International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 7 (2018) pp. 4802-4806

Independent Component Analysis and Support Vector Neural Network for Face Recognition

Dr. Abdulqadir Ismail Abdullah
Computer Science Department, Knowledge University, KRG-Iraq.

Mustafa Zuhaer Nayef Al-Dabagh
Computer Science Department, Knowledge University, KRG-Iraq.

Mustafa H. Mohammed Alhabib
Communications and Computer Engineering Department, Cihan University-Erbil, KRG-Iraq.

Abstract
With increasing rate of security threats in last decade, attentions toward the field of face detection and recognition have increased rapidly. Many face recognition algorithms have been introduced and most of them are focused on increasing the accuracy rate of the recognition system. In this paper, face recognition system using Independent Component Analysis (ICA) for features extraction and Support Vector Neural Network (SVNN) for classification is suggested. Also, a comparison between each of SVNN, Support Vector Machine (SVM) and Artificial Neural Networks (ANNs) is given as a demonstration of the reliability of the proposed method. The experiments are implemented using Yale databases and the results reveal that the suggested method has a classification accuracy of % 95.238, which is higher than the results of (ICA+SVM) and (ICA+ANNs) methods, respectively.

Keywords: Face Recognition, ICA, SVM, ANN.

INTRODUCTION

In order to keep up with the growing pace of technological evolution, the need for a robust and effective recognition system has significantly increased. In general, face recognition is one of the systems that drew a lot of attention from researchers [1]. Facial recognition is one of the applications that can be typically used in security systems to identify a person based on image or video frame. The most commonly implemented technique for this task is by comparing the extracted facial features of the captured image with features from the facial image of the same person in a database [2-3].

Normally, Face recognition systems consist of two important stages: feature extraction and classification, in addition to other preprocessing steps. Feature extraction is an important and effective part in the face recognition algorithm and the method's ability in expressing the face may improve the accuracy rate and reduce the computation time and storage space [4].

It is possible to apply many methods in order to extract features from the facial image such as Principle Component Analysis (PCA), Linear Discriminant Analysis (LDA) and Independent Component Analysis (ICA). PCA [5] is one of the first techniques that is applied in feature extraction stage, but in fact, this technique is more appropriate for image reconstruction because it doesn't provide any consideration for the isolation of different classes but rather seeks an optimal division of the feature subspace. In order to overcome this deficiency, LDA method can be applied instead, which is considered as a supervised method. It aims to reduce the variances among the samples of the same class and maximize them between different classes [6]. But, this method has a problem with the small sample size, therefore, the traditional model of this method can’t be applied directly because the matrix of the within-class scatter is singular.

ICA is another well-known robust method that might be used in face recognition. This method aims at finding the best basis by identifying the high-order relations among the image pixels [7]. It has a number of advantages over other methods. First of all, orthogonality is not mandatory for its based vectors, which decreases the amount of reconstruction error. Secondly, higher order statistics are being used by its feature vectors as well as covariance matrix information [8].

Classification is the second important stage of the recognition system. Many of classifiers are currently proposed in the field of machine learning like artificial neural network (ANN) and Support Vector Machine (SVM) [9].

SVM is usually used to classify patterns in small and large representation space. In this method, hyperplanes for pattern classification are determined and their number is proportional to the number of classes. To distinguish the hidden elements in one class, the later probabilities in 1-vs-1/1-vs-rest paradigm are computed [10].

Artificial Neural Networks (ANN) are applied in order to obtain the weight set to be used in the classification stage. ANN compensate for the limitation of SVM by using an appropriate threshold in the case of outside patterns classification. If the patterns are not related to any class in L given classes, ANN recognize and send the results to the external given classes [11].

The contents of this paper are arranged as follows: In part 2, the preprocessing, feature extraction and classification stages are introduced, while the practical tests and study of the results are given in part 3. Finally, the conclusions of the paper and future work are summarized in part 4.

THE STRUCTURE OF THE PROPOSED SYSTEM

The structure of the proposed system consists of three major stages, preprocessing, face feature extraction and face classification. Fig. 1 displays the structure of the proposed recognition system in which ICA method is used to extract features from the image and a new training method named Support Vector Neural Network is applied for classification. Finally, number of experiments and comparison with other training methods is discussed. These steps are elaborated as follows:

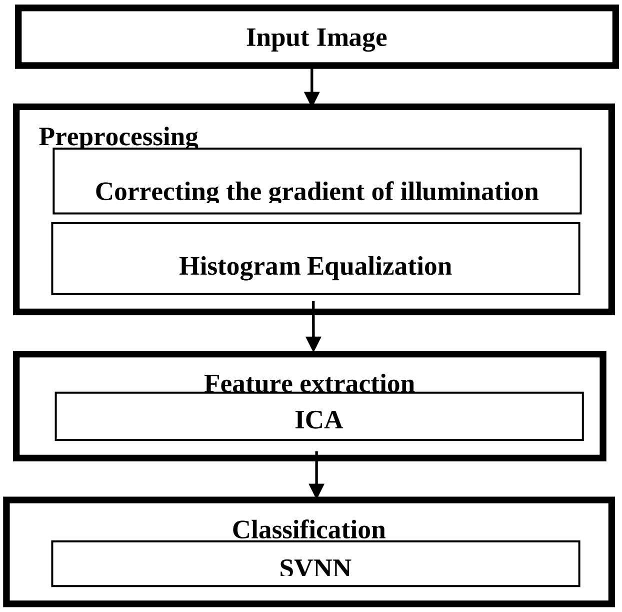

Figure1: Face Recognition System Process

Image preprocessing

The aim of this step is to decrease the light projected on the images, therefore, it is necessary to implement some of preprocessing steps before feature extraction. The preprocessing stage involves two steps:

Correction of the gradient of illumination

This step calculates the fittest value of luminance in the image and then subtract this value from the pixels of the original image. The aim of this step is to reduce the heavy shadows that appear duo to the effect of extreme lighting angle [12].

Histogram equalization

Histogram equalization is a type of transformation that can be used to evenly spread the image histogram in order to neutralize the effect of changes in illumination brightness and also the variations in the camera response curves. [13].

Feature Extraction

The feature extraction stage aims at extracting a group of representative features from the image that can be used in the classification stage. Without using this stage the classification process will be very difficult and impossible to implement since the original face image has a huge amount of information and most of these information cannot be used for classification. By using this stage the information that are required to represent the face image is reduced.

Independent Component Analysis (ICA)

ICA is an unsupervised learning rule method inspired by the optimal sigmoid neuron standard. Let's consider x as an input signal, y as the output and g as a nonlinear squashing function [14].

$$y = g(u) = \frac{1}{1 + e^{-u}} \qquad (1)$$

$$u = wx + w_0 \qquad (2)$$

Here, x is applied as input to the function g(x). The data of the output signal fy(y) rely on matching the mean to the slop and the threshold of the a nonlinear squashing function g(x), then repeating the same process again with the variance of fx(x). The right side of figure (2) is plotted against various weight (w) values, where the optimal weight (wopt) send out the majority of the data.

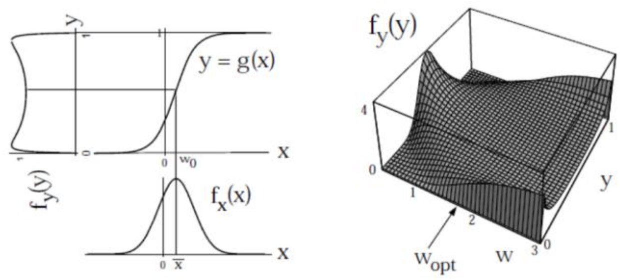

Figure 2: Optimal information flow in sigmoidal neurons[14].

For perfectly matching the probability density of the input x to the nonlinearity slope of g(x), information transfer rate must be maximized. This can be accomplished by applying the optimal weight of w on the input x.

The entropy property can be calculated from the relation of the output y with respect to weight w, who's gradient ascent value can determine the optimal weight of w. The output's joint entropy will be maximized if the case multiple inputs by outputs (MIMO) is found. Hence the chance of statistical independence for individual outputs will increase.

In the case of similarity between the form of the nonlinear transfer function g and the underlying independent components' cumulative density functions (CDF) up to a certain limit, the mutual information between the ui will be minimized due to the maximization of the mutual information between the input x and the output y.

For the weight matrix (W), the update rule for multiple inputs can be given by the equation (3):

$$\Delta W = (I + y' u^T) W \quad \text{where } y' = \frac{\partial}{\partial y_i}\frac{\partial y_i}{\partial u_i} = \frac{\partial}{\partial u_i} ln \frac{\partial y_i}{\partial u_i}. \tag{3}$$

The following logistic transfer function is applied:

$$g(u) = \frac{1}{1+e^{-u}}; \quad \text{, given } y' = (1 - 2y_i) \tag{4}$$

Here, a "sphering" step is inserted in the algorithm before the learning phase. The rows mean values are being subtracted from the dataset X, the later is then applied to a zero phase whitening filters (Wz), which equals the inverse square root of the covariance matrix multiplied by two as shown below:

$$W_z = 2 * \langle XX^T \rangle^{-\frac{1}{2}}. \tag{5}$$

By using this method, the first and the second statistics order of the data can be removed. The values of mean and covariance are set to zero while equalizing the variances. The matrix multiplication between the sphering matrix and the learned matrix, namely WI = W * WZ dedicates the result of the full transform from the zero-mean input. The shape of the ICA algorithm pre-whitening filter is Mexican-hat of retinal ganglion fields of cell receptive, much of the variability caused by lighting effect can be eliminated by using this method.

CLASSIFICATION

Finally, after preprocessing the face image and extracting its features, the time comes for the classification stage to be carried out. The image is to be classified based on the features that been extracted from every image. Support Vector Neural Network (SVNN) is used to carry out this task. This method aims to maximize the margin in the training phase through minimizing the values that are related with the parameters of the classifier model, this way allows to minimize the complexity of classifier without affecting the accuracy of the training data.

Support vector Neural Network (SVNN)

The main idea of the SVNN method isn't based only on the maximal-margin principle but it is also set to avoid nonlinear SVM kernels [15].

The SVM decision function can be given as shown in equation (6) [15]:

$$c(x) = \text{sgn}\left(\sum_{i=1}^{N_{sv}} y_i \alpha_i K(x_i, x) + b\right) \tag{6}$$

Where αi and b are SVM parameters, (xi, yi) is the ith support vector data pair, the output of the sgn function is 1 if the summation result is more than zero and −1 otherwise, and K is a non-linear kernel function. Using the equation above, the classification needs long time in the case of many support vectors. This problem is solved by using SVNN.

To understand the scheme, it is suitable to consider the optimization problem of the soft margin SVM, as shown below:

$$\min_{w,\xi_i}\left(\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i\right) \tag{7}$$

subject to

$$\forall i | y_i (w \cdot x_i - b) \geq 1 - \xi_i \tag{8}$$

$$\forall i | \xi_i \geq 0 \tag{9}$$

Where ŷ is given by equation (13), C1 is a regularization hyper-parameter, also as explained above, yi is the objective class of the ith training case, and ξi are slack variables, which determine the level of misclassification of the vector xi:

$$y_h = \varphi(W_1 \cdot x + b_1) \tag{10}$$

$$\hat{y} = W_2^T y_h + b_2 \tag{11}$$

By replacing the constrained optimization problem of equation (10)-(12) with the corresponding unconstrained optimization problem (14)[18] that has a discrete objective function Φ, which disables the gradient-based optimization techniques. Therefore, a real-coded GA (Genetic Algorithm) is executed to work out for (14) while using Φ as a fitness function [15].

$$\min_{W_1, W_2, b_1, b_2} \Phi \tag{12}$$

Where

$$\Phi = \lambda_{min} + \lambda_{max} + \frac{C_1}{N}\sum_{i=1}^{N} H(y_i \hat{y}_i) \tag{13}$$

and H(t) = max(0, 1 − t) is the Hinge loss. In the algorithm below, the chromosome of each individual is coded into a vertical vector composed by the concatenation of all the columns of W1 with W2, b1, and b2 [16]. The algorithm is started by generating a randomly initial population of Npop individuals in a uniform distribution, according to the Nguyen-Widrow criterion [20]. By passing throughout generations, the fitness value of each individual is computed on the training dataset according to (14) and (15). Then, individuals are graded depending on their fitness values, and a crossover operator is performed to produce new individuals by arbitrarily picking the parents by their grades, depending on the random variable p ∈ [1, Npop] that proposed in [15]:

$$p = \left(N_{pop} - 1\right)\frac{e^{a\vartheta} - 1}{e^{a} - 1} + 1 \tag{14}$$

Where ϑ ∈ [0, 1] is defined as an arbitrary variable with uniform distribution and a > 0 determines the selective pressure, particularly, the larger value of a means bigger likelihood of small values of p, which are linked to high-graded individuals..

EXPERIMENTAL RESULTS AND ANALYSIS

In this section, Experiments using three methods are implemented on Yale database [16] to evaluate the performance of the system. These methods are ICA pulse SVNN, ICA pulse ANN and ICA pulse SVM. The Yale database consists of 165 individual images of 15 persons with 11 different images under different light, exposure, and perspective for each person. The resolution of the images are 243 x 320 pixels with 265 gray levels per pixel, samples of the images are shown in Figure (1):

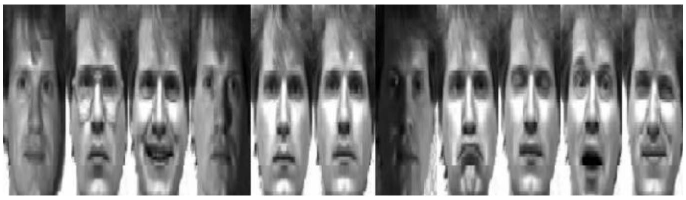

Figure 3: The face images samples from Yale database

Figure (1): The face images samples from Yale database

In the first experiment, The ICA pulse ANN is applied. Whereas ICA is used for features extraction and ANN is used as classifier. The accuracy rate that is obtained from combining these two methods is 93.333 %. While in the second experiment, ICA pulse SVM is used. Here, the accuracy rate is enhanced comparing to the previous experiment, reaching 94.286 %. Finally, ICA pulse SVNN is implemented. As before, ICA is used to obtain the important features that can be used in the following classification stage, while SVNN is used to classify these features and also to give the ability to overcome to the a big amount of time needed by traditional SVM method to classify the large amount of the support vectors. Figure (2) displays the training results.

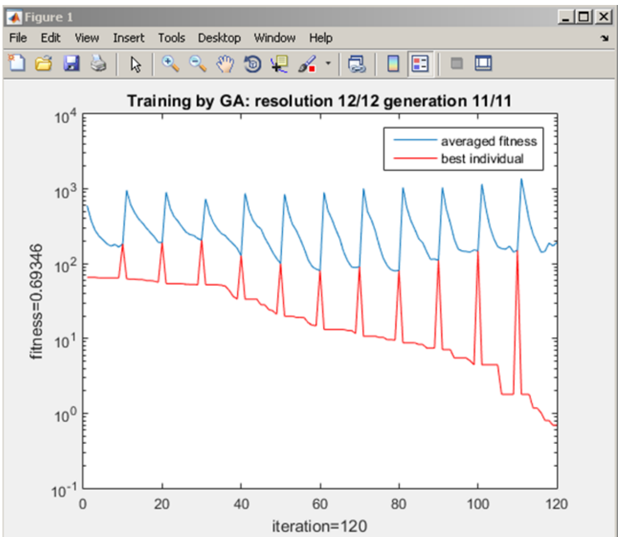

Figure 4: The tanning results of SVNN

By using this method the results enhancement by 0.9523 compare than traditional SVM and 1.9047 compared to the ANN. where the accuracy rate that obtained by this method reached to 95.2380. The table (1) displays the obtained results:

Table 1: The experiments result on Yale database

Method	No. of training Sample	No. of testing Sample	Accuracy Rate
ICA+ANN	60	105	93.3333
ICA+SVM	60	105	94.2857
ICA+SVNN	60	105	95.2380

CONCLUSION AND THE FUTURE WORK

In this paper, a system for face recognition based on ICA and SVNN is proposed. In addition, a comparison among three types of classifiers (ANN, SVM and SVNN) is introduced. The ICA is a very popular feature extraction method that is widely used in image processing applications. A new method called SVNN has been introduced as a classifier. This method is applied to increase the classification margin and also to decrease the time needed by SVM method when classifying large numbers of support vectors. To evaluate the performance, number of experiments are implemented on Yale university face database. The results show an improvement when using ICA+SVNN method that offers a recognition score higher than 95%. In the future work, the powerful ICA's modifications like Kernel ICA will be used. Moreover, other types of methods will be combined, such as K-Nearest Neighborhood, to improve the classification.

REFERENCES:

[1] H. Huang, J. Li, and J. Liu, Enhanced semi-supervised local Fisher discriminant analysis for face recognition, Future Generation Computer Systems 28 (1) (2012) 244-253.

[2] M.Soundararajan and Dr.S.Selvamuthukumaran," Feature Extraction using various Pose Angles in Three Dimensional Face Recognition", International Journal of Applied Engineering Research, Volume 12, Number 21 ,2017, pp. 10797-10801.

[3] Mustafa Zuhaer Nayef Al-Dabagh,"Face Recognition Using LBP, FLD and SVM with Single Training Sample Per Person", in International Journal of Scientific & Engineering Research, Volume 5, Issue 5, May-2014.

[4] Xingang Liu, Lingyun Lu, Zhixin Shen, Kaixuan Lu, "A novel face recognition algorithm via weighted kernel sparse representation", Future Generation Computer Systems, Volume 80, 2018, pp. 653-663,

[5] Vedantham Ramachandran and Dr. E. Srinivasa Reddy," Facial expression recognition with enhanced feature extraction using PSO & EBPNN",

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 11, Number 10 (2016) pp 6911-6915.

[6] G. P. Hegde, Shubhada V. P., M. Seetha and Nagaratna P. Hegde,” Recognition of Expressions Based on Kernel Global and Local Symmetrical Weighted Fisher Discriminant Nonlinear Subspace Approach”, International Journal of Applied Engineering Research ISSN 0973-4562 Volume 11, Number 5 (2016) pp 3558-3569.

[7] Y. Wang, H. Li and Y. Guo, "Face recognition based on ICA and SPSO-ELM," 2017 IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chengdu, 2017, pp. 602-606.

[8] M. M. Karimi and H. Soltanian-Zadeh, "Face recognition: A Sparse Representation-based Classification using Independent Component Analysis," 6th International Symposium on Telecommunications (IST), Tehran, 2012, pp. 1170-1174.

[9] Mustafa Zuhaer AL-Dabagh, Dr. Firas H. AL-Mukhtar."Breast Cancer Diagnostic System Based on MR images Using KPCA-Wavelet Transform and Support Vector Machine", International Journal of Advanced Engineering Research and Science(ISSN : 2349-6495(P) | 2456-1908(O)),vol.4,no. 3, pp.258-263,2017.

[10] Ting-Fan Wu, Chih-Jen Lin, and Ruby C. Weng, “Probability Estimates for Multi-class Classification by Pairwise Coupling”. J. Mach. Learn. Res. 5, 975-1005, 2004.

[11] Tran Son Hai, Le Hoang Thai, and Nguyen Thanh Thuy,” Facial Expression Classification Using Artificial Neural Network and K-Nearest Neighbor”, I.J. Information Technology and Computer Science, vol. 03,pp. 27-32,2015.

[12] X. Zou, J. Kittler and K. Messer, "Illumination Invariant Face Recognition: A Survey," 2007 First IEEE International Conference on Biometrics: Theory, Applications, and Systems, Crystal City, VA, 2007, pp. 1-8.

[13] Venkat R Peddigari, Phanish Srinivasa and Rakesh Kumar, "Enhanced ICA based Face Recognition using Histogram Equalization and Mirror Image Superposition," 2015 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, 2015, pp. 625-628.

[14] Bartlett M.S. (2001) Independent Component Representations for Face Recognition. In: Face Image Analysis by Unsupervised Learning. The Springer International Series in Engineering and Computer Science, vol 612. Springer, Boston, MA

[15] SVNN Oswaldo Ludwig, Urbano Nunes, Rui Araujo, “Eigenvalue decay: A new method for neural network regularization”, Neurocomputing, Volume 124, 2014, Pages 33-42,

[16] http://cvc.yale.edu/projects/yalefaces/yalefaces.html