Evolutionary Fuzzy Clustering and Parallel Neural Networks based Human Identification using Face Biometrics

Vivek Srivastava, Bipin K Tripathi, Vinay K Pathak
Department of Computer Science & Engineering,
Harcourt Butler Technological Institute, Kanpur, INDIA

ABSTRACT

This paper proposed a face recognition technique using evolutionary fuzzy clustering and parallel neural networks. Evolutionary fuzzy clustering is used for optimal distribution of images into the corresponding clusters which are generated through fuzzy clustering. Parallel neural network is used for training and recognition process is done based on the outcome of the parallel neural network. For evaluation of performance of the proposed technique, we use AT & T bell lab face dataset. Experimental analysis shows the efficacy of the proposed method and effect of various parameters on recognition rate.

General Terms

Pattern recognition, security.

Keywords

Biometrics, Face Evolutionary algorithm, fuzzy clustering, parallel neural network.

1. INTRODUCTION

Computational intelligence is an emerging area of research and is widely used for solving many real world problems dealing with large amount of data. Computational intelligence based techniques are applied for biometric applications [1] [3] [4] over the last decades. The strength and effectiveness of these techniques have been shown in various literature [5][11][14][16]. Fuzzy clustering has proven its better ability for various classification problems [7] [12] over traditional clustering techniques. Combining fuzzy clustering with evolutionary computation is quite efficient for solving classification and recognition problems. Very few literatures are devoted in which fuzzy clustering is used for biometric recognition such as face or iris recognition. Computational intelligence based techniques have been well introduced for solving the human recognition problems successfully over the last decades. But still some of them techniques suffer from huge computational cost and relatively lower recognition rate. Present work is intended for development of computational intelligence based technique which gives better recognition rate and less computational cost. In conventional techniques, statistical approach is frequently used for human recognition [8] [13].On other hand, neural network based techniques have proven many advantages over the conventional approaches because of their fast learning abilities and good generalization capabilities. Multilayer perceptron has been used in [1] for human face recognition which shows the comparatively analysis between MLP and its variant fuzzy MLP. Generally back propagation algorithm is used for learning of neural network. This still suffers with network scaling problem as we increase the number of neurons or layers in network, computational efficiency decreases rapidly because of slow convergence and occurrence of local minima. For avoiding this, we propose an evolutionary fuzzy clustering technique which decides the number of hidden neurons in the parallel neural network. Parallel neural networks are the combination of the conventional neural network which works as similar to the classical neural network. In our proposed technique evolutionary algorithm is used for optimal distribution of images into number of clusters generated from conventional fuzzy c-means algorithm. Parallel neural network is used for training and recognition of the training and testing image sets respectively.

Rest of the paper is as following- Section 2 deals with present methodology while the Section 3 shows experimental results and analysis. Finally Section 4 is conclusion.

2. PRESENT WORK

Present work is intended for face recognition using evolutionary fuzzy clustering with parallel neural network. We propose a novel approach for human recognition in which feature vector of images is obtained using well known principal component analysis and then based on this images are distributed to different clusters which are obtained by the evolutionary searching algorithm. Number of parallel network is decided based on the number of clusters. Maximum number of members per cluster is decided based on the maximum recognition rate. Outcome of these parallel networks are recorded and fed into the integrator network which combines result of all parallel networks known as gating network. Final recognition is done based on the result of the integrator.

2.1 Feature Extraction and distribution of images using Evolutionary fuzzy clustering

Various techniques have been proposed for feature extraction from the image. Principal component analysis (PCA) is widely used for feature extraction. In this work, we adapt PCA for feature extraction. Suppose an image I is $m \times m$ array of intensity values. An image can be also written as m^2 linear vector. If there are n images { $I_1, I_2, ...I_n$ } then covariance matrix is defined as follows:

$$Cov = \frac{\sum_{i=1}^{n}(I_i - \sum_{i=1}^{n} I_i)(I_i - \sum_{i=1}^{n} I_i)^T}{n}$$

Eigen vectors E={ $E_{1,}E_2,...,E_n$ }are calculated and *k* eigen vectors are selected based on *k* largest eigen values. Eigen face based features *A* can be obtained by projecting *I* into eigenface space as.

$$A = E^T I$$

Evolutionary computation involves the three basic approaches genetic algorithm, evolutionary programming and evolutionary strategies which are all biologically inspired. Evolutionary algorithms are similar as the process of natural evolution which is the driving process of emergence and well adapted organic structure. A single individual of the population is affected by the other individuals of population therefore according to the survival of the fittest only individuals who can perform under these conditions having greater chance to live longer.

Evolutionary searching is generally used for finding the best possible solution among the existing ones. In fuzzy clustering, initial partitioning is created randomly which satisfies the equation (1), (2) and (3), therefore different runs of the same algorithm may produce different partitions of the same dataset [2]. Objective function also plays a vital role for proper partitioning. Objective function describes the inter classes similarities which is minimized in each iteration. Various kinds of evolutionary algorithms have been proposed for overcoming these limitations such as multi objective evolutionary clustering [6] and ensemble based evolutionary clustering. There are two major categorizations of evolutionary algorithms in regards of fuzzy clustering. First category applies the evolutionary algorithms when the numbers of clusters are unknown while the second is applied for searching the best partition when the numbers of clusters are pre known. In first case, evolutionary algorithms are adopted for searching the best number of clusters. These algorithms are designed with underlying assumption that best number of clusters is unknown and it is presumed that number of clusters is inherent to the dataset. Rather using conventional fuzzy c-means clustering (FCM) algorithm we use evolutionary fuzzy clustering with minkowski distance algorithm, we call this technique as EFC-MD [15]. Motivation for using minkowski distance in stead of Euclidian distance is its more generalized nature. Also it does not restrict the shape of the clusters generated. EFC-MD involves the evolutionary search approach on different runs of the Fuzzy clustering in which minkowski distance is used as similarity/dissimilarity measure for different values of number of clusters(C). In this section we define steps of EFC-MD as the membership function, chromosome representation, population initialization, computation of fitness function and selection process of best population.

Let X= { $x_1, x_2, x_3,.........x_N$ } is input dataset having each elements of *n*-dimensions. Fuzzy c-mean clustering algorithm divides N datasets to C clusters with fuzzy partition matrix U of size C×N which is known as membership function.

Membership function is defined as U = [μ_{ik}] is C×N matrix which satisfied the following constraints –

$$\mu_{ik} \in [0,1] \quad \forall\ 1 < i < C \quad \text{And} \quad 1 < k < N \tag{1}$$

$$0 < \sum_{k=1}^{N} \mu_{ik} < N \quad \forall\ 1 < i < C \tag{2}$$

$$0 < \sum_{i=1}^{C} \mu_{ik} = 1 \qquad \forall 1 \le k \le N \quad \text{Where} \quad 2 \le C \le N \tag{3}$$

Each chromosome is a sequence of attribute values representing C clusters. Let $\Theta = \{C_{ij}\}$ where C_{ij} is defines as

C_{ij} = { 1 if jth data set belongs to ith cluster,

0 Otherwise }

Where $1 \le i \le C$ and $1 \le j \le N$

Initially C clusters are encoded in each chromosome and population is initialized randomly .Therefore in each run different initial population is generated.

The Objective function for EFC-MD is defined as following

$$f = \left[\frac{1}{\sum_{k=1}^{N} \sum_{i-1}^{C} (\mu_{ik})^m d^{2\beta}(x_{k,} O_i)} \right] \tag{4}$$

$$d^{\beta}(x_{k,} O_i) = \left(\sum_{j=1}^{n} \| x_{kj} - O_{ij} \|^p \right)^{\beta / p} ,1 \le p \le \infty,$$

$$0 < \beta \le 1 \tag{5}$$

Where *m* is a weighting exponent which is known as fuzzifier. Generally, the value of *m* lies between one to infinity. Value of m greatly influences the performance of FCM algorithm. When *m* approaches to infinity, the solution will be the center of gravity of whole dataset and when *m*=1 it behaves like classical c-means. There fore selection of suitable fuzzifier m is very important for implementation of FCM. In [9], it has been shown that a proper weighting exponent value depends on data itself.

$d^{\beta}(x_{k,} O_i)$ is minkowski distance and in objective function we use squared minkowski distance. Main motivation behind using minkowski distance in stead of using Euclidian distance is that the shape of the clusters decided for the given problem mainly depends upon the distance measure taken. The exact nature of these parameters depends on the shape of clusters to be generated, which may be boxes, ellipsoids, spheres and others. Selection of this distance measure does not tend the shape of cluster spherical which is often in Euclidian distance. The introduction of power β allows controlling the loss function against outliers [9]. In fuzzy clustering, we minimize the objective function which means the fitness function is inversely proportional to the objective function. Hence Higher value of *f* gives survival to the fittest population and best population is selected among the various offsprings generated on different runs. EFC-MD algorithm is iterated through the necessary conditions for minimizing the objective function with the following updates in member function and centre of the clusters:

$$\text{O}_i = \frac{\sum_{k=1}^{N} {\mu_{ik}}^m x_k}{\sum_{k=1}^{N} {\mu_{ik}}^m} \tag{6}$$

$$\mu_{ik} = \frac{d(x_k, O_i)^{\frac{1}{m-1}}}{\sum_{j=1}^{C} d(x_k, O_i)^{\frac{1}{m-1}}} \qquad i = 1,2.....,C \qquad (7)$$

Let $J(\mu, O)^{(t)}$ is the objective function at *t* th iteration then The EFC-MD algorithm termination condition is as following-

$\| J(\mu,O)^{(t+1)}$ - $J(\mu,O)^{(t)} \| < \Im$ Where $\Im$ is the pre specified threshold. (8)

Let $P^{(0)}, P^{(1)}, ...P^{(k)}$ be the *k* populations generated by the k iterations of EFC-MD algorithm. For *k=1,….,t*
Where *t* is the total number of runs generate the offsprings $P^{(k+t)}$ using equation (4) and termination criteria (8). Selection is done based on the elitism function which selects the best chromosome first.
EFC-MD iteratively finds the best population among the various population generated by different runs. After getting the best partitioning for training sets, testing is performed by matching the unknown test data with the centre of the clusters generated.

2.2 Training and Recognition using parallel neural network

As defined previous, parallel neural network is analogous to the classical neural network but it differs from other neuron models in terms of functional modules in which each module runs a neural network based on various kinds of characteristics of network such as number of hidden layers, activation function, and number of neurons per layer and learning algorithms. In this network each module performs independently and an integrator which is a gating network combines the outcome of different parallel networks to produce some decision. Integrator or gating network is a functional unit which aggregates the result of each module to decide final result.
Overall methodology is summarized as following-

1. Allocate feature vectors of the training set using EFC-MD.
2. Each parallel network corresponds to each cluster generated from above step. In each cluster maximum number of members is decided based on maximum recognition rate.
3. For each parallel network, Back propagation learning algorithm is used in training phase.
4. For testing, feature vector of unknown image is computed and fed into each parallel neural network. Output of each module is computed and results having maximum output from each module are selected and its corresponding unit is recorded.
5. Outcome of all the modules are integrated through gating network. Recognition of unknown test image is done based on the Euclidian distance measure between of the final outcomes of each module and the input test image.

3. EXPERIMENTAL RESULTS

To evaluate the proposed technique, we perform the experiment with well known dataset AT & T bell laboratories face dataset [10] which is formerly known as ORL face data. Recognition rate is calculated as the ratio between correctly recognized image and total images to identify. The AT & T Laboratories face database contains 400 face images from 40 individuals captured over the span of a 2-year period from subjects aged between 18 and 81. There are 10 different images of a single person based on variations of position, scale, rotation and expression (fig 1). Out of 400 images, we use 160 images randomly for

Fig. 1 AT & T bell laboratories Cambridge database

training and 240 images for testing which contains 6 images per person. Using PCA, we computed eigen faces of training set. In Training set, mean vector is calculated as the average of four images and we get 40 mean vectors which are distributed into the 10 clusters through EFC-MD algorithm. Final distribution of mean vector is shown in table I.

TABLE 1
ALLOCATION OF IMAGES ON VARIOUS CLUSTERS BASED ON EFC-MD FOR AT & T BELL LAB. FACE DATASET

	Members													
Cluster1	I_6	I_{10}	I_{11}	I_{13}	I_{16}	I_{18}	I_{22}	I_{25}	I_{26}	I_{31}	I_{33}	I_{37}	I_{38}	I_{39}
Cluster2	I_6	I_{10}	I_{13}	I_{16}	I_{18}	I_{22}	I_{25}	I_{26}	I_{31}	I_{33}	I_{34}	I_{37}	I_{38}	I_{39}
Cluster3	I_1	I_3	I_8	I_{14}	I_{15}	I_{17}	I_{20}	I_{21}	I_{24}	I_{27}	I_{28}	I_{29}	I_{32}	I_{40}
Cluster4	I_2	I_4	I_6	I_{11}	I_{12}	I_{16}	I_{18}	I_{19}	I_{22}	I_{26}	I_{33}	I_{34}	I_{35}	I_{36}
Cluster5	I_1	I_3	I_7	I_8	I_9	I_{14}	I_{17}	I_{20}	I_{21}	I_{24}	I_{27}	I_{28}	I_{29}	I_{32}
Cluster6	I_2	I_6	I_{11}	I_{12}	I_{15}	I_{16}	I_{18}	I_{19}	I_{22}	I_{26}	I_{33}	I_{34}	I_{35}	I_{36}
Cluster7	I_1	I_3	I_7	I_8	I_9	I_{14}	I_{17}	I_{20}	I_{21}	I_{24}	I_{27}	I_{28}	I_{29}	I_{32}
Cluster8	I_6	I_{10}	I_{11}	I_{13}	I_{16}	I_{18}	I_{22}	I_{25}	I_{26}	I_{31}	I_{33}	I_{37}	I_{38}	I_{39}
Cluster9	I_1	I_3	I_7	I_8	I_9	I_{14}	I_{17}	I_{20}	I_{21}	I_{24}	I_{27}	I_{28}	I_{29}	I_{32}
Cluster10	I_2	I_3	I_5	I_7	I_8	I_{12}	I_{15}	I_{19}	I_{23}	I_{30}	I_{32}	I_{35}	I_{36}	I_{40}

TABLE II
OUTPUT OF EACH MODULE FOR IMAGE NO 189

Module	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
Max. Output	0.782	0.296	0.754	0.953	.912	.712	.898	.912	.774	.936
Image	I_{12}	I_{19}	I_{21}	I_6	I_1	I_{35}	I_{14}	I_{31}	I_{29}	I_{23}

TABLE III
DISTANCE MEASURE FROM INPUT TEST IMAGE NO 189 IN GATING NETWORK

I_k	I_{12}	I_{19}	I_{21}	I_6	I_1	I_{35}	I_{14}	I_{31}	I_{29}	I_{23}
Distance measure	0.699	0.039	0.783	0.876	1.298	1.433	0.911	1.087	0.581	0.247

It has been observed that maximum recognition is obtained when we select fourteen members in each cluster as shown in fig. 3.We construct each parallel network correspond to each cluster, means that a single parallel neural network is composed of 40 input neurons which are the feature vectors taken based on maximum recognition rate as shown in fig 2 and 40 hidden neurons and 14 output neurons.

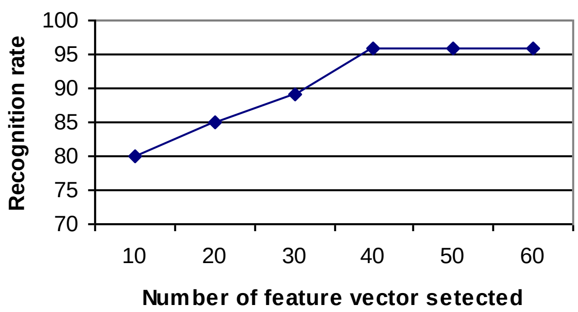

Fig. 2 Effect of selection of feature vector on recognition rate

Once the each network is trained for 160 test images then we perform testing on rest of 240 images. For unknown test image, we compute feature vector using PCA and select 40 major dimensions then this is fed into each parallel network and output is computed. For test image 189, maximum output of all parallel networks is shown in table II. Gating network collects maximum output of each module and finds its distance measure with the input vector using Euclidian distance. It has been observed that maximum recognition rate is achieved when the fuzzifier is close to one. On raising value of *m*, recognition rate decreases (fig 4).

Total 6000 iterations are executed for achieving the desired training pattern. We get 96% recognition rate which is quite acceptable rate.

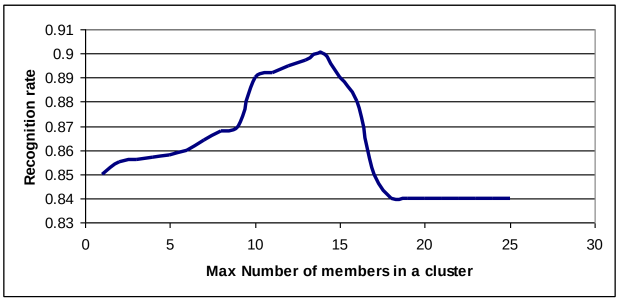

Fig. 3 Plot between recognition rates vs. max number of members in clusters selected

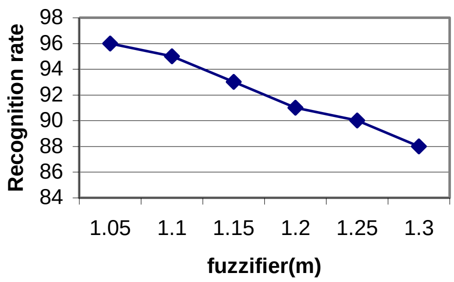

Fig. 4 Effect of *m* on recognition rate

It has been thoroughly observed that the present method yield comparatively better accuracy in terms of recognition rate. Also the impact of various parameters has been shown in this paper. Experimental results demonstrate the efficacy of the neural network configuration selection and lesser number of iterations.

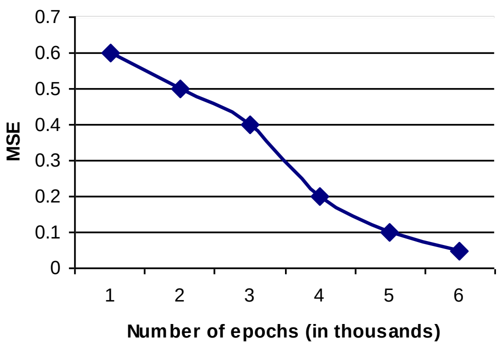

Fig. 5

Plot between MSE and number of epochs

4. CONCLUSION

In this paper an evolutionary fuzzy clustering and parallel neural network based approach is presented for human face recognition. Evolutionary fuzzy clustering is used for optimal distribution of images into clusters through fuzzy clustering. Parallel neural network is used for training and a gating network is used for final recognition. Experimental results demonstrate the performance of the present method using AT & T face dataset. 96 % recognition accuracy has been obtained using 160 images for training and 240 images for testing.

5. ACKNOWLEDGMENT

Present work is supported by the department of computer science and engineering, H.B.T.I., Kanpur and Gautam Buddh Technical University ,Lucknow.

6. REFERENCES

[1] D Bhattacharjee, D K Basu, M Nasipuri, M Kundu, “Human face recognition using fuzzy multilayer perceptron”, vol 14 ,pp 559-570, soft computing 2010

[2] E R Hruschka, R J G B Campello, A A Freitas, A C P L F de Carvelho,” A survey of evolutionary algorithms for clustering”, IEEE Trans on system man and cybernetics-Part C:applications and reviews, vol 39, no 2, pp-133-155, 2009

[3]Guan-Chun Luh, Chun-Yi Lin ,”PCA based immune networks for human face recognition”, 11(2011), 1743-1752, applied soft computing , 2011

[4] F Gaxiola ,P Melin,” Parallel neural networks for person recognition using segmentation and the iris biometric measurement with image preprocessing”,IEEE world congress on computational intelligence ,spain 2010.

[5] B K Tripatrhi, P K Kalra,”Functional mapping with higher order compensatory neuron model”, ”,IEEE world congress on computational intelligence ,spain 2010.

[6] K Deb, “Multi-Objective Optimization using evolutionary algorithms,New york: Wiley, 2001

[7] P Fazendeiro and J V de Oliveira, “A semantic driven evolutive fuzzy clustering algorithm”, in Proc IEEE Int Conf Fuzzy Syst., pp. 1-6, 2007.

[8] Er MJ, Shiquian w,Juwei L,Hock LT,”Face recognition with radial basis function neural networks,IEEE trnas neural network 13(3),697-709,2002

[9] P J F Groenen, K Jajuga”, Fuzzy clsuetring with squared minkowski distances”, fuzzy sets and systems, vol 120,pp. 227-237, 2001

[1[10]AT & T bell laboratories Cambridge database, URL: http://www.camorl.co.uk/facedatabase.html.

[11] B K Tripathi, P K Kalra ,” The novel aggregation function based neuron models in complex domain”, 14, 1069-1081, soft computing , springer, 2010

[12] Sueli A. Mingoti , Joab O. Lima, “Comparing SOM neural network with Fuzzy c-means, K-means and traditional hierarchical clustering algorithms”, European Journal of Operational Research 174 (2006) 1742–1759, 2006

[13] B L Zhang, H Zhang and S S Ge, " Face recognition by applying wavelet subband representation and kernel associative memory", IEEE trans neural network, vol 15, no 1, pp 166-177, jan 2004

[14] V Srivastava, V K Pathak, "Human recognition in passive environment using bidirectional associative memory", Intl journal of computer application, vol 1, no 15, pp-1-3, 2010

[15] V Srivastava, B K Tripathi, V K Pathak, "Evolutionary fuzzy clustering with minkowski distances", proceeding of 18th international conference on neural information processing ICONIP 2011, Part II, LNCS 7063, pp. 753--760. Springer, Heidelberg (2011)

[16] Tripathi B K , Kalra P K , "On efficient learning machine with root power mean neuron model in complex domain", IEEE trans on neural network, vol 22, no 5, pp 727-738, 2011