recognition rate is shown in figure 6. From these figures it is clear that Image size of 9×9 with sample dimension 35 is giving maximum recognition rate of 100%. The test face image and recognized images are shown in figures 11(a) and 11(b).
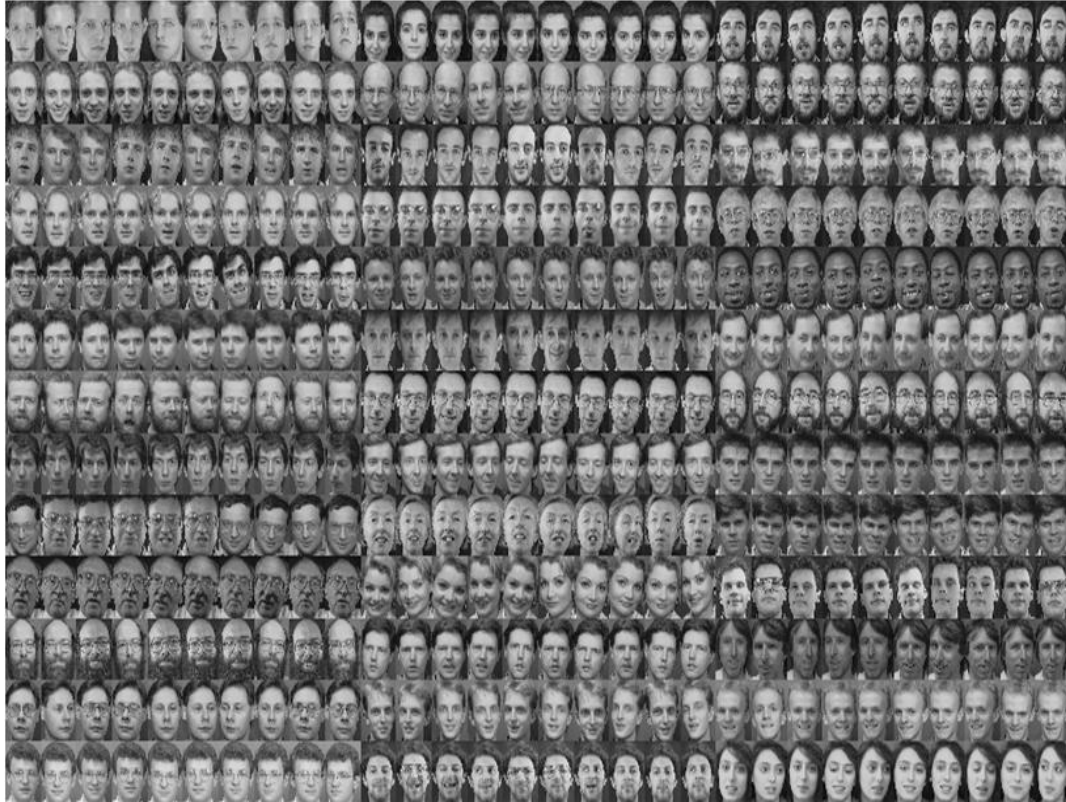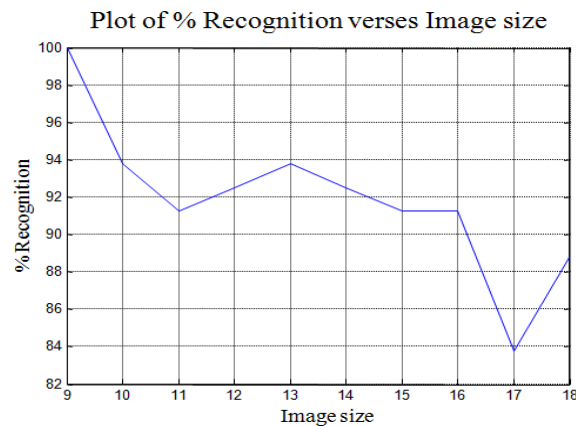
Figure 4. ORL Face Databse

Figure 5. Plot of Recognition Rate verses image size

The feature vectors of the face images obtained at the output of the Tchebichef polynimials are given to the LDA as input. LDA produces 39 most discriminate features per image that leads better classification. The discriminant features of the face images produced by the LDA are given to the Probabilistic Neural Network for classification. Probabilistic Neural network (PNN) gives fast and accurate classification of face images. In simulation three different combinations of training and test image samples are used as

i)   280 training images and 120 test images
ii)   320 training images and 80 test images
iii)   360 training images and 40 test images

Simulation was done with the above three sets of data and the obtained results are summarized in the tabular form as shown in table 1.
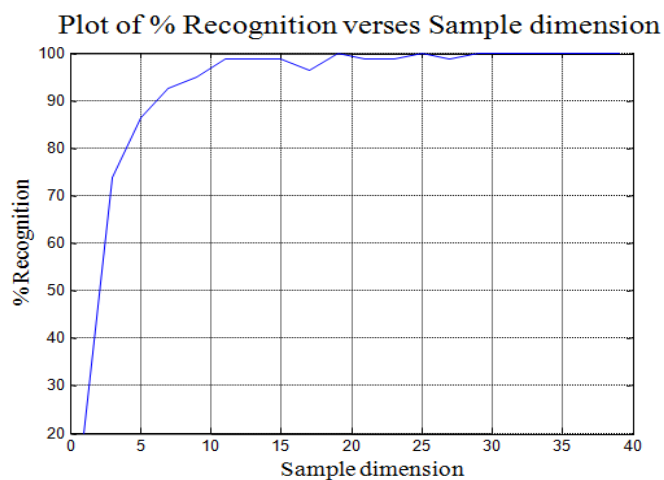

Figure 6. Plot of Recognition Rate versed Sample dimension

No of features per image used in Tchebichef polynomials as input = 81
No of features per image used in LDA as input = 81
No of features per image used for classification = 39

Figure 7(a). Test Face   Figure 7(b). Recognized Face

For 320 training samples, the average training time taken is 6.2590 sec. For 80 test samples the average testing time taken is 12.42 sec that means 0.1553 sec time is taken for the testing of one sample. This is very small classification time when compare to any other Neural Network classification time.

Table – 1 Recognition Rate for different Training and Test samples for AT&T (ORL) database

No of classes	No of Training Samples	Samples per class	No of Test Samples	Samples per class	Recognition Rate (Max)
40	280	7	120	3	97.5
40	320	8	80	2	100
40	360	9	40	1	100

5.1.2. Students Database

This database consists of 14 classes and each class consists of 5 samples. One sample of each class is shown in figure 8. Before doing the experiment the size of the face images are reduced to 4×4 using matlab. These reduced size images are used as inputs to the discrete Techebichef polynomials and it gives 16 features per image as output. But experimental results shows that 14 features per face image gives maximum recognition hence only 14 features per image are used for further use. The graph drawn between Image size vs recognition rate is shown in figure 9 and the graph drawn between number of features per sample (Sample dimension) vs recognition rate is shown in figure 10.

From these figures it is clear that Image size of 4×4 with sample dimension 9 is giving maximum recognition rate of 100%. The test face image and recognized images are shown in figures 11(a) and 11(b).

Figure 8. One sample from each Class of the Student database

The feature vectors of the face images obtained at the output of the discrete Tchebuchef polynomials are given to the LDA as input. LDA produces 13 most discriminate features per image that leads better classification. The discriminant features of the face images produced by the LDA are given to the Probabilistic Neural Network for classification. In simulations two different combinations of training and test image samples are used as

   i)   56 training images and 14 test images
   ii)   42 training images and 28 test images

Simulation was performed on the above two sets of data and the obtained results are summarized in the tabular form as shown in table 2.

No of features per image used in Tchebichef polynomials as input = 16
No of features per image used in LDA as input = 14
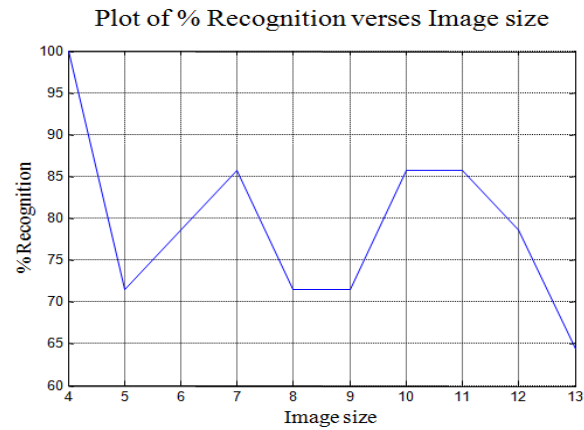No of features per image used for classification = 9

Plot of % Recognition verses Image size

Figure 9. Plot of Recognition Rate verses image size
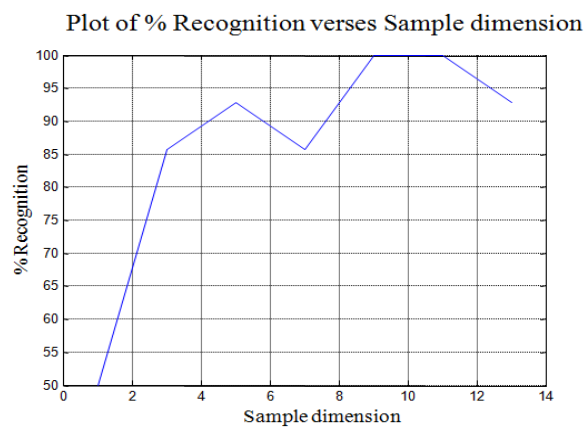
Plot of % Recognition verses Sample dimension

Figure 10. Plot of Recognition Rate verses Sample dimension

Table – 2 Recognition Rate for different Training and Test samples for Student database

NO of Classes	No of Training Samples	Samples per class	No of Test Smples	Samples per class	Recognition Rate (MAX)
14	42	3	28	2	89.2857 (25 out of 28)
14	56	4	14	1	100

Fig – 11(a) Test Face   Fig – 11(b) Recognized Face

For 52 training samples, the average training time taken is 1.666 sec. For 13 test samples the average testing time taken is 3.524 sec that means 0.2514 sec time is taken for the testing of one sample. This is very small classification time when compare to any other Neural Network classification time.

6.   CONCLUSION

In this paper, a new Face recognition method is presented. This new method is a combination of Discrete Tchebichef moments, Linear Discriminant Analysis and Probabilistic Neural Network. By using these algorithms an efficient face recognition method was constructed with maximum recognition rate of 100%. Simulation results using AT & T (ORL) face database and student's database demonstrated the ability of the proposed method for optimal feature extraction and efficient face classification. The new face recognition algorithm can be used in many applications such as security systems.

The ability of our proposed face recognition method is demonstrated on the basis of obtained results on AT & T face database and student's database. For generalization, the proposed method should achieve 100% Recognition rate on other face databases and also on other combinations of training and test samples.

REFERENCES

[1]   A. Jain, R. Bolle, S. Pankanti Eds, "BIOMETRIC – Personal Identification in Networked Society".*KluwerAcademicPublishers, Boston/Dordrecht/London*,1999.
[2]   A.K. Jain, R.P.W. Duin, J. Mao, Statistical pattern recognition: a review, *IEEE Trans. Pattern Anal. Mach.Iintell*. 22 (2000) 4 – 37.
[3]   K. Fukunaga, Introduction to Statistical pattern recognition, II nd Eddition. *Academic Press, New York*, 1990.
[4]   R. O. Duda, P.E. Hart, Pattern Classification and Scene Analysis, *Wiley, New York*, 1973.
[5]   Nur Azman Abu, Siaw Lang Wong, Hidayah Rahmalan, Shahrin Sahid, Fast abd Effective 4×4TchbichefMomentImage Compression.*Majlesi Journal of Electrical Engineering Vol. 4, No. 3, September* 2010.
[6]   J. P. Ananth, Dr V. Subbiah Bharathi, Face Image Retrieval using Tchbichef Moments, *Indian Journal of Computer Science and Engineering,* Vol. 2 No.6 Dec 2011-jan 2012.
[7]   N. Kwak, and C.H. Choi, "Input Feature Selection for classification problem", *IEEE Transactions on Neural Networks*, 13(1), 143-159,2002.
[8]   W. zhao, R. Chellappa, P. J. Phillips, Subspace linear discriminant Analysis for face recognition, *Technical Report* CAR – TR -914, *Center for Automation Research, University of Maryland, College Park*,1999.
[9]   H. Jain, P.C. Yuen, C.Wen-Sheng, *A novel subspace LDA Algorithm for recognition of face   images with illumination and pose variation,* Proceedings of the International Conference on Mission Learning and Cibernetics, Vol. 6, 2004,pp. 3589-3594.
[10] W. Xiaogang, T. Xiaoou, A unified framework for subspace face recognition*, IEEE Trans. Pattern Anal. Math Intell.* 26 (9) (2004) 1222 – 1228.
[11] Alok Sharma, Kuldip K. Paliwal, Godfrey C. Onwubolu, Ckass – dependent PCA, MDC AND LDA: A combined classifier fir pattern classification, 2006 *Pattern Recognition Society. Published by Elsevier Ltd.* Doi:10,1016/j.patcog.2006.02.001.
[12] Lu, K, Plataniotis, K.N. & Venetsanopoulos, A.N. (2003) Face Recognition Using LDA – based algorithm, *IEEE trans. Neural Network*, vol. 14, No 1, pp. 195- 199, January 2003
[14] M. Er, S. Wu, J. Lu, L. H. Toh. " Face recognition with Radial Basis Function(RBF) Neural Networks", IEEE Trans. Neural Network, Vol. 13, No. 3,pp 697-710.
[15] D.F. Specht, "Probabilistic Neural Networks for classification, mapping, or associative memory*", Proceedings of IEEE International Conference on Neural Networks*, Vol 1, IEEE Press, New York, pp. 525-532, june 1988.
[16] D.F. Specht, *"Probabilistic Neural Networks" Neural Networks, Vol. 3*, No 1,pp.109-118,1990.
[17] W. zhao, R. Chellappa, N. Namdhakumar, Emperical performance analysis of linear discriminant classifiers, *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognitio*, 1998,pp, 164-169.
[18] Mohd Fauzi Othman and Mohd Ariffanan Mohd Basri, "Probabilistic Neural Network For Brain Tumor Classification", *Second International Conference on Intelligent Systems, Modeling and Simulation,* IEEE *Computer Society*, 2011, DOI 10.1109/ISMS. 2011.32.

BIBLIOGRAPHY OF AUTHORS

D. Sridhar He is currently persuing his Ph.D. degree from Jawaharlal Nehru Technological University, Hyderabad, India. His research interests include pattern recognition, neural networks, image processing, computer vision, machine learning and their applications.

Dr I.V. Murali Krishna received the M.Tech degree from IIT, Madras and P.hD degree from IISC, Bangalore, India. He was former Scientist NRSA, ISRO, India. Former Director R & D and Center for Atmospheric Science and Weather Modification Technologies, JNTU, Hyderabad, India. His research interests include remote sensing, image processing and pattern recognition and their applications.