Face Recognition Using Firefly Model

Mahdi Nasrollah Barati[1], Maryam Khademi[2]
[1]Karaj Branch, Islamic Azad University, Karaj, Iran
[2]South Tehran Branch, Islamic Azad University, Tehran, Iran

ABSTRACT

This paper presents a technique for facerecognition. To thiseffect, Principal ComponentAnalysis (PCA)is used to decrease the data content and extract features of face images from ORL databases to train and test the neural network. In the next stage, a classifierbased on feed forwardback-propagation artificial neural network is employed Firefly-optimized learning rate and hidden layer size. The accuracy of classification is 99.83% on the ORLdatabase. This result shows that the proposed method is effective.

Keywords: -*Facerecognition,Firefly, principalcomponent analysis, multi-layerperceptron.*

---

1. INTRODUCTION

Biometrics is identification of an individual on the basis ofunique physiological and behavioral patterns. It is fastreplacing other means of authentication like passwords andkeys due to the inherent drawbacks in them and increasingeffectiveness and reliability of the biometric modalities. Thepasswords can be forgotten or hacked, while keys can be lost.

But the individual's unique physiological or behavioralcharacteristics are hard to forged or lost [1].Nowadays, biometric technologies such as face recognition(FR) [2], fingerprint recognition [3], iris recognition [4], andpalm-print recognition [5] are commonly used. Facialrecognition's advantages outweigh other biometric techniques.

One of the advantages of face recognition system is that it isnon-invasive which means that it does not require a person tobe isolated from the crowd to be examined. Although facerecognition system has many challenges such as head pose,illumination, aging and occlusion but because of its capabilityin collecting convenient, high user public acceptability and applicability, it has been captured more and more attention [6].

Face is the most significant part of the human body for ourdaily mutual interaction [7]. A general statement of facerecognition problem can be formulated as follows: First,giving still or video images of a scene, then identifying orverifying one or more persons in the scene using a storeddatabase of faces [6].

The primary task in a FR system is the extraction of thefeatures [8]. The "feature extraction" is considered as a preclassification step in most image classification systems. Thefeatures are usually local or global structural descriptors of theimage. The subsequent classification step then works in thefeature space, where

a large number of classifiers may beemployed. In most cases, the extracted features restrict theperformance of image classification systems because theyneed to be selected with great care and can affect theclassifiers. For example, since images or objects are oftenshifted, scaled and rotated, it is desirable to define (or design)the features so that they are invariant or robust to thesechanges [9]. There are also powerful machine learningalgorithms, such as classic artificial neural networks [10] andconvolution neural networks [11], that can be employed toautomatically discover good features from a large number oftraining images.
The proposed system is based on Principal Component Analysis (PCA) andFirefly Algorithm (FA).

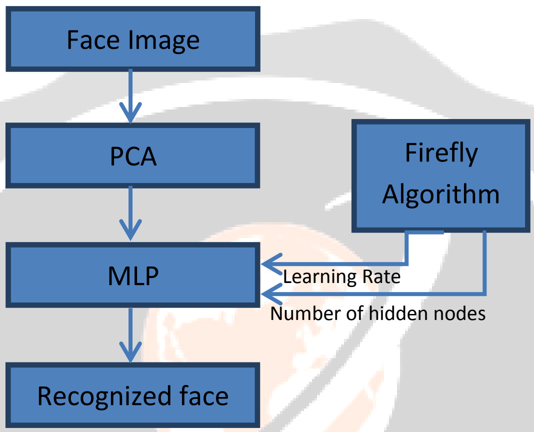

Fig -1: Block diagram of the proposed method

Multi-Layer Perceptron (MLP) neural network, whosestructural and learning parameters are optimized by FA, isemployed as the classifier (Fig -1). The performance results ofproposed mothed are compared with the similar models, whichis based on GeneticAlgorithm (GA) and Particle Swarm Optimization (PSO).

2. RELATED WORK

Researchers for FR problem (see [7] for a survey has developed a wide variety of approaches. From apoint of view, these various approaches are fallen into twogeneral groups: feature-based approaches and holisticapproaches. Feature-based approaches analyze shapes andgeometrical relationships of the individual facial featuresincluding eyes, mouth and nose whereas holistic approachesanalyze the face images as two-dimensional holistic patterns.
Among statistical approaches, PCA and linear discriminantanalysis (LDA) are two powerful statistical tools for featureextraction. Kirby and Sirovich [12] were the first to employKarhunen–Loeve Transform (KLT) to represent facial images.
Afterwards, Turk and Pentland [13] developed a PCA-basedapproach namely "eigenface". Etemad and Chelappa [14],Belhumeur et al. [15]and Zhao et al. [16] then proposed theLDA "Fisherface" method to extract features that are mostefficient for classification. Because of some limitations of thePCA and LDA, a

variety of modifications has been planned[17]. The advantages of deterministic transforms make theman interesting class of feature extraction approaches. DiscreteFourier Transform (DFT) [2], Discrete Cosine Transform(DCT) [18], DWT [19], Curvelet Transform [20] andContourlet Transform [21] are the important approaches ofthis class.

Combination of statistical and deterministic transformsconstructs a new type of feature extraction approach with bothbenefits. In this type, transforms decrease the dimension ofdata to avoid singularity and decrease the computationalload of statistical methods. Researchers have surveyed numerous combinations of the DCT, DFT, PCA and LDA. Ramasubramanian and Venkatesh[22] used amixture of the DCT, PCA and the characteristics of the humanvisual system for encoding and recognition of faces. Also,development of Curvelet[23] that offers enhanced directionaland edge representation has prompted researchers to applythem to several areas of image processing.Curvelet-basedPCA [5] and curvelet-based LDA [24] are some recentcurvelet-based face recognition approaches.

As other sample reported face recognition systems in therecent decade, we can mention the following systems:

Chitaliya and Trivedi[25] proposed an efficient facerecognition method based on the CT using PCA and theEuclidean distance classifier. They decomposed each faceusing the CT. PCA is then applied to reduce thedimensionality of the feature vector. Finally, the reducedfeature vector is adopted as the face classifier. Liau and Isa[26] proposed face-irismultimodal biometric system based onfusion at matching score level using Support Vector Machine(SVM). They employed DCT as a feature extractor and usedPSO to obtain an ‘‘optimized’’ subset of those features.

3. PRELIMINARIES

The foundation of techniques that are used inthe proposedmethod is reviewed as follows:

3.1 Principal Component Analysis

PCA, which linearly transforms the original signals intonew uncorrelated features, has been a famous method forfeature extraction. As described in face recognitionapplication, PCAs are used with two main purposes. First, itreduces the size of data to computationally possible size.Second, it extracts the most illustrative features out of theinput data so that although the size is reduced, the mainfeatures remain, and still be able to denote the original data[27].

The eigenface procedure is as follows:

1. Obtain N training images $I_1, I_2, I_3, \ldots, I_N$.
2. Represent each image I_i as a vector discussed above, eachimage is of size n.
3. Find the mean face vector ψ for N images:

$$\psi = \frac{1}{N}\sum_{i=1}^{N} I_i$$

4. Subtract the mean face from each face vector I_i to get a setof vectors Φ_i. The purpose of subtracting the mean imagefrom each image vector is to be left with only thedistinguishing features from each face:

$$\Phi_i = I_i - \psi$$

5. Find the covariance matrix C:

$$C = AA^T = A^T A$$

Where $A = [\Phi_1, \Phi_2, \dots, \Phi_N]$

Note that the covariance matrix is simply made by puttingone modified image vector obtained in each column.

6. Calculate the eigen vectors u andeigenvalues d of C.

7. Multiply the mean subtracted images with correspondingeigen vectors.

8. Select the top Leigen vectors with the highest eigen values.

9. The eigen values corresponding to the eigen vectors arecalled eigenfaces.

However, the PCA always carry information on bothsignal and disturbance, so the problem that how manyprincipal components should be chosen remains a significantquestion [28].

3.2Firefly Algorithm

The Firefly Algorithm was developed by Yang ([29] [30]), and it was based on the following idealized behavior of the flashing characteristics of fireflies:

• All fireflies are unisex so that one firefly is attracted to other fireflies regardless of their sex;

• Attractiveness is proportional to their brightness, thus for any two flashing fireflies, the less bright one will move towards the brighter one. The attractiveness is proportional to the brightness and they both decrease as their distance increases. If no one is brighter than a particular firefly, it moves randomly;

• The brightness or light intensity of a firefly is affected or determined by the landscape of the objective function to be optimized.

For a maximization problem, the brightness can simply be proportional to the objective function. Other forms of brightness can be defined in a similar way to the fitness function in genetic algorithms.

Firefly Algorithm

Objective function $f(x)$, $x = (x_1, \dots, x_d)^T$

Initialize a population of fireflies x_i(i = 1, 2, ..., n)

Define light absorption coefficient γ

while (t<MaxGeneration)

for$i$ = 1: n all n fireflies

for$j$ = 1: i all n fireflies

Light intensity Ii at xi is determined by $f(x_i)$

if ($I_j > I_i$)

Move firefly i towards j in all d dimensions

end if

Attractiveness varies with distance r via exp $[-\gamma r^2]$

Evaluate new solutions and update light intensity

end for *j*

end for *i*

Rank the fireflies and find the current best

end while

Postprocess results and visualization

The movement of a firefly *i* is attracted to another more attractive (brighter) firefly j is determined by

$$x_i^{t+1} = x_i^t + \beta_0 e^{-\gamma r_{ij}^2}(x_i^t - x_i^t) + \alpha\varepsilon_i^t$$

Whereβ_0 is the attractiveness at $r = 0$, the second term is due to the attraction, while the third term is randomization with the vector of random variables ε_i being drawn from a Gaussian distribution. The distance between any two fireflies *i* and *j* at x_i and x_j can be the Cartesian distance$r_{ij} = \|x_i - x_j\|_2$ or the l_2-norm. For other applications such as scheduling, the distance can be time delay or any suitable forms, not necessarily the Cartesian distance. For most cases in our implementation, we can take $\beta_0 = 1$,$\alpha \in [0, 1]$, and $\gamma = 1$. In addition, if the scales vary significantly in different dimensions such as -10^5 to 10^5 in one dimension while, say, -10^{-3} to 10^3 along others, itis a good idea to replace α by αS_k where the scaling parameters $S_k(k = 1, . . ., d)$ in the d dimensions should be determined by the actual scales of the problem of interest. In essence, the parameterγcharacterizes the variation of the attractiveness, and partly controls how the algorithm behaves. It is also possible to adjust γso that multiple optima can be found at the same during iterations.

3.3Multi Layer Perceptron

The MLP [31] is the most widely used neural network thatconsists of three main layers: input layer, hidden layer(s), andoutput layer. The hidden and output layers contain nodes that receive signals flowing from nodes in the previouslayer, whereas the input layer contains nodes that receive theinput features directly. In the hidden and output layers, the netinput to node *j* is represented by:

$$a_j = \sum_{i=1}^{d} w_{ji} x_i + w_{j0}$$

Where x_i is the input, $w_{ji}$is the weight related to each nodeconnection, and $w_{j0}$is the bias connected to node *j*. This sumis sent through a transfer function. Thus, the output of thenode is:

$$y_j = \frac{1}{1 + e^{-a_j}}$$

Generally, the transfer functions employed in MLP includelog-sigmoid function, tan-sigmoid function, and linearfunction. The number of input and output nodes is uniquelydetermined by the number of input features and output classes.The important issue is how to suitably set the number ofhidden nodes. There are no specific guidelines to determinethe optimum number of hidden nodes, except based on one'sexperience. It is generally understood only that setting too fewor too many hidden nodes causes lack-of-fit or over-fitting inthe network. Although,the trial-and-error method is normallyexploited to set the network parameters. In this paper, Fireflyalgorithm is employed to set the number of hidden nodes andlearning rate in MLP. In this case, we search the space of (10,50) and (0, 0.9) for the number of hidden nodes and learningrate, respectively.

4. SIMULATION AND EXPERIMENTAL RESULTS

In order to compare the performance of various facerecognition algorithms, a complete, analytically annotateddatabase is required. A database contains face images thathave been taken at diversity of pose angles, with an inclusivevariety of illumination angles.

This paper has employed the publicly available standardOlivetti-Oracle Research Lab (ORL) database as the face dataset. ORL consists of 400 frontal faces. There are 10 tightlycroppedimages of 40 individuals. All images are of grey scalewith a 48*48 pixels resolution. The face images are slightlyvaried in lighting conditions, pose, scale, face expression andpresence or absence of glasses. All images were taken under adark background. The faces are regularly positioned in theimage frame, and very little background is visible.Fig -2: depicts sample images from ORL database.

Fig -2: Sample images from ORL database

In this simulation, the test data is separated from training datarandomly and 70% of images (i.e., 280 images) in database areused for training and the remaining images (i.e., 120 images)are considered for testing. By the consumption ofPCA, the eigen vector and eigen value of a specific image arecalculated and 50 of the best features are extracted. To trainthe network, the features which selected by FA are fed intothe neural network input. Therefore, the input matrix of neuralnetwork is 280*N (in which N is optimized by FA). Thecolumns of this matrix are the best features and the rows arethe images which are enthusiastic to be trained. The value ofoutput target for each row distinguishes the specific value of image class. As far as theORL database has 40 different images, the output of thenetwork should be 40 neurons, that each of them indicatesspecific person. FA optimizes the number of hidden layer neurons and the required learning rate of the network,through the back-propagation algorithm process. Then, theobtained optimized value is applied to train the network.Therefore, the network is trained by the optimized number ofhidden layer neurons and learning rate to achieve its suitableweights and bias values.

In the test phase, with the observance of the feature vectorand mean value of training image, Eigen values are extractedfrom the test images. In this case, the input matrix of neural network has120 rows and N columns (in which N is optimized by FA). This matrix is fed into the input of neural network toobtain the class of each image. The output value of the neuralnetwork is between [0, 1] (because of sigmoid activationfunction in output layer). Therefore, to specify the output class,the maximum of output is determined and in this process, themaximum value is replaced by 1 and the remaining outputs arereplaced by 0.

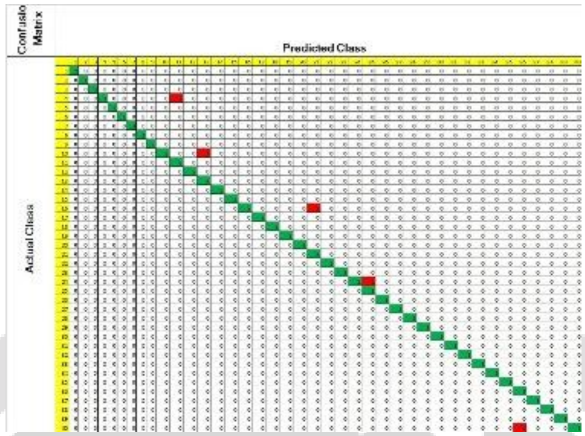

Fig -3: Confusion matrix for 40 classes

The confusion matrix for 40 classes is depicted inFig -3.In the artificial intelligence field, this matrix is the one thatdescribes the performance of the classification algorithm.Each column of the matrix explains the estimated value, whilethe rows show the actual value. In this way, the detection rate (DR) is calculated as:

$$DR = \frac{TP}{TP + FP}$$

In this research, the simulations are run on a PC poweredby an AMD Athlon™ 7750 Dual-Core Processor 2.71 GHzCPU, and 4 GB of RAM.

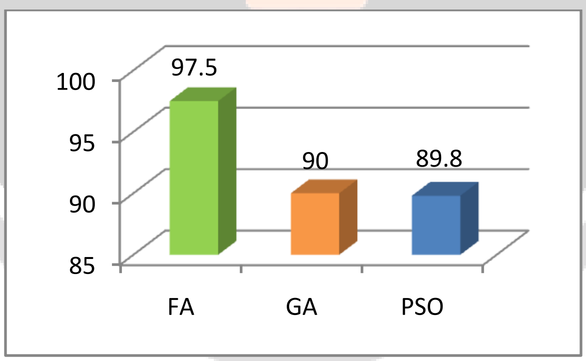

Fig -4: The Performance of simulated models

Fig -4: illustrates the results of proposed method. As can be seen, FA+PCA method hashigher detection rate.

5. CONCLUSION

Feature extraction, feature selection, and classification arethe three stages of most face recognition systems. At the classification stage, the MLP neuralnetwork is used. The learning rate and number of hidden layernodes in MLP are optimized by Firefly algorithm.Experimental results show that PCAequipped with Firefly algorithm for optimizing feature selectionprocess and MLP's parameters results in 99.83% recognition rate that is higher than GA and PSO simulated insimilar conditions.

REFERENCES

[1] M. Hassan, M. Mustafa, and S. A. Khan, M. A. Bin, A feature level multimodal approach for palmprint identification using directional subband energies.: Network and Computer Applications, vol. 34, pp., 2011.

[2] B. Liu and F. Wang, Face recognition approach based on 2D discrete fractional Fourier transform.: Proc. Int. Conf. Commun. Software and Networks, pp. 656-660, 2011.

[3] Y. Peng, D. Peng, and L. Wenda, F. Dongjin, A novel probabilistic model based fingerprint recognition algorithm.: model based fingerprint recognition algorithm," Procedia Engineering, 2012.

[4] K. Grabowski, M. Napieralska, W. Sankowski, M. Zubert,and A. Napieralski, R. Szewczyk, A reliable iris recognition algorithm based on reverse biorthogonal wavelet transform.: Pattern Recognition Letters, vol. 33, pp. 1019-1026, 2012.

[5] T. Mandal and Q. M. J. Wu, Face recognition using curvelet based PCA.: Proc. Int. Conf. Pattern Recognition, pp. 1-4, 2008.

[6] N. Veeranjaneyulu, and K. V. K. Kishore, B. Jyostna Devi, A novel face recognition system based on combining eigenfaces with fisher faces using wavelets.: rocedia Computer Science, vol. 2, pp. 44-51, 2010.

[7] R. Chellappa, P. J. Phillips, and A. Rosenfeld, W. Zhao, Face recognition: a literature survey.: ACM Comput. Surv., vol. 35, pp. 399-458, 2003.

[8] N. Farheen, K. Manikantan, and S. Ramachandran, V. Vidya, Face recognition using threshold based DWT feature extraction and selective illumination enhancement technique.: Procedia Technology, vol. 6, pp.334-343, 2012.

[9] D. G. Lowe, Distinctive image features from scale-invariant keypoints.: Int. J. Comput. Vision, vol. 60, pp. 91-110, 2004.

[10] G. E. Hinton, and R. J. Williams, D. E. Rumelhart, Learning internal representations by error propagation. Cambridge: in Parallel Distributed Processing:Explorations in the Microstructure of Cognition, vol. 1; Foundations, 1986.

[11] D. Steinkraus, and J. C. Platt, P. Y. Simard, Best practices for convolutional neural networks applied to visual document analysis.: Proc. Int. Conf. Document Analysis and Recognition, pp. 958-963, 2003.

[12] M. Kirby and L. Sirovich, Application of the Karhunen-Loeve procedure for the characterization of human faces.: IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 12, pp. 103-108, 1990.

[13] M. Turk and J. A. Pentland, Eigenfaces for recognition.: Cognitive Neurosci., vol. 3, pp. 71-86, 1991.

[14] K. Etemad and R. Chellappa, Discriminant analysis for recognition of human face images.: J. Opt. Soc. Am., vol. 44, pp. 1724-1733, 1997.

[15] J. P. Hespanha, and D. J. Kriegman, P. N. Belhumeur, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection.: IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, pp. 711-720, 1997.

[16] A. Krishnaswamy, R. Chellappa, D. L. Swets, and J. Weng, W. Zhao, Discriminant analysis of principal components for face recognition.: Proc. Int. Conf. Face and Gesture Recognition, pp. 336-341, 1998.

[17] L. Xiaoliang, and H. Cheng, W. Fulong, An improved kernel Fisher discriminant analysis for face recognition.: Proc. Int. Conf. Computational Intelligence and Security, pp. 353-357, 2009.

[18] G. Gunlu and H. S. Bilge, Face recognition by using 4D discrete cosine transform.: Proc. Int. Conf. Signal Processing, Commun. and Applications, pp. 1-4, 2008.

[19] H. Hu, Variable lighting face recognition using discrete wavelet transform.: Pattern Recognition Letters, vol. 33, pp. 1526-1534, 2011.

[20] S. Prasad, and G. Bakul, G. K. Verma, Robust face recognition using curvelet transform.: Proc. Int. Conf. Commun., Computing and Security, pp. 239-242, 2011.

[21] L. Weifeng, and L. Qingmin, W. Biao, Face recognition based on nonsubsampled contourlet transform and block-based kernel Fisher linear discriminant.: Proc. Int. Conf. Acoust., Speech and Signal Processing, pp. 1533-1536, 2012.

[22] D. Ramasubramanian and Y. V. Venkatesh, Encoding and recognition of faces based on the human visual model and DCT.: Pattern Recognition, vol. 34 , pp. 2447-2458, 2001.

[23] D. L. Donoho and M. R. Duncan, Digital curvelet transform: strategy, implementation, and experiments. Stanford University: Technical Report No. 2000-12, 2000.

[24] S. Ghouzali, M. El Hassouni, M. Rziza, and D. A. El Aroussi, M. El Aroussi, Curvelet-based feature extraction with B-LDA for face recognition.: Proc. Int. Conf. Computer Systems and Applications, pp. 444-448, 2009.

[25] N. G. Chitaliya and A. I. Trivedi, “An efficient method for face feature extraction and recognition based on contourlet transforms and principal component analysis.: Procedia Computer Science, vol. 2, pp. 52-61, 2010.

[26] H. F. Liau and D. Isa, Feature selection for support vector machinebased based face-iris multimodal biometric system.: Expert Systems with Applications, vol. 38, pp. 11105-11111, 2011.

[27] N. G. Chitaliya and A. I. Trivedi, An efficient method for face feature extraction and recognition based on contourlet transforms and principal component analysis.: Procedia Computer Science, vol. 2, pp. 52-61, 2010.

[28] W. Li, Y. Zhang, B. Sun, Y. Zhong, J. Liu, D. Yu, and M. Li, D. Qu, Support vector machines combined with wavelet-based feature extraction for identification of drugs hidden in anthropomorphic phantom.: Measurement, vol. 46, pp. 284-293, 2013.

[29] X.S. Yang, Nature-Inspired Metaheuristic Algorithms. UK: Luniver Press, 2008.

[30] X.S. Yang, Firefly algorithms for multimodal optimization, in: O. Watanabe, T. Zeugmann (Eds.), Stochastic Algorithms: Foundations and Appplications. Berlin: SAGA 2009, Lecture Notes in Computer Science, vol. 5792, Springer-Verlag, 2009, pp. 169–178., 2009.

[31] G. E. Hinton, and R. J. Williams D. E. Rumelhart, Learning representations by back-propagating errors.: Nature, vol. 323, pp. 533-536, 1986.