Hybrid Feature Extraction Technique for Face Recognition

Sangeeta N. Kakarwal

Department of Computer Science and Engineering

P.E.S. College of Engineering

Aurangabad, India

Ratnadeep R. Deshmukh

Department of Computer Science and IT

Dr. Babasaheb Ambedkar Marathwada University

Aurangabad, India

Abstract— **This paper presents novel technique for recognizing faces. The proposed method uses hybrid feature extraction techniques such as Chi square and entropy are combined together. Feed forward and self-organizing neural network are used for classification. We evaluate proposed method using FACE94 and ORL database and achieved better performance.**

Keywords-Biometric; Chi square test; Entropy; FFNN; SOM.

I. INTRODUCTION

Face recognition from still images and video sequence has been an active research area due to both its scientific challenges and wide range of potential applications such as biometric identity authentication, human-computer interaction, and video surveillance. Within the past two decades, numerous face recognition algorithms have been proposed as reviewed in the literature survey. Even though we human beings can detect and identify faces in a cluttered scene with little effort, building an automated system that accomplishes such objective is very challenging. The challenges mainly come from the large variations in the visual stimulus due to illumination conditions, viewing directions, facial expressions, aging, and disguises such as facial hair, glasses, or cosmetics [1].

Face Recognition focuses on recognizing the identity of a person from a database of known individuals. Face Recognition will find countless unobtrusive applications such as airport security and access control, building surveillance and monitoring Human-Computer Intelligent interaction and perceptual interfaces and Smart Environments at home, office and cars [2].

Within the last decade, face recognition (FR) has found a wide range of applications, from identity authentication, access control, and face-based video indexing/ browsing; to human-computer interaction. Two issues are central to all these algorithms: 1) feature selection for face representation and 2) classification of a new face image based on the chosen feature representation. This work focuses on the issue of feature selection. Among various solutions to the problem, the most successful are those appearance-based approaches, which generally operate directly on images or appearances of face objects and process the images as two-dimensional (2-D) holistic patterns, to avoid difficulties associated with three-dimensional (3-D) modeling, and shape or landmark detection [3]. The initial idea and early work of this research have been published in part as conference papers in [4], [5] and [6].

A recognition process involves a suitable representation, which should make the subsequent processing not only computationally feasible but also robust to certain variations in images. One method of face representation attempts to capture and define the face as a whole and exploit the statistical regularities of pixel intensity variations [7].

The remaining part of this paper is organized as follows. Section II extends to the pattern matching which also introduces and discusses the Chi square test, Entropy and FFNN and SOM in detail. In Section III, extensive experiments on FACE94 and ORL faces are conducted to evaluate the performance of the proposed method on face recognition. Finally, conclusions are drawn in Section IV with some discussions.

II. PATTERN MATCHING

A. Pattern Recognition Methods

During the past 30 years, pattern recognition has had a considerable growth. Applications of pattern recognition now include: character recognition; target detection; medical diagnosis; biomedical signal and image analysis; remote sensing; identification of human faces and of fingerprints; machine part recognition; automatic inspection; and many others.

Traditionally, Pattern recognition methods are grouped into two categories: structural methods and feature space methods. Structural methods are useful in situation where the different classes of entity can be distinguished from each other by structural information, e.g. in character recognition different letters of the alphabet are structurally different from each other. The earliest-developed structural methods were the syntactic methods, based on using formal grammars to describe the structure of an entity [8].

The traditional approach to feature-space pattern recognition is the statistical approach, where the boundaries between the regions representing pattern classes in feature space are found by statistical inference based on a design set of sample patterns of known class membership [8]. Feature-space methods are useful in situations where the distinction between different pattern classes is readily expressible in terms of numerical measurements of this kind. The traditional goal of feature extraction is to characterize the object to be recognized by measurements whose values are very similar for objects in

the same category, and very different for objects in different categories. This leads to the idea of seeking distinguishing features that are invariant to irrelevant transformations of the input. The task of the classifier component proper of a full system is to use the feature vector provided by the feature extractor to assign the object to a category [9]. Image classification is implemented by computing the similarity score between a target discriminating feature vector and a query discriminating feature vector [10].

B. Chi Square Test

Chi-square is a non-parametric test of statistical significance for analysis. Any appropriately performed test of statistical significance lets you know the degree of confidence you can have in accepting or rejecting a hypothesis. Typically, the hypothesis tested with Chi Square is whether or not two different samples (of people, texts, whatever) are different enough in some characteristic or aspect of their behavior that we can generalize from our samples that the population from which our samples are drawn are also different in the behavior or characteristics.

On the basis of hypothesis assumed about the population, we find the expected frequencies Ei (I =1,2,…,n), corresponding to the observed frequencies Oi(i=1,2,…,n) such that $\sum$ Ei = $\sum$ Oi. It is known that

$$\chi^2 = \sum_{i=1}^{n} \frac{(Oi - Ei)^2}{Ei}$$

follows approximately a χ^2 - distribution with degrees of freedom equal to the number of independent frequencies. To test the goodness of fit, we have to determine how far the difference between Oi and Ei can be attributed to fluctuations of sampling and when we can assert that the differences are large enough to conclude that the sample is not a simple sample from the hypothetical population[11][12].

C. Entropy

The entropy is equivalent (i.e., monotonically functionally related) to the average minimal probability of decision error and is related to randomness extraction. For a given fuzzy sketch construction, the objective is then to derive a lower bound on the min entropy of the biometric template when conditioned on a given sketch, which itself yields an upper bound on the decrease in the security level measured as the min-entropy loss, which is defined as the difference between the unconditional and conditional min entropies [13] Shannon gave a precise mathematical definition of the average amount of information conveyed per source symbol, which is termed as Entropy [14].

Consider two random variables and having some joint probability distribution over a finite set. The unconditional uncertainty of can be measured by different entropies, the most famous of which is the Shannon entropy. Some of them have been given practical interpretations, e.g., the Shannon entropy can be interpreted in terms of coding and the min entropy in terms of decision making and classification [15]

Entropy is a statistical measure that summarizes randomness. Given a discrete random variable, its entropy is defined by

$$H(X) = -E_x[\log P(X)]$$

$$= - \sum_{xi \in \Omega x} P(X = xi) \log(X = xi) \qquad \text{…(1)}$$

Where Ωx is the sample space and xi is the member of it. P(X=xi) represents the probability when X takes on the value xi. We can see in (1) that the more random a variable is, the more entropy it will have.

D. Artificial Neural Network

In recent years, there has been an increase in the use of evolutionary approaches in the training of artificial neural networks (ANNs). While evolutionary techniques for neural networks have shown to provide superior performance over conventional training approaches, the simultaneous optimization of network performance and architecture will almost always result in a slow training process due to the added algorithmic complexity [16].

1) Feed Forward Network

Feed forward networks may have a single layer of weights where the inputs are directly connected to the output, or multiple layers with intervening sets of hidden units. Neural networks use hidden units to create internal representations of the input patterns [17].

A Feed forward artificial neural network consists of layers of processing units, each layer feeding input to the next layer in a Feed forward manner through a set of connection weights or strengths. The weights are adjusted using the back propagation learning law. The patterns have to be applied for several training cycles to obtain the output error to an acceptable low value.

The back propagation learning involves propagation of the error backwards from the input training pattern, is determined by computing the outputs of units for each hidden layer in the forward pass of the input data. The error in the output is propagated backwards only to determine the weight updates [18]. FFNN is a multilayer Neural Network, which uses back propagation for learning.

As in most ANN applications, the number of nodes in the hidden layer has a direct effect on the quality of the solution. ANNs are first trained with a relatively small value for hidden nodes, which is later increased if the error is not reduced to acceptable levels. Large values for hidden nodes are avoided since they significantly increase computation time [19].

The Back propagation neural network is also called as generalized delta rule. The application of generalized delta rule at any iterative step involves two basic phases. In the first phase, a training vector is presented to the network and is allowed to propagate through the layers to compute output for each node. The output of the nodes in the output layers is then compared against their desired responses to generate error term. The second phase involves a backward pass through a network during which the appropriate error signal is passed to each node and the corresponding weight changes are made. Common practice is to track network error, as well as errors associated with individual patterns. In a successful training session, the network error decreases with the number of

iterations and the procedure converges to a stable set of weights that exhibit only small fluctuations with additional training. The approach followed to establish whether a pattern has been classified correctly during training is to determine whether the response of the node in the output layer associated with the pattern class from which the pattern was obtained is high, while all the other nodes have outputs that are low [20].

Backpropogation is one of the supervised learning neural networks. Supervised learning is the process of providing the network with a series of sample inputs and comparing the output with the expected responses. The learning continues until the network is able to provide the expected response. The learning is considered complete when the neural network reaches a user defined performance level. This level signifies that the network has achieved the desired accuracy as it produces the required outputs for a given sequence of inputs [21].

2) Self Organizing Map

The self-organizing map, developed by Kohonen, groups the input data into cluster which are, commonly used for unsupervised training. In case of unsupervised learning, the target output is not known [17].

In a self-organizing map, the neurons are placed at the nodes of a lattice that is usually one or two dimensional. Higher dimensional maps are also possible but not as common. The neurons become selectively tuned to various input patterns or classes of input patterns in the course of a competitive learning process. The locations of the neurons so tuned (i.e., the wining neurons) become ordered with respect to each other in such a way that a meaningful coordinate system for different input features is created over the lattice. A self-organizing map is therefore characterized by the formation of a topographic map of the input patterns in which the spatial locations of the neurons in the lattice are indicative of intrinsic statistical features contained in the input patterns, hence the name "self-organizing map"[22]. The algorithm of self-organizing map is given below:

Algorithm SelfOrganize;
- Select network topology;
- Initialize weights randomly; and select D(0)>0;
- While computational bounds are not exceeded, do
  1. Select an input sample i_l;
  2. Find the output node j* with minimum $\sum_{k=1}^{n}$ (i$_{l,k}$(t)-w$_{j,k}$(t))^2;
  3. Update weights to all nodes within a topological distance of D(t) from j*, using
         w$_j$(t+1)= w$_j$(t) +η(t)(i$_l$(t)-w$_j$(t)),
     where 0< η(t)≤ η(t-1)≤1;
  4. Increment t;
  End while.

Figure 1.   Algorithm of Self Organizing Map

III. EXPERIMENTAL RESULTS AND DISCUSSION

In order to assess the efficiency of proposed methodology which is discussed above, we performed experiments over Face94 and ORL dataset using FFNN and SOM neural network as a classifier.

A. Face94 Dataset

Face94 dataset consist of 20 female and 113 male face images having 20 distinct subject containing variations in illumination and facial expression. From these dataset we have selected 20 individuals consisting of males as well as females [23].

Face94 dataset used in our experiments includes 250 face images corresponding to 20 different subjects. For each individual we have selected 15 images for training and 5 images for testing.

Figure 2.   Some Face Images from FACE94 Database

B. ORL

The Olivetti Research Lab (ORL) Database [4] of face images provided by the AT&T Laboratories from Cambridge University has been used for the experiment. It was collected between 1992 and 1994. It contains slight variations in illumination, facial expression (open/closed eyes, smiling/not smiling) and facial details (glasses/no glasses). It is of 400 images, corresponding to 40 subjects (namely, 10 images for each class). Each image has the size of 112 x 92 pixels with 256 gray levels. Some face images from the ORL database are shown in figure3

For both database, we selected 50 images for testing genuine as well imposter faces. To extract the facial region, the images are normalized. All images are gray-scale images.

Figure 3.   Some Face images from ORL Database

C. Steps used in Face Recognition

- Read input image, convert it into gray scale image then resize it to 200x180 pixels.

- Divide image into 4x4 blocks of 50x45 pixels.

- Obtain hybrid features from face by combining values of Chi Square test and Entropy together.

- Classify the images by Feed forward neural network and Self organizing map neural network.

- Analyse the performance by computing FAR and FRR.

D. Performance Evaluation

The accuracy of biometric-like identity authentication is due to the genuine and imposter distribution of matching. The overall accuracy can be illustrated by False Reject Rate (FRR) and False Accept Rate (FAR) at all thresholds. When the parameter changes, FAR and FRR may yield the same value, which is called Equal Error Rate (EER). It is a very important indicator to evaluate the accuracy of the biometric system, as well as binding of biometric and user data [25].

A typical biometric verification system commits two types of errors: false match and false non-match. Note that these two types of errors are also often denoted as false acceptance and false rejection; a distinction has to be made between positive and negative recognition; in positive recognition systems (e.g., an access control system) a false match determines the false acceptance of an impostor, whereas a false non-match causes the false rejection of a genuine user. On the other hand, in a negative recognition application (e.g., preventing users from obtaining welfare benefits under false identities), a false match results in rejecting a genuine request, whereas a false non-match results in falsely accepting an impostor attempt.

The notation "false match/false non-match" is not application dependent and therefore, in principle, is preferable to "false acceptance/false rejection." However, the use of false acceptance rate (FAR) and false rejection rate (FRR) is more popular and largely used in the commercial environment [26].

Traditional methods of evaluation focus on collective error statistics such as EERs and ROC curves. These statistics are useful for evaluating systems as a whole. Equal-Error Rate (EER) denotes the error rate at the threshold t for which false match rate and false non-match rate are identical: FAR(t) = FRR(t) [27].

FAR and FRR values for all persons with different threshold values. The FRR and FAR for number of participants (N) are calculated as specified in Eq. (2) and in equation Eq. (3) [28]:

$$FRR = \frac{1}{N}\sum_{n=1}^{N} FRR\ (n) \qquad \dots (2)$$

$$FAR = \frac{1}{N}\sum_{n=1}^{N} FAR\ (n) \qquad \dots (3)$$

When the experiment was carried out on ORL database 96% result is obtained with FFNN. In case of FACE94 database, the result obtained with SOM is 94%. Table1 and Table2 give the performance of hybrid feature extraction technique for FFNN and SOM respectively.

In addition to this experimentation was also carried out to recognize impostor faces. Graph1 and Graph2 illustrate the result of genuine and impostor face recognition.

CONCLUSION

This paper investigates the feasibility and effectiveness of face recognition with Chi square test and Entropy. Face recognition based on Chi square test and Entropy is performed by supervised and unsupervised network. Experimental results on Face94 and ORL database demonstrate that the proposed methodology outperforms in recognition.

TABLE I.   PERFORMANCE OF FACE RECOGNITION FOR CHI SQUARE TEST+ENTROPY AND FFNN

Face datab ase	No. of test faces recognized	Rate of recognit ion	FRR	No. of impostor faces recognized	Rate of recognit ion	FAR
FACE 94	46	92	0.08	39	78	0.22
ORL	48	96	0.04	26	52	0.48

TABLE II.   PERFORMANCE OF FACE RECOGNITION FOR CHI SQUARE TEST+ ENTROPY AND SOM

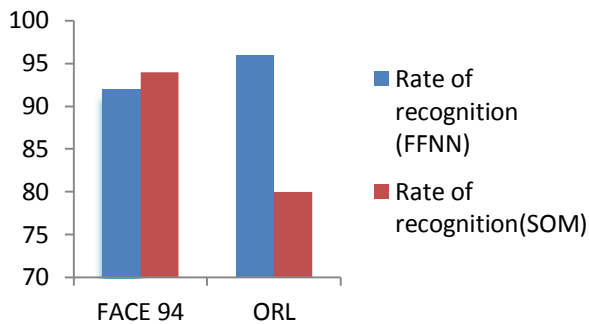| Face datab ase	No. of test faces recognized	Rate of recognitio n	FRR	No. of test impostor faces recognized	Rate of recognitio n	FAR
FAC E 94	47	94	0.06	35	70	0.3
ORL	40	80	0.2	26	52	0.48

Figure 4.   Graph1: Performance of Genuine faces using Chi Square+Entropy for FFNN and SOM
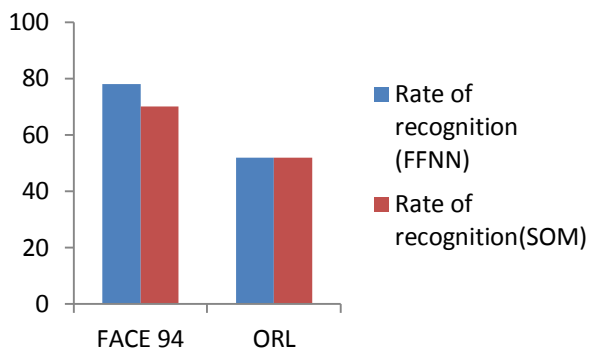

Figure 5.   Graph2: Performance of Imposter faces using Chi Square+Entropy for FFNN and SOM

REFERENCES

[1] Yu Su, Shiguang Shan, Xilin Chen, and Wen Gao, "Hierarchical Ensemble of Global and local Classifiers for   Face Recognition", IEEE Transactions On Image Processing,Vol. 18, No. 8, August 2009, pp 1885 – 1896.

[2] Chengjun Liu, Harry Wechsler," Independent Componenet Analysis of Gabor Features for Face Recognition", IEEE Transactions on Neural Networks, Vol. 14, July 2003, pp 919-928.

[3] Juwei Lu, Konstantinos N. Plataniotis, Anastasios N. Venetsanopoulos: Face Recognition Using Kernel Direct Discriminant Analysis Algorithms, IEEE Transactions On Neural Networks, Vol. 14, No. 1, January 2003, pp 117-126.

[4] S.N. Kakarwal, S.D. Sapkal, P.J.Ahire, Dr. D.S. Bormane: Analysis of Facial Image   Classification using Discrete Wavelet Transform, Proc. International Conference ICSCI, 2007,pp 700-705.

[5] S.N. Kakarwal, Mahananda Malkauthekar, Shubhangi Sapkal, Dr. Ratnadeep   Deshmukh: Face Recognition using Fourier Descriptor and FFNN, Proc. IEEE   International Advance Computing Conference, 2009, pp 2740-2742.

[6] S.N. Kakarwal, Dr. R.R. Deshmukh: Wavelet Transform   based Feature Extraction for Face Recognition, pp 100-104, IJCSA Issue-I June 2010, ISSN 0974-0767

[7] Bai-Ling Zhang, Haihong Zhang and Shuzhi Sam Je: Face Recognition by Applying Subband Representation and Kernel Associative Memory, IEEE Transaction on Neural Networks, Vol. 15, Jan 2004, pp 166-177.

[8] Daisheng Luo: Pattern Recognition and Image Processing (Horwood Publishing Limited   1998), pp 2-3

[9] Richard O. Duda, Peter E. Hart, David G. Stork: Pattern Classification (John Wiley 2001),   pp 11-12

[10] Chengjun Liu: Learning the Uncorrelated, Independent, and Discriminating Color Spaces for Face Recognition, IEEE Transactions On Information Forensics and Security, Vol. 3, No. 2, June 2008, pp 213-222.

[11] T. Veerarajan, "Probability, Statistics and Random Processes", TMH, 2003,pp. 311-312.

[12] Chuang, K.-S.; Huang, H.K., Comparison of Chi- Square and Join-Count Methods for Evaluating Digital   Image Data, IEEE Transaction on Medical Imaging, Vol. 11, No. 1, March 1992,   33.

[13] Su Hongtao, David Dagan Feng, Zhao Rong-chun, Wang Xiu-ying, "Face Recognition Method Using Mutual Information and Hybrid Feature", 0-7695-   1957-1/03 © 2003 IEEE.

[14] Richard Wells, Applied Coding and Information Theory for engineers, Pearson Education, pp. 10

[15] Jovan Dj. Golic´ and Madalina Baltatu, "Entropy Analysis and New Constructions of Biometric Key Generation Systems", IEEE Transactions On Information Theory, Vol. 54, No. 5, May 2008, pp.2026-2040

[16] Chi-Keong Goh, Eu-Jin Teoh, and Kay Chen Tan, Hybrid Multiobjective Evolutionary Design for Artificial Neural Networks, IEEE Transactions on Neural Networks, Vol. 19, no. 9, September 2008

[17] S.N. Sivanandanam, S. Sumathi, S. N. Deepa, Introduction to Neural Networks using MATLAB 6.0, TMH, pp 20

[18] B. Yegnanarayana, Artificial Neural Networks, PHI, pp. 117-120.

[19] Kevin Stanley McFall and James Robert Mahan, Artificial Neural Network Method for Solution of Boundary Value Problems with Exact Satisfaction of Arbitrary Boundary Conditions, IEEE Transactions on Neural Networks, Vol. 20, No. 8, pp. 1221-1233, August 2009

[20] Rafel C. Gonzalez, Richard E. Woods, Digital Image Processing, pp. 896

[21] S. Jayaraman, E. Esakkirajan, T. Veerakumar,   Digital Image Processing, pp. 425

[22] Simon Haykin, Neural Networks, LPE, pp. 465-466

[23] Dr Libor Spacek Computer Vision Science   Research Projects, Face94Dataset http://dces.essex.ac.uk/mv/allfaces /faces94.zip

[24] The   ORL   Database   of   Faces,   Available: http://www.uk.research.att.com/facedatabase.html

[25] Stan Z. Li, Anil K. Jain, Encyclopedia of Biometrics, Springer, pp 75

[26] Davide Maltoni, Dario Maio, Anil K. Jain, Salil Prabhakar: Handbook of Fingerprint Recognition (Springer),pp 3

[27] Neil Yager and Ted Dunstone: The Biometric Menagerie, IEEE Transactions on Pattern Analysis and Machine Inte-ligence, VOL. 32, No. 2, FEBRUARY 2010, pp 220-230

[28] Website:http://www.bromba.com/faq/biofaqe