Face Recognition System Using Independent Components Analysis and Support Vector Neural Network Classifier

Mustafa A. Shihab[1*], *Mustafa* Zuhaer Nayef Al-Dabagh[2*], *Ali* jawad kadhim Alrubaie[3], *Murthad* Al-Yoonus[4], and *Mohammed* Ghazal[5]

[1]Department of Computer Science, College of Science, Tikrit University, Tikrit, Iraq
[2]Department of Computer Science, College of Science, Knowledge University, Erbil, Kurdistan Region, Iraq
[3]Research and Studies Unit, Al-Mustaqbal University College, 51001 Hillah, Babil, Iraq
[4]Computer Science Department, Engineering and Science College, Bayan University, Kurdistan Region, Iraq
[5]Computer Technology Engineering Department, Technical Engineering College of Mosul, Northern Technical University, Mosul, Iraq

Abstract. With an increasing number of security threats in recent years, the field of automatic facial recognition has seen many new developments. The introduction of many new face recognition algorithms focuses on increasing the accuracy rate of the recognition system. This paper introduces a face recognition system using Independent Component Analysis (lCA) for feature extraction and a Support Vector Neural Network (SVNN) for classification. As well as introducing a comparison between SVNN and Support Vector Machine (SVM) and Artificial Neural Network (ANN) classifiers, they are applied to prove the reliability of the proposed method. The implemented experiments use Yale databases, and the results prove that the proposed approach has a higher recognition rate than the (ICA+SVM) and (ICA+ANN) approaches for face recognition.

1 Introduction

Facial recognition systems use computer-coded algorithms to spontaneously recognize or verify a person using digital images or video frames [1]. Many methods compare selected facial image features with a facial database [2]. Typically, this technology pertains to security systems and is analogous to other biometrics solutions, such as fingerprints or eye-iris recognition systems [3].

The PCA method [4] is unquestionably advantageous for feature extraction, but it is better suited for image reconstruction because it does not take into account the various classes of separability. In contrast, the LDA method (Linear Discriminate Analysis) is better than the PCA because it addresses the feature subspace deficiencies of the PCA in terms of separability [5].

---

* Corresponding author: mustafa.zuhaer.nayef@gmail.com

On the other hand, the ICA method provides an improved basis for recognizing high-order relationships between the image's pixels [6]. Most recently, popular machine learning methods use large margin classifiers, for instance, SVM [7]. SVM pattern classification even applies within an ample representation space. Establishing the hyperplanes for pattern classification with that method is imperative. The derivation of the hyperplanes is a ratio of the number of classes. The identification of unseen class elements can be accomplished by calculating posterior probabilities in 1-vs-1/1-vs-rest binary SVMs (Piatt's method) or multi-class SVMs [8].

Training of an ANN uses patterns from the classification process to ascertain the weight collection. ANN is implementing a threshold classification method to define an outside pattern. If the ANN ascertains that the patterns do not apply in any training of an ANN, it uses patterns from the classification process to ascertain the weight collection. ANN is implementing a threshold classification method to define an outside pattern. If the ANN ascertains that the patterns do not apply in any given class, it will report the results of outside classes [9].

The organization of the contents of this paper is as follows: Section 2 introduces the preprocessing, feature extraction, and classification stages, while Section 3 outlines the experiments and analysis of the results. Finally, the conclusions summarise the paper and future work in section 4.

2 The structure of the proposed system

The structure of the proposed system consists of three significant steps: preprocessing, extracting facial features, and classification. Fig. 1 displays the structure of the proposed recognition system in which the ICA method is used to get features from the image, and a new training method named SVNN [9] is applied for classification. Finally, several experiments are sure to be compared with other training methods. The elaboration of these steps is as follows:
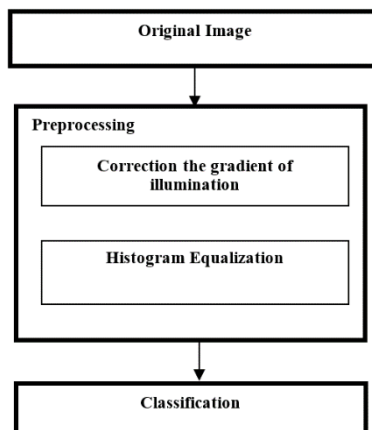

Fig. 1. Face Recognition System Process.

2.1 Image preprocessing

The image preprocessing step aims to reduce the light on the images; therefore, is necessary to implement some preprocessing steps before feature extraction. The preprocessing stage involves two steps:

2.1.1 Correction the gradient of illumination

This step calculates the best-fit brightness value in the image and subtracts this value from the original image's pixels. By using this step, the heavy shadows from extreme lighting angles can be reduced [10].

2.1.2 Histogram equalization

It is utilized to flatten the image's histogram to counterbalance the changes in illumination brightness and the differences in the camera response curves [11].

2.2 Feature Extraction

This stage aims to extract characteristic features from the image in order to employ them in the classification stage. Without applying this stage, the recognition process becomes complicated and impossible because the original face image has a tremendous amount of information, and most of this information is unusable for classification. This stage minimizes the amount of information needed to represent the facial image.

2.2.1 Independent component analysis (ICA)

ICA [12] is one of the unsupervised learning rules derived from the principle of the optimal sigmoid neuron. Let us consider that x is the signal input, y is the output, and g is a nonlinear squashing function.

$$u = wx + w_0 \tag{1}$$

$$y = g(u) = \frac{1}{1 + e^{-u}} \tag{2}$$

This input x is processed by the function g(x). The output density of data depends on three components. The first is the mean, the second is the variance of fx (x) to the slope, and the third is the threshold of a nonlinear squashing function g(x). To maximize information transfer requires deriving the optimal weight (w) on (x), which best agrees with the probability density of x to the nonlinearity gradient; optimal w presents the flattest output density.

The ascending gradient of entropy of the output y concerning weight w derives the optimal weight. Multiple inputs concerning outputs necessitate maximizing the joint entropy of the output, which encourages forward statistical independence of the individual outputs. When the cumulative density functions of the underlying independent components equal the nonlinear transfer function g (up to scaling and translation), an increase in the correlative information between the input x and the output y also reduces the correlative information between the two. Where equation (3) gives the update rule for the weight matrix (W) for multiple inputs:

$$\Delta W = (I + y' u^T) W \tag{3}$$

$$\text{where } y' = \frac{\partial}{\partial y_i} \frac{\partial y_i}{\partial u_i} = \frac{\partial}{\partial u_i} \ln \frac{\partial y_i}{\partial u_i}$$

The logistic transfer function is employed:

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

The algorithm incorporates a "sphering" action before learning [12]. Then, by subtracting the means of the row, the dataset X is derived. Then a zero-phase whitening (Wz) filter is applied to X, which is the inverse square root of the covariance matrix doubled:

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

This method can eliminate the first and second statistical order of the data. Then, by setting the mean and covariance to zero, they equalize the variances. Evaluating the complete transformation from the zero-mean input provides the product of the sphering matrix and the learned matrix, W1 = W * Wz. The lCA algorithm's pre-whitening filter has the Mexican-hat shape of retinal ganglion cell receptive fields; that excludes much of the variability caused by illumination.

2.3 Classification

This step aims to classify the image depending on the features extracted from every image. Then SVNN is used to classify. This method provides a maximal-margin training method and aims to maximize the margin by minimizing the values related to the parameters of the classifier model; this way reduces the complexity of the classifier without affecting the accuracy of the training data.

2.3.1 SVNN

The main idea of the SVNN [9] is not based just on the maximal-margin principle but also avoids the nonlinear SVM kernels [7]. The given SVM decision function can be shown in equation (6):

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

Where αi and b are SVM parameters, (x_i, y_i) is the ith support vector data pair, the values of the (sgn) function are one of the argument is more significant than zero and minus one if it is less than zero, and K is a nonlinear kernel function. Using the equation above, the classification needs a long time when the number of support vectors is large. Using SVNN overcomes this problem. To better understand the method, it is convenient to consider the soft margin SVM optimization problem, as follows:

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

subject to

$$\forall i \, | y_i (w \cdot x_i - b) \geq 1 - \varepsilon_i$$ (8)

$$\forall i \, | \varepsilon_i > 0$$ (9)

Separating hyperplane com poses w and b, where C is a constant, yi is the target class of the ith training case, and ξi are slack variables, which equate to the degree of misclassification the vector xi. The optimization is a trade-off between a high margin (min ||w||2), and a tiny error penalty (min C $\sum_{i=1}^{N}$ ξi).). This approach is trained the NN by solving a similar optimization problem:

$$\min_{w,\varepsilon_i} = \left[\lambda_{min} + \lambda_{max} + \frac{C_1}{N} \sum_{i=1}^{N} \varepsilon_I \right] \qquad (10)$$

Subject to

$$\forall i \, |y_i \hat{y} \geq 1 - \varepsilon_i \qquad (11)$$
$$\forall i \, |\varepsilon_i > 0 \qquad (12)$$

Where y is given by y =, W-2-T., y-h.+, b-2., and C is a regularisation hyperparameter, yi is the objective class of the ith preparation example, and i and are slack variables, as alluded to before, which ascertain the degree of classification error vector xi. By replacing the constrained optimization problem equations (10) and (12) within the equivalent unconstrained optimization problem (13), the objective function ϕ becomes discontinuous, rendering gradient-based optimization methods ineffective. As a result, a real-coded GA (Genetic Algorithm) is implemented to determine (13), utilizing as a fitness function ϕ and H(t) = max (0, 1 t) is the Hinge loss.

$$\min_{w_1, w_2, b_1, b_2} \Phi \qquad (13)$$

Where

$$\Phi = \lambda_{min} + \lambda_{max} + \frac{C_1}{N} \sum_{I=1}^{N} H(y_i \hat{y}_i) \qquad (14)$$

3   Experimental Results and Analysis

Experiments using three methods are implemented on the Yale database [13] to evaluate the system's performance: ICA plus SVNN, ICA with ANN, and ICA with SVM. The Yale database consists of 165 images, with different images under different lights, expressions, and perspectives for each person. The resolution of the images is 243 x 320 pixels with 265 grey levels per pixel.

The ICA with ANN is applied in the first experiment using ICA as a feature extraction method and the ANN as a classifier. The accuracy rate obtained using these two methods is 93.3333. While the second experiment uses ICA with SVM, the accuracy rate of enhancement compares favourably to the previous experiment and reaches 94.2857.

Finally, implementing ICA with SVNN is implemented. ICA is used to obtain the essential features that can be used in the next stage (classification), while SVNN is used to classify these features and overcome the significant amount of time needed by a traditional SVM method to classify a large number of support vectors. This method enhanced the results by 0.9523, which compares favourably to traditional SVM and ANN. Where the accuracy rate obtained by this method reaches 95.2380, table (I) 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

4 Conclusion and the Future Work

This paper proposes the combination of ICA and SVNN algorithms as an optimal model for a face recognition system. This is done by comparing three types of classifiers: ANN, SVM, and SVNN. The ICA method is an accepted feature extraction method widely used in image

processing applications. The new training method called SVNN is introduced as a classifier and improves the other methods. This method is applied to increase the classification margin and decrease the time needed by the SVM method when classifying a large number of support vectors. To evaluate the performance requires the implementation of several experiments on the Yale Face database. The results then improve when applying the ICA with SVNN method, resulting in a more than 95% recognition score. Future work could use powerful ICA modifications, such as Kernel ICA. Moreover, other methods could be combined, such as K-nearest neighbour, to improve the classification.

References

1. P. M. Kumar, U. Gandhi, R. Varatharajan, G. Manogaran, R. Jidhesh, T. Vadivel, *Intelligent face recognition and navigation system using neural learning for smart security in Internet of Things*, Cluster Comput., 22, 4 (2019)

2. M. Z. N. Al-Dabagh, M. I. Ahmad, M. N. M. Isa, S. A. Anwar, *Face Recognition System Based on Fusion Features of Local Methods Using CCA*, in 2020 8th International Electrical Engineering Congress (iEECON), Chiang Mai, Thailand (2020)

3. I. Q. Mundial, M. S. U. Hassan, M. I. Tiwana, W. S. Qureshi, E. Alanazi, *Towards facial recognition problem in COVID-19 pandemic*, in 2020 4th International Conference on Electrical, Telecommunication and Computer Engineering (ELTICOM), Medan, Indonesia (2020)

4. B. M. S. Hasan, A. M. Abdulazeez, *A Review of Principal Component Analysis Algorithm for Dimensionality Reduction*, J. Soft Comput. Data Min., 2, 1 (2021)

5. S. V. Kumar, T. V Rajinikanth, S. V. Raju, *Heart Attack Classification Using SVM with LDA and PCA Linear Transformation Techniques*, Mach. Learn. Technol. Appl. Proc. ICACECS 2020, (2021).

6. V. Vapnik, *Statistical Learning Theory*, John Wiley, Sons, New York, 1998.

7. H. Wang, Y. Shao, S. Zhou, C. Zhang, N. Xiu, *Support Vector Machine Classifier via L0/1 Soft-Margin Los*s, IEEE Trans. Pattern Anal. Mach. Intell. (2021)

8. H. Tran, T. Le, N. Thuy, *Facial Expression Classification Using Artificial Neural Network and K-Nearest Neighbor*, Int. J. Inf. Technol. Comput. Sci., 7 (2015)

9. O. Ludwig, U. Nunes, R. Araujo, *Eigenvalue decay: A new method for neural network regularization*, Neurocomputing, 124 (2014)

10. S. Luo, H. Shen, H. Li, Y. Chen, *Shadow removal based on separated illumination correction for urban aerial remote sensing images*, Signal Processing, 165 (2019)

11. J. Raghavan, M. Ahmadi, *Preprocessing Techniques to Improve CNN based Face Recognition System*, in CS & IT Conference Proceedings, Kanton Zurich, Switzerland (2021)

12. A. I. Abdullah, M. Z. N. Al-Dabagh, M. H. ALHABIB, *Independent Component Analysis and Support Vector Neural Network for Face Recognition*, Int. J. Appl. Eng. Res., 13, 7 (2018)

13. P. N. Belhumeur, J. P. Hespanha, D. J. Kriegman*, Eigenfaces vs. fisherfaces: Recognition using class specific linear projection*, IEEE Trans. Pattern Anal. Mach. Intell., 19, 7 (1997)