Face Recognition Using Fuzzy Neural Network Classifier

Dhanya S. Pankaj and M. Wilscy

Department Of Computer Science, University Of Kerala
Kariavattom, Thiruvananthapuram-695581, Kerala, India
{dhanyaspankaj,wilsyphilipose}@gmail.com

Abstract. Given a still image or a video, a face recognition application identifies or verifies face images using a stored database of faces. In this paper a method for face recognition using a fuzzy neural network classifier based on the Integrated Adaptive Fuzzy Clustering (IAFC) method has been proposed. IAFC forms the cluster boundaries by a combined similarity measure and by integrating the advantages of the fuzzy c-means (FCM), the adaptive resonance theory, and a fuzzified Kohonen-type learning rule. The proposed system achieved a recognition rate of 98.75% and 99.39% for the AT & T and Yale databases respectively, which is better compared to the Back Propagation Neural Network (BPNN) system. Considering the rejection rate for the non-registrants, the system achieved an equal error rate of 3.7 % and 1.3% for the AT & T and the Yale databases respectively which is better compared to most of the existing systems.

Keywords: Face Recognition, PCA, LDA, Integrated Adaptive Fuzzy Clustering, Fuzzy Neural Network.

1 Introduction

Face Recognition 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 [1]. Given a still image or a video, a face recognition application identifies or verifies one or more persons in the scene using a stored database of faces.

Various appearance-based approaches to face recognition have been proposed in literature [2]. The appearance based methods extract features that optimally represent the faces belonging to a class and separate faces from different classes [3]. A class of face recognition algorithms employ various classifiers such as probabilistic [4], hidden Markov models (HMMs) [5], neural networks (NNs) [6], and support vector machine (SVM) [7] and feature extraction methods like Principal Component Analysis (PCA)[16], Linear Discriminant Analysis (LDA)[8], Discrete Cosine Transform (DCT)[9].

Most of the face recognition algorithms are designed as classifiers. The automatic classification of human faces is still a challenging problem due to two main factors - large intra-subject variations due to change in pose, illumination, facial expression, aging, occlusion etc and small inter-subject variations due to the similarity of

D. Nagamalai, E. Renault, and M. Dhanushkodi (Eds.): PDCTA 2011, CCIS 203, pp. 53–62, 2011.

individual appearances. Hence, forming well defined class boundaries is a major challenge for the existing face recognition algorithms based on classifiers.

A number of neuro-fuzzy clustering algorithms have been proposed in literature [11-15]. However, all of these algorithms developed suffer from restrictions in identifying the exact decision boundaries of the clusters in proximity [10]. The integrated adaptive fuzzy clustering (IAFC) developed by Y.S.Kim and S.Mitra[10] addresses this issue and forms better decision boundaries in the case of closely located clusters. A new similarity measure for the vigilance criterion and a new learning rule based on fuzzification of Kohonen- learning rule is used in IAFC.

The face recognition classifiers often face the problem of forming class boundaries in the case of similar face images. This paper proposes a new architecture for the classification of faces for the face recognition task based on the IAFC model [10] to address this challenge. The face images are pre-processed, dimensionality reduction is performed using PCA [16], feature extraction is performed using LDA [8] and then classified using the fuzzy neural network based on IAFC.

A practical face recognition system should identify a known individual as well as reject an unknown individual accurately. This is a challenging task since the system has to deal with the large intra-class variations and the small inter-class variations. The proposed system effectively rejects an unknown person by adaptively forming a new class for the unknown person.

The proposed system has been tested using the publicly available AT&T and Yale Face databases. A recognition rate of 98.75% and 99.39% was achieved for the AT&T and Yale databases respectively, which is better compared to the Back Propagation Neural Network (BPNN) system. The system achieved an equal error rate of 3.7 % and 1.3% for the AT & T and the Yale databases respectively.

The details of the face recognition system are discussed in the remainder of this paper. Section 2 covers the Proposed Face Recognition system. The design of the fuzzy neural network classifier is explained in section 3. In Section4, experimental results of evaluating the developed techniques and discussions are presented. Finally, conclusions are summarized in Section 5.

2 Proposed Face Recognition System

In the proposed face recognition system, the face images are first preprocessed to compensate for the intensity variations. PCA is applied to reduce the dimensionality of the face images and LDA is performed for extracting the features. The feature vectors obtained are then given to the neural network. The neural network classifies the input vector into one of the existing classes if some criteria are met or into a new class, otherwise. Thus, if the person in the input image is present in the database used for training (registrant), the person is identified as one of the persons in the database. If the person in the input image is a non-registrant (not present in the training database) then the person is rejected and classified into a new class by the neural network classifier.

2.1 Preprocessing

The face images are preprocessed by histogram equalization to account for the intensity variations in the face images. The two-dimensional face image of size p x q is converted into a vector of size m (m = p x q).

2.2 Feature Extraction

The face images are represented by a feature vector which helps the classifier to efficiently classify the face images. The PCA [16] is the most widely used feature for representing face images. However, PCA maximizes the intra class as well as the inter class scatter while performing the dimensionality reduction. For efficient classification, a feature like LDA [8] which maximizes the inter class scatter and minimizes the inter class scatter is more efficient. However the performance of LDA is affected by the Small Sample Size (SSS) problem and to deal with this, often PCA and LDA are combined. In this paper, the face images are represented by a feature vector obtained by combining PCA and LDA.

Principal Component Analysis (PCA)

The face images represented as vectors are subject to dimensionality reduction to reduce computational complexity. Principal Component Analysis (PCA) [16] is used for dimensionality reduction. Let the database of n training images be represented by n vectors $Z = (Z_1, Z_2, \ldots, Z_n)$ of size m each. The mean vector $\overline{Z}$ and the covariance matrix are calculated as in (1-2).

$$\overline{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i \tag{1}$$

$$\Gamma = \frac{1}{n}\sum_{i=1}^{n}\left(Z_i - \overline{Z}\right)\left(Z_i - \overline{Z}\right)^T = \phi\phi^T \tag{2}$$

The eigen values and eigen vectors of the covariance matrix Γ are calculated. Let $E = (E_1, E_2, .., E_t)$ be the t eigen vectors corresponding to the t largest eigen values. For the n patterns Z, their corresponding eigen face-based features X can be obtained by projecting Z into the eigen space as follows:

$$X = E^T Z \tag{3}$$

Thus the patterns of dimension m (m=p x q) is reduced to the dimension t. (t < m, t is taken to be (No of classes * No of training patterns per class) - No of classes.)

Linear Discriminant Analysis (LDA)

The objective of Linear Discriminant Analysis (LDA) is to perform dimensionality reduction while preserving as much of the class discriminatory information as possible. It is also more capable of distinguishing image variation due to identity from variation due to other sources such as illumination and expression. Two scatter matrices are calculated as in (4-5).

$$S_w = \sum_{j=1}^{R} \sum_{i=1}^{M_j} (x_i^j - \mu_j)(x_i^j - \mu_j)^T \tag{4}$$

$$S_b = \sum_{j=1}^{R} (\mu_j - \mu)(\mu_j - \mu)^T \tag{5}$$

Where S_w is called the within-class scatter matrix while S_b is called the between lass scatter matrix. j denotes the class while i denotes the image number. μ_j is the mean of class j while μ is the mean of all classes. M_j is the number of images in class j and R is the number of classes. LDA computes a transformation that maximizes the between-class scatter while minimizing the within-class scatter.

$$\text{maximize } |S_b| / |S_w| \tag{6}$$

The linear transformation is given by a matrix U whose columns are the eigenvectors of $S_w^{-1} S_b$ (called Fisherfaces). There are at most R - 1 non-zero generalized eigenvectors. However, in practice, S_w is often singular since the data are image vectors with large dimensionality while the size of the data set is much smaller. To alleviate this problem, PCA is first applied to the data set to reduce its dimensionality.

3 Proposed Fuzzy Neural Network Classifier

The design of the proposed fuzzy neural network classifier is explained in this section. The IAFC clustering algorithm based on which the proposed network is built is explained in section 3.1. The proposed network is explained in section 3.2.

3.1 Integrated Adaptive Fuzzy Clustering(IAFC)[10]

The IAFC [10] model is a fuzzy neural network similar to ART-1 that finds the cluster structure embedded in data sets. IAFC finds the actual decision boundaries of closely located clusters by incorporating a new similarity measure for the vigilance criterion and a new learning rule into a neural network.

The new learning rule, developed in IAFC, incorporates a fuzzy membership value μ_i, an intra-cluster membership value π, and a function of the number of iterations f(l)) into a Kohonen-type learning rule. The fuzzy membership value used in IAFC is based on the FCM model. The use of an intra-cluster membership value guarantees the fast convergence of the weights [10]. IAFC consists of three major procedures: deciding a winning cluster, performing the vigilance test, and updating the centroid of a winning cluster [10].The IAFC algorithm is explained in Fig.1.

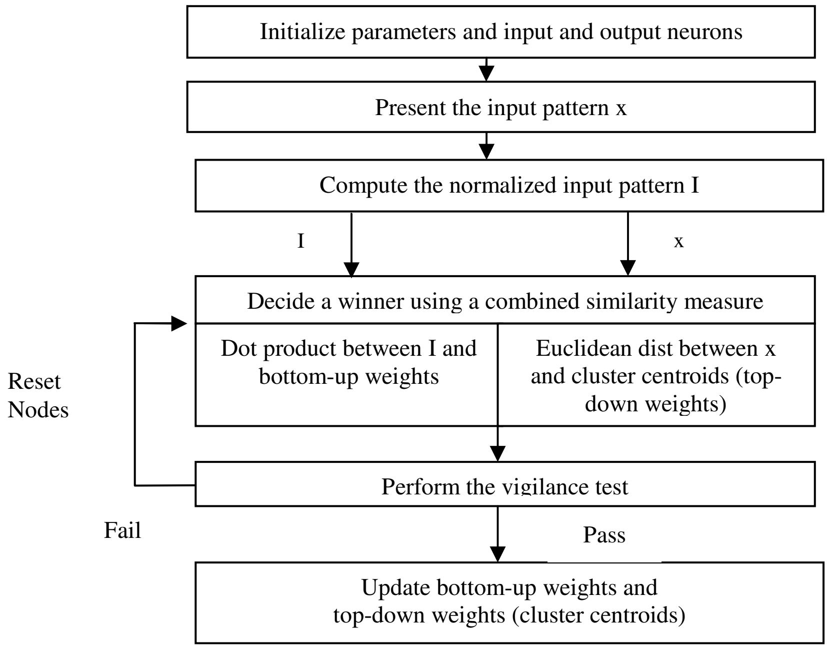

Fig. 1. The IAFC Algorithm

The input pattern and the normalized input pattern are presented to the fuzzy neural network in parallel. The dot product between the normalized input pattern and the bottom up weights is performed as in the equation

$$I \bullet b_i = \frac{x \bullet v_i}{\|x\| \bullet \|v_i\|} \tag{7}$$

where b_i is the bottom-up weight, which is the normalized version of the i^{th} cluster centroid (v_i) , from the input neurons to the i^{th} output neuron (cluster).

The output neuron that receives the largest value for the dot product in (7) wins the competition. Here, the winner is decided by the angle between the input pattern and the cluster centroids. This may lead to misclassifications because a cluster of which the direction of the centroid vector has the smallest angle with the input vector wins the competition regardless of its location with respect to the cluster centroids. In such a case, Euclidean distance can be used as a better similarity measure to determine a winner. However, cluster centroids cannot approach appropriate locations during the early stage of learning, thus causing poor performance of clustering algorithms. To prevent both problems, the IAFC algorithm uses a combined similarity measure to decide a winner.

In the IAFC algorithm, the winner is first decided by dot product as in (7). If the fuzzy membership value of the input pattern in the winning cluster is less than the value of the parameter σ, a threshold of the fuzzy membership value, the IAFC algorithm finds the winning cluster using the Euclidean Distance criterion. Otherwise the winner is same as the one obtained by the dot product.

If the parameter σ is low the angle between the input pattern and the cluster centroid is the dominant similarity measure to decide a winner. On the other hand, if the parameter σ is high, the Euclidean distance between the input pattern and the cluster centroid is the dominant similarity measure to decide a winner.

Once the winning cluster is selected, the IAFC algorithm performs the vigilance test using the un-normalized input data pattern according to the vigilance criterion.

$$e^{-\gamma\mu_i}\left\|x - v_i\right\| <= \tau \tag{8}$$

where τ is the vigilance parameter and the γ is a constant . τ controls the size of clusters and γ controls the shape of clusters [10]. If the value of the vigilance parameter is large, the size of clusters is large and vice versa. The fuzzy membership value μ_i is calculated as in FCM as in (9). In (9), n is the number of currently existing clusters which is updated during clustering and m is a weight exponent called the fuzzifier whose value is experimentally set to 2[10].

$$\mu_i = \left(\frac{1}{\left\|x - v_i\right\|^2}\right)^{1/m-1} \Bigg/ \sum_{j=1}^{n}\left(\frac{1}{\left\|x - v_j\right\|^2}\right)^{1/m-1} \tag{9}$$

If the winning cluster satisfies the vigilance test, its centroid is updated as in (10).

$$v_i^{new} = v_i^{old} + \lambda_{fuzzy}(x - v_i^{old}) \tag{10}$$

where λ_{fuzzy} is [f(l). π(x; v_i(old); τ). μ_i^2]. f(l) is a function of the number of iterations l, and π(x; v_i(old); τ) decides the intra-cluster membership value of the input pattern x in the i^{th} winning cluster. f(l) and π are calculated as in [10].

The centroid of each cluster is used as the top down weight related with each cluster. If the vigilance test is not satisfied, the winning output neuron is temporarily reset and the similarity test is performed using the remaining neurons. If all committed output neurons are reset, the first uncommitted output neuron is activated to form a new cluster.

3.2 Proposed Neural Network Classifier Based on IAFC

In this paper, a face recognition system based on IAFC is proposed. The training face images are clustered using IAFC and the trained network is then used for classifying the test face images. In the proposed system, an initial direction for the convergence of the weights and cluster centroids is provided by initializing the cluster centroids using the training face patterns. As a result of this, convergence of the algorithm is faster and classification accuracy is improved. This aids in a better classification of the face images.

The proposed system can be explained as follows:

1. In the training phase, the average training face patterns are used to initialize the weights and the centroids of the clusters. The number of classes is initialized to the number of persons in the database.
2. The network is then trained using the training face patterns. The refinement of the weights and the centroids is performed using the equation (10) as in IAFC.
3. Once the clustering is performed, the training of the network is completed.
4. The test face patterns are given as input to the trained network. The weights and the cluster centroids are not updated in the testing phase. The winner is computed using the combined similarity measure and the vigilance test is performed to find the class of the input pattern.

If the vigilance test is satisfied for any of the existing classes, the input pattern is classified into one of the classes in the database. Otherwise, a new class is adaptively formed which indicates that the input pattern is not present among the training classes. In the first case, the input pattern is a registrant (person in the database) and in the second case the pattern is a non-registrant (not present in the database).

4 Experimental Results and Discussion

Experiments have been carried out using the publicly available AT & T [17] and Yale [18] databases. The AT & T database also known as the ORL database of faces contains ten different images of each of 40 distinct subjects. The Yale Face Database contains 165 grayscale images in GIF format of 15 individuals. There are 11 images per subject.

4.1 Results

Recognition rate is considered for the registrants (i.e. the persons in the training classes) and the rejection rate is considered for the non-registrants (i.e. the persons not present in the training database.)[16]. Recognition Rate (for registrants) is the ratio of the number of patterns correctly classified to the total number of patterns tested multiplied by 100. Rejection Rate (for non-registrants) is the ratio of the Number of rejected patterns to the total number of patterns tested multiplied by 100.

From the AT & T database, 20 persons were selected as registrants and 20 persons as non-registrants. From the 10 patterns of each individual, 5 patterns were used for training the network and all the 10 patterns were used for testing the network. Hence 200 images from the database were used for training and all the 400 images were used for testing.

From the Yale Face Database of 15 persons, 7 persons were considered as registrants and 8 persons as non-registrants. Out of the 11 patterns of each person, 6 patterns were used for training and all the 11 patterns were used for testing the network. Hence 42 patterns were used for training and 165 patterns were used for testing the system.

The Fig. 2 shows the rejection rate and recognition rate obtained by the system for different values of the vigilance parameter for the 2 databases.

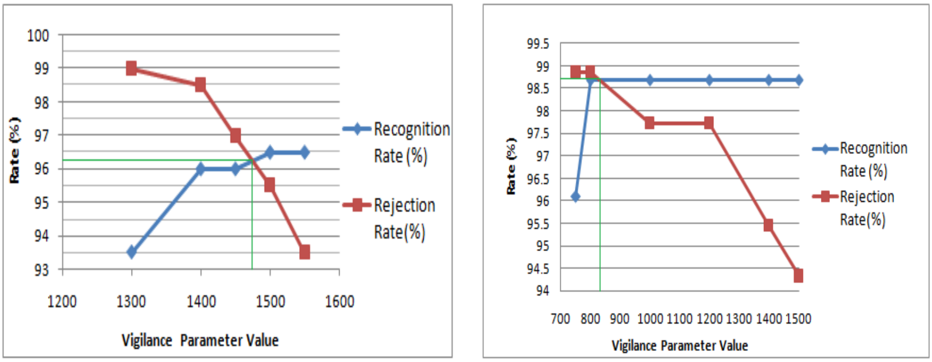

Fig. 2. Recognition and Rejection Rates for different values of τ for AT&T and Yale Databases

For the AT & T database, the system achieved a False Acceptance Rate (FAR) of 4% and False Rejection Rate (FRR) of 1.5 % for the vigilance parameter value of 1400 and the gamma value of 1.3. An Equal Error Rate (ERR) of 3.7 % is achieved. The proposed system achieved an FAR of 1.3% and FRR of 1.14 % for the vigilance parameter value of 800 and the gamma value of 1.3 for the Yale Face database. An Equal Error Rate (ERR) of 1.3 % is achieved.

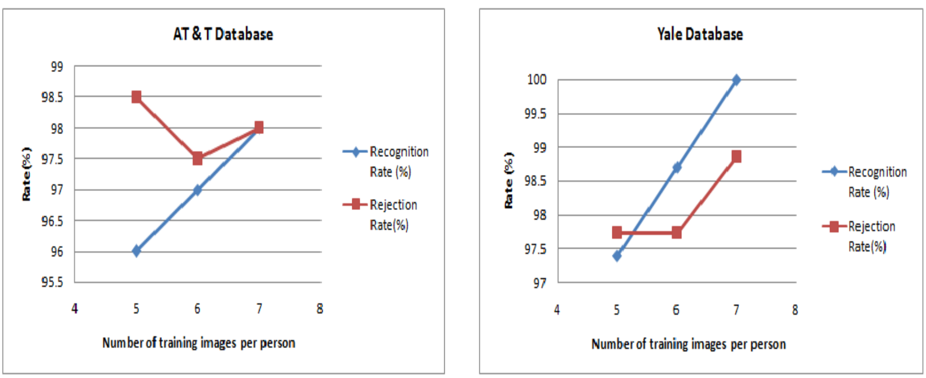

Fig. 3. Recognition and Rejection Rates for different number of training images per subject

The Fig.3 shows the performance achieved by the system when the number of training images is varied. The results were obtained on the AT & T and the Yale Face Databases.

The proposed face recognition system has been compared with the recognition results obtained by a face recognition system with Back Propagation Neural Network (BPNN) as the classifier and PCA+LDA as the feature extractor. The BPNN maps the face patterns of a non-registrant to the closest match in the database and hence the entire database has been considered as registrants for testing. The proposed system has also been tested considering all the persons in the database as registrants. The Fig. 4 shows the performance comparison of the proposed system and the BPNN system. The results show that the proposed system achieves better recognition rates compared to the BPNN system.

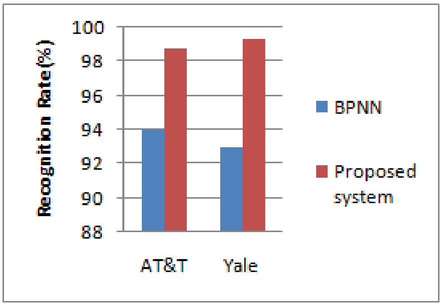

Fig. 4. Comparison of Recognition Rates for BPNN and proposed system

5 Conclusion

In this paper, a face recognition system using a fuzzy neural network classifier based on IAFC is proposed. The IAFC integrates the advantages of the fuzzy optimization constraint in fuzzy c-means (FCM), the control structure of adaptive resonance theory and a fuzzified Kohonen-type learning rule. The decision boundaries for the classes are obtained by a combined similarity measure and a vigilance criterion. The face features are extracted using a combination of PCA and LDA. LDA maximises the between class to within class scatter ratio and provides good discrimination information required for face recognition. The face features are given as input to the fuzzy neural network classifier. The network classifies the input face image as a registrant or a non-registrant in the database. The proposed system achieves better recognition rates compared to the BPNN system. A recognition rate of 98.75% and 99.39% is obtained by the proposed system for the AT & T and Yale databases respectively. The system achieved an equal error rate of 3.7 % and 1.3% for the AT & T and the Yale databases respectively, considering the rejection rate for non-registrants also. Practical applications of face recognition often require the rejection of non-registrants as well as the recognition of registrants reliably and the proposed system performs better in this respect compared to many of the existing systems.

References

1. Su, Y., Shan, S., Chen, X., Gao, W.: Hierarchical Ensemble of Global and Local Classifiers for Face Recognition. IEEE Transactions on Image Processing 18(8) (2009)
2. Zhao, W., Chellappa, R., Rosenfeld, A., Phillips, P.J.: Face Recognition: A Literature Survey. ACM Computing Surveys, 399–458 (2003)
3. Sahoolizadeh, H., Ghassabeh, Y.A.: Face recognition using eigen-faces, fisher-faces and neural networks. In: 7th IEEE International Conference on Cybernetic Intelligent Systems, pp. 1–6 (2008)

4. Moghaddam, B.: Principal manifolds and probabilistic subspaces for visual recognition. IEEE Trans. pattern Anal. Machine Intel. 24(6), 780–788 (2002)
5. Othman, H., Aboulnasr, T.: A separable low complexity 2D HMM with application to face recognition. IEEE Trans. Pattern. Anal. Machine Intel. 25(10), 1229–1238 (2003)
6. Er, M., Wu, S., Lu, J., Toh, L.H.: Face recognition with radial basis function (RBF) neural networks. IEEE Trans. Neural Networks 13(3), 697–710 (1999)
7. Lee, K., Chung, Y., Byun, H.: SVM based face verification with feature set of small size. Electronic Letters 38(15), 787–789 (2002)
8. Zhao, W., Chellappa, R., Krishnaswamy, A.: Discriminant analysis of principal component for face recognition. IEEE Trans. Pattern Anal. Machine Intel. 8 (1997)
9. Er, M.J., Chen, W., Wu, S.: High speed face recognition based on discrete cosine transform and RBF neural network. IEEE Trans. on Neural Network 16(3), 679–691 (2005)
10. Kim, Y.S., Mitra, S.: An adaptive integrated fuzzy clustering model for pattern recognition. Journal Fuzzy Sets and Systems (65), 297–310 (1994)
11. Huntsbergerd, T.L., Ajjimarangsee, P.: Parallel selforganizing feature maps for unsupervised pattern recognition. Int. J. General Systems 16(4), 357–372 (1990)
12. Carpenter, G.A., Grossberg, S., Rosen, D.: Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. Neural Networks 4, 759–771 (1991)
13. Simpson, P.K.: Fuzzy min-max neural networks Part 2: clustering. IEEE Trans. on Fuzzy systems 1(1), 32–45 (1993)
14. Newton, S.C., Mitra, S.: Self-organizing leader clustering in a neural network using a fuzzy learning rule. In: SPIE Proc., vol. 1565, pp. 331–337 (1991)
15. Newton, S.C., Pemmaraju, S., Mitra, S.: Adaptive fuzzy leader clustering of complex data sets in pattern recognition. IEEE Trans. on Neural Networks 3(5), 794–800 (1992)
16. Lu, J., Yuan, X., Yahagi, T.: A Method of Face Recognition Based on Fuzzy c-Means Clustering and Associated Sub-NNs. IEEE Transactions on Neural Networks 18(1) (2007)
17. AT & T. The Database of Faces, `http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html`
18. Yale Face Database, `http://cvc.yale.edu/projects/yalefaces/yalefaces.html`