International Journal of Applied Mathematics Electronics and Computers

IJAMEC ISSN:2147-8228 http://dergipark.gov.tr/ijamec **Original Research Paper**

Individual Recognition System using Deep network based on Face Regions

Abdelouahab ATTIA*[1], Mourad CHAA[2]

Accepted : 01/08/2018 Published: 30/09/2018

Abstract: Biometric based face recognition is a successful method for automatically identifying a person using her face, with a high confidence. For that reason, this paper introduces an efficient method for face recognition based on deep networks. It considers the three face regions: eye, mouth, and face. First, we have built one sparse autoencoder for every single region their outputs will be concatenated together and fed into another sparse autoencoder. After that, the softmax layer has been employed in the classification step. However, with a deep network method known as the softmax layer has been formed by stacking the encoders from the autoencoder. Followed by formed the full deep network. Finally, the results have been generated on the test set based on the deep network. In the experimental stage, the Yale B database and the AR database and JAFFE database have been used to test the proposed individual recognition system. Experimental findings have clearly proven that the performance of the introduced algorithm is very encouraging and can respond to the security requirements.

Keywords:*Deep network; sparse autoencoder; hybrid face regions; individual recognition system; face recognition*.

1. Introduction

During the past few decades, face recognition it has become an important research field including computer vision, machine learning as well as pattern recognition. Several activities in this field rely on its applications including different areas such as security, criminal identification, surveillance, commercial and credit card verification. However, Facial recognition presents numerous advantages more than other biometric technologies: it is natural, non-intrusive and easy to use. The performance of a person face Recognition system design depends on the step of features extraction that is an essential operation before applying an algorithm of classification.

Various methods have been developed to extract facial features. Mainly, the existing methods in the literature are classified into two categories: in the holistic methods including the Principal Component Analysis (PCA) [1], Linear Discriminant Analysis (LDA) [2], Laplacian faces [3], Multilinear Principal Component Analysis (MPCA) [4], and Independent Component Analysis (ICA) [5]. In the appearance based methods, one of them that have been generally used to extract information appearance that based on the Gabor wavelet [6, 7]. However, the major limitation of Gabor filters is a very large number of features that it generated, it is time-consuming as well it requires convolving face images with a Gabor filter bank in order to extract multi-scale coefficients and multi-orientation. Recently, methods of texture description and classification are known as Local Binary Pattern (LBP) [8] and local quantization phase (PLQ) [9] have been studied to extract facial appearance-based features. a new decorrelated neural network ensemble algorithm for face recognition has been introduced in [10]. However, in this technique, the 2D-FNNs have been used as base components and the negative correlation learning (NCL) scheme is united with the two-dimensional (2D) feed-forward neural networks (2D-NNRW) algorithm to construct ensembles (DNNE 2D-NNRW). The current paper introduces a novel technique for face recognition based on the deep network. Before features extraction techniques step, first the detection of significant information in the face image which captured in other word select the face itself. However, hair, background, and skin have not influence in steps of analysis and identification. Then, the two regions from the face: eye and mouth are extracted from the face selected image. After that, sparse autoencoder (SAE) has been constructed for each distinct region (face, eye, and mouth). The outputs of SAE have been combined together and fed into another sparse autoencoder. Then, the softmax layer has been employed to classify the obtained feature vectors. It is worth to mention that a deep network is formed by stacked the encoders from the autoencoder with the softmax layer. Also with the full deep network formed, the results on the test set have been computed. Finally, the individual recognition system has been tested on the three useful databases Yale B, AR, and JAFFE. The rest of this paper is organized as follows: section two, describe the feature extraction method. And the classification process used in the proposed system, including a brief description of the Softmax classifier and Fine-tuning. Experimental results are given and discussed in section three. The conclusion and further works are drawn in section four.

2. Proposed method

2.1. Feature Extraction

The aim of the current work is the use of the autoencoder so that to extract features vectors from a different region of face cited

[1] *Computer Science Department - Faculty of Mathematics and informatics Mohamed El Bachir El Ibrahimi University of Bordj Bou Arreridj, 34000, Algeria*
[2] *Lab. ELEC – Faculty of new technology of information and communication Ouargla university,Ouargla 30 000, Algeria*
* *Corresponding Author:Email: attia.abdelouahab@gmail.com*

above. However, an autoencoder is an unsupervised learning algorithm refers to an artificial neural network that is trained to reconstruct its input for its output (encoding). Recently, an autoencoder has been used as a tool for dimensionality reduction, image representation and learn deep neural network [11]. Architecturally, the autoencoder is made up of two main parts: the encoder and the decoder. In the current study, three layers have been used: an input layer, an output layer, and one hidden layer.

Given the input to an autoencoder be an image $x \in \mathbb{R}^D$, and the encoder maps of the images x to a different data $z \in \mathbb{R}^Q$ as given in the following formulae:

$$z^{(1)} = f^{(1)}\left(w^{(1)}x + b^{(1)}\right) \qquad (1)$$

However, if $Q < D$, the decoder maps and the encoder representation z followed an estimate of the original input matrix,x, by resolving the following system:

$$\hat{x} = g^{(2)}\left(w^{(2)}x + b^{(2)}\right) \qquad (2)$$

Where the weight matrices are represented by $w^{(1)} \in \mathbb{R}^{D\times Q}$ between the input and the hidden layer (encoder). and the weight matrices between the hidden and the output layer (decoder)are represented by $w^{(2)} \in \mathbb{R}^{Q\times D}$. While the notifications $b^{(1)} \in \mathbb{R}^D$ and $b^{(2)} \in \mathbb{R}^D$ are bias vectors for the hidden and the output layer, respectively. Here,$f^{(1)}g^{(2)}$stand for the transfer function of the encoder and the decoder respectively. In the current work, the Logistic sigmoid activation function has been applied to both the encoder and the decoder respectively. It is calculated by:

$$\log sig(v) = \frac{1}{(1+exp(-v))} \qquad (3)$$

Hence, no labeled data is required; also training process is based on the optimization of a cost function and the training of an autoencoder is unsupervised. However, the cost function generates the error between the input $x \in \mathbb{R}^D$ and its reconstruction at the output$\hat{x} \in \mathbb{R}^D$. Thus, training a sparse autoencoder consists of adjusting the mean squared error function as follows:

$$E = \frac{1}{N}\sum_{n=1}^{N}\sum_{m=1}^{M}(x_{mn} - \hat{x}_{mn})^2 + \alpha \cdot \lambda_{weights} + \beta \cdot \lambda_{sparsity} \qquad (4)$$

Where α and β refer to the coefficients of the regularization term and the sparsity regularization term respectively. In this paper, $\lambda_{weights}$ is called the L_2 regularization term and is calculated as follows:

$$\lambda_{weights} = \frac{1}{2}\sum_{h}^{H}\sum_{j}^{n}\sum_{i}^{m}\left(w_{ji}^{(h)}\right)^2 \qquad (5)$$

Where H refers to the number of hidden layers. The number of observations represented by n and m stands for the number of the training data. The sparsity regularization term $\lambda_{sparsity}$ can be calculated as described in the following formulae:

$$\lambda_{sparsity} = \sum_{i=1}^{D}\xi\log\left(\frac{\xi}{\hat{\xi}_i}\right) + (1-\xi)\log\left(\frac{1-\xi}{1-\hat{\xi}_i}\right) \qquad (6)$$

Where $\hat{\xi}_i$ denotes the average output activation measure of a neuron i that is given by:

$$\hat{\xi}_i = \frac{1}{n}\sum_{j=1}^{n} h\left(w_i^{(1)T}x_j + b_i^{(1)}\right) \qquad (7)$$

In the above equation, η represents the total number of training samples. x_j represents the j the training sample. $w_i^{(1)T}$ and $b_i^{(1)}$are the i row of the weight matrix and the i entry of the bias vector respectively. The parameters of the sparse autoencoder have been empirically selected as $\alpha = 0\cdot 5, \beta = 10$. To form the feature vector based on the proposed method, features have been extracted from face and eyes and concatenated with mouth features. As clearly shown in Figure.1, the obtained feature has been fed into another sparse autoencoder the final hidden feature vector has been used for the classification stage.

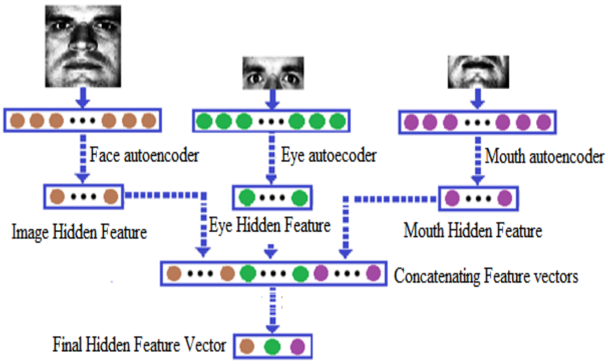

Fig.1 Block diagram of the feature extraction methods

2.2. Classification using the softmax layer

The softmax layer has been used in the classification stage. To train the softmax layer the Scaled Conjugate Gradient back-propagation has been used in this process the weight and bias values have been updated. Training the softmax layer is consist to minimize the cost function so called the mean squared error defined as :

$$E = \frac{1}{N}\sum_{j=1}^{n}\sum_{i=1}^{k}\left(t_{ij} - y_{ij}\right)^2 \qquad (8)$$

Where n is the number of training data; k stands for the number of class; y_{ij} are the outputs of autoencoder; $t_{ij}$are inputs from the target matrix. Finally, a deep network has been formed by stacking the encoders from the autoencoder with the softmax layer. Then the results on the test set have been computed after the full deep network.

3. Experiment results

3.1. Databases and processing

This subsection describes the datasets used in the evaluation of the performances of the proposed method. Three face datasets have been used in the experiments, (i) the Yale B database [12], (ii) the AR database [13] and (ii) Japanese Female Facial Expression (JEFFE) database (http:// www.kasrl.org/jaffe.html).

However, the Yale face database B contains 5760 images taken from 10 participants under 576 viewing conditions that are 9 poses and 64 illumination conditions. For this database 640 images for 10 individuals representing 64 illumination conditions under the frontal pose have been selected.

While the AR face database contains 4000 color images of 126 persons. However, 100 persons have been chosen was separated as 50 male and 50 female. With, 14 images for each person have been selected. All images of this database are converted to gray-level images.

The JAFFE database encloses 213 face images collected from 10 subjects. The resolution of each image is 256×256. We select 20 images for each subject.

Furthermore, face model as shown in Figure 3. Has been used for all databases, each face image has been cropped and resized to remove the background and the hair. In addition, the histogram equalization has been applied to reduce the influence of the illumination variation (see Figure.4).

The examples of eyes and mouths that are used for testing and training set is shown in **Figure 5, Figure 6** and **Figure 7** respectively. The best parameters for face recognition are given in **Table.1**. These parameters are empirically selected.

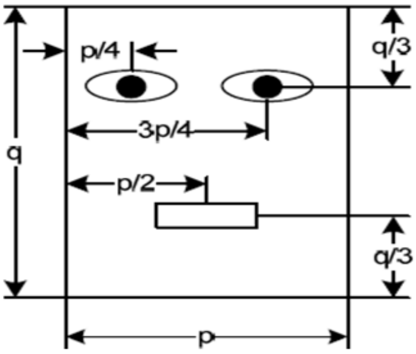

Fig2. Face model to extract the eyes and mouth

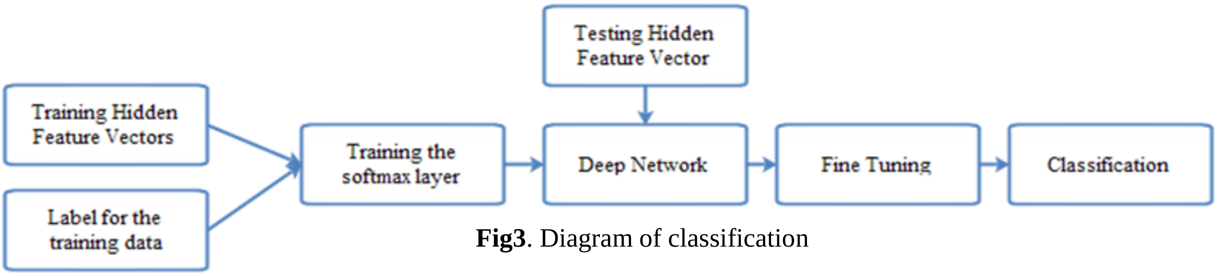

Fig3. Diagram of classification

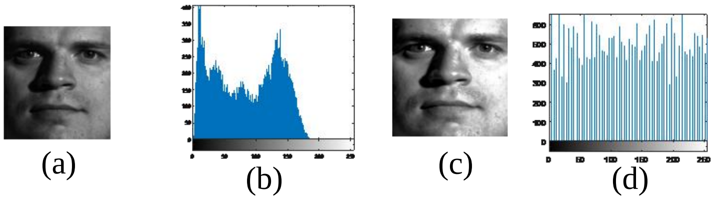

Fig4. : (a) Example image face from Yale B database, (b) their histogram, (c) their images face after equalization histogram, (d) the histogram of the equalized face images

Fig.5: Example of faces from AR database with their localization of eyes and mouth

Fig.6. Example of faces from the Yale B database with their localization of eyes and mouth

Fig.7: Example of faces from JEFFE database with their localization of eyes and mouth

Table1 The best parameters for face recognition purposes.

		Size image	Hidden Size	Max Epochs to train an autoencoder	Max Epochs to train the Softmax Layer
Eyes	**Yale-B**	30*60	200	490	39
	AR	30*60	200	5000	39
	JEFFE	30*60	250	5000	29
Face	**Yale-B**	64*56	200	500	18
	AR	60*43	200	5000	41
	JEFFFE	32*32	250	3000	24
Mouth	**Yale- B**	32*64	200	450	20
	AR	20*60	200	5000	49
	JEFFE	20*60	250	5000	45
Fusion	**Yale-B**	--------	200	10	51
	AR	--------	350	27	1000
	JEFFE	--------	250	200	20

3.2. Results and discussion

In the conducted experiments, for each database, half of the images per class have been chosen randomly for training and the remainder for testing. Therefore, several tests have been performed. The obtained results are point up in **Table 2**. From this table, it is clear that the system accuracy is increased from 95.00% and 90.71% on the basis of the whole face image to 98.13% and 92.71% based on the proposed method for Yale-B and AR databases. Generally, the accuracy has improved with nearly 3% for Yale-B and 2% for AR databases. Using the introduced approach, the testing time is 0.17 second and 0.21 second for Yale-B and AR databases respectively. In case of using faces, the testing time is 0.22 second and 0.35 second. It can be concluded that the proposed system is faster than the system that uses faces. Comparison between the confusion matrix for Yale-B and JEFFE database using face, eyes and our method is shown in Figure .8 and Figure .9 respectively.

Table 2. The recognition rate with the testing time of different facial regions by the proposed method.

Databases		**Face**	**Eyes**	**Mouth**	**Fusion Face Eyes Mouth accuracy**
Yale- B	**RR (%)**	95.00%	97.50%	85.94%	98.13%
	Testing time	0.22s	0.11s	0.21s	0.17s
AR	**RR (%)**	90.71%	89.00%	80.14%	92.71%
	Testing time	0.35s	0.33s	0.25s	0.21s
JEFFE	**RR(%)**	97.00%	96.00%	90.00%	99.00%
	Testing time	0.36s	0.32s	0.30s	0.17s

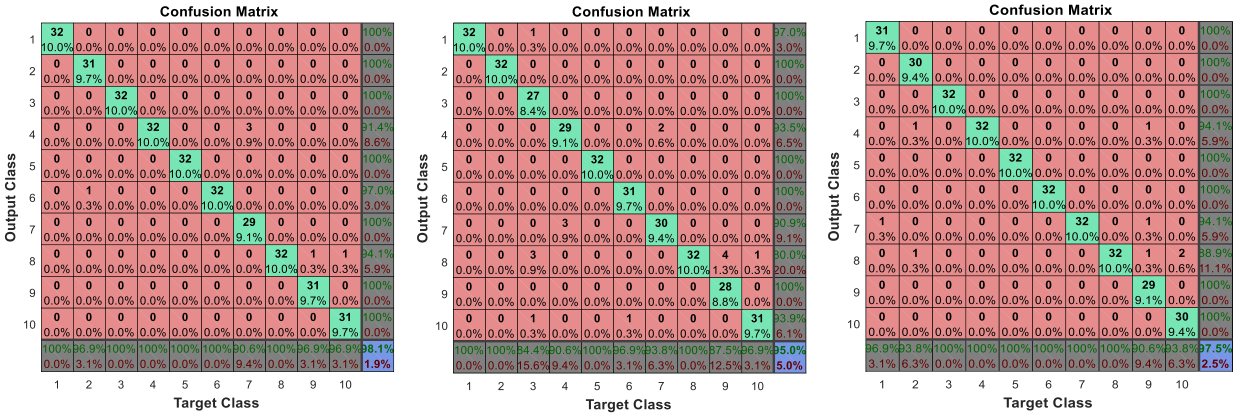

Fig.8 Confusion Matrix for Yale-B database: (a) our method, (b) Face, (c) eyes

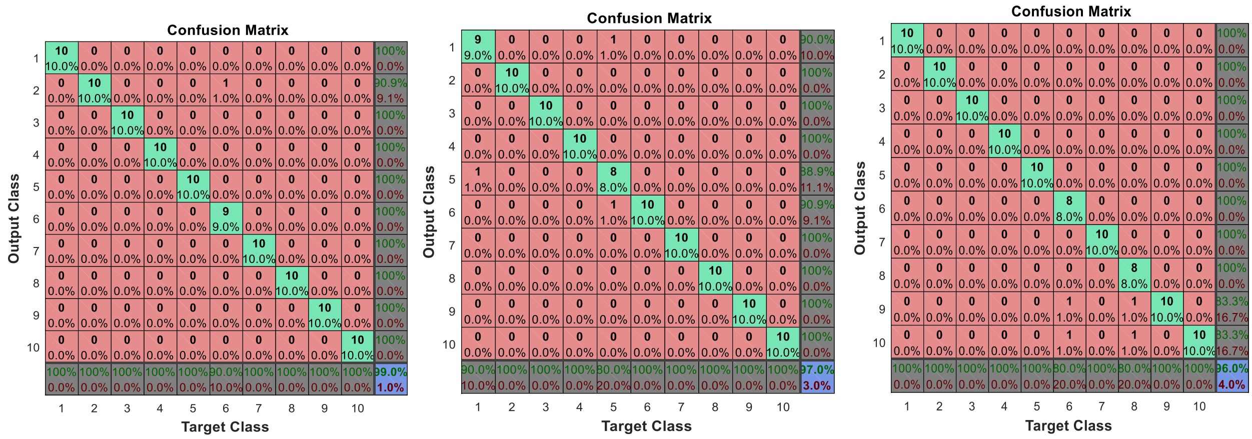

Fig.9 Confusion Matrix for JEFFE database: (a) our method, (b) Face, (c) eyes

Table 3 Comparison of our study with the study in [10,14].

Methods	datasets	Recognition rate
Ref[10]	**AR**	88.60%
	Yale B	97.20%
	JEFFE	100%
Ref[14]	**Yale B**	94.80%
Ref[15]	**Yale B**	82.61%
Ref[16]	**AR**	92.50%
	Yale B	79.30 %
This work	**AR**	**98.13%**
	Yale B	**92.71%**
	JEFFE	**99.00%**

To show the efficiency of the introduced method, a comparison with [10],[14],[15],[16] has been done. This comparison is clearly provided in table 3. It can be viewed that the presented system gives the highest recognition rate for the YaleB and AR databases. For JEFFE database the introduced system achieves a good recognition rate.

4. Conclusion

This paper introduced an efficient face recognition system based on deep networks that take into consideration the three regions of the face. First, after sparse autoencoder built feature from every single region(face, eye, mouth), their outputs will be concatenated together and fed into another sparse autoencoder. The obtained feature vectors from SAE were used as an input to the softmax layer in the classification step. Finally, a deep network was built by stacking the encoders from the autoencoder with the softmax layer. The results were computed on the test set after the full deep network. The evaluation of the performance of the proposed method on Yale B, AR and JAFFE database was realized. From the experimental results, it is clearly noticed that the attained performance is very encouraging and indicating that the introduced system can respond to the security requirements.
Overall, we found that the proposed method is robust for face recognition system. Also for more improvement of the proposed system, other biometrics modalities will be integrated with this modality such as fingerprint or palmprint in order to get a security system with the highest accuracy.

References

[1] Turk, M.; Pentland, A. Eigenfaces for recognition. J. Cogn. Neurosci.1991, 3, 71–86.

[2] Belhumeur, P.N.; Hespanha, J.P.; Kriegman, D. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Trans. Pattern Anal. Mach. Intell. 1997, 19, 711–720.

[3] He, X.; Yan, S.; Hu, Y.; Niyogi, P.; Zhang, H.J. Face recognition

using Laplacian faces. IEEE Trans. Pattern Anal. Mach. Intell. 2005, 27, 328–340.

[4] Lu, H.; Plataniotis, K.N.; Venetsanopoulos, A.N. MPCA: MultilinearPrincipalComponent Analysis of Tensor Objects. IEEE Trans. Neural Netw.2008, 19, 18–39.

[5] Yuen, P.C. and J.H. Lai, Face representation using independent component analysis. Pattern Recognition, 2002. 35(6): p. 1247--1257.

[6] L. Shen, L. Bai, Information theory for Gabor feature selection for face recognition, Eurasip Journal on Applied Signal Processing, in press, doi:10.1155/ASP/2006/30274.

[7] P. Yang, S.G. Shan, W. Gao, S.Z. Li, D. Zhang, Face recognition using ada-boosted Gabor features, in: Sixth IEEE International Conference on Automatic Face and Gesture Recognition, Proceedings, 2004, pp. 356–361.

[8] T. Ahonen ,A. Hadid ,M. Pietikainen, Face Description with Local Binary Patterns: Application to Face Recognition,IEEE Transactions on Pattern Analysis and Machine Intelligence,2006,pp. 2037 – 2041.

[9] T. Ahonen ;E. Rahtu ;V. Ojansivu ;J. Heikkila, Recognition of blurred faces using Local Phase Quantization Pattern Recognition, 2008. ICPR 2008.19th International Conference.

[10] Kankan D, Jianwei Z, Feilong C. A novel decorrelated neural network ensemble algorithm for face recognition. Knowledge-Based Systems.2015; 89 ,541–552

[11] Changjie Hu, XiaoliHou ;YonggangLu ,''Improving the Architecture of an Autoencoder for Dimension Reduction'',Ubiquitous Intelligence and Computing, IEEE 14th Intl Conf on Scalable Computing and Communications and Its Associated Workshops ,pp.855 – 858. 2014.

[12] A.S. Georghiades, P.N. Belhumeur, D. Kriegman, From few to many: illumination cone models for face recognition under variable lighting and pose, IEEE Trans. Pattern Anal. Mach. Intell. 23 (6) (2001) 643–660.

[13] A. Martinez and R. Benavente, "The AR face database," Technical Report, CVC, Univ. Autonoma Barcelona, Barcelona, Spain (1998).

[14] Ahdid R., Taifi K., Fakir M., Safi S., and Manaut B., "Two-Dimensional Face Recognition Methods Comparing with a Riemannian Analysis of Iso-Geodesic Curves," Journal of Electronic Commerce in Organizations, vol. 13, no. 3, pp. 15-35, 2015

[15] XU, Yong, ZHONG, Zuofeng, YANG, Jian, et al. A New Discriminative Sparse Representation Method for Robust Face Recognition via l_2 Regularization. IEEE transactions on neural networks and learning systems, 2017, vol. 28, no 10, p. 2233-2242.

[16] SUN, Ya'nan et WANG, Huiyuan. Face Recognition Based on Circularly Symmetrical Gabor Transforms and Collaborative Representation. In : Multimedia and Image Processing (ICMIP), 2017 2nd International Conference on. IEEE, 2017. p. 103-107.

.