Face Recognition Based on DWT Feature for CNN

Ridha Ilyas Bendjillali
Laboratory of TIT,
Department of Electrical Engineering,
Tahri Mohammed University,
Bechar, Algeria
rayben43@gmail.com

Mohammed Beladgham
Laboratory of TIT,
Department of Electrical Engineering,
Tahri Mohammed University,
Bechar, Algeria
beladgham.tlm@gmail.com

Khaled Merit
Laboratory of TIT,
Department of Electrical Engineering,
Tahri Mohammed University,
Bechar, Algeria
merit1984@gmail.com

ABSTRACT

In the last decade, facial recognition techniques are considered the most important fields of research in biometric technology. In this research paper, we present a Face Recognition (FR) system based on the Viola-Jones face detection algorithm, discrete wavelet transform (DWT), facial image enhancement using histogram equalization (HE) algorithm, and deep convolution neural network. Extraction results of facial features using DWT are used directly to train the CNN network, this network composed of three convolution layers, two pooling layers, a fully-connected layer, and one softmax regression layer. The face recognition rate based on this network is 99.85% and 99.80%. The face recognition rate of the ORL face database and the AR face database based on this network achieved 99.85% and 99.80% respectively.

Keywords

Face recognition (FR), Discrete wavelet transform (DWT), Convolutional Neural Network (CNN).

ACM Reference format:

R.I. Bendjillali, M. Beladgham and K. Merit. 2019. SIG Proceedings Paper in Word Format. In Proceedings of ACM ICIST conference, Cairo, Egypt, March 2019 (ICIST'19), 5 pages.
https://doi.org/10.1145/3361570.3361584

1 INTRODUCTION

In recent years, high security has become a major problem, not only in military environments but also in many civilian applications. Person identification by face recognition gives a number of solutions to this problem. The study of automatic face recognition has begun to be studied since the early 1970s. The work done by Kanade [1] during his thesis is considered the first in this field.

ICIST '19, March 24–26, 2019, Cairo, Egypt

ACM ISBN 978-1-4503-6292-4/19/03…$15.00
https://doi.org/10.1145/3361570.3361584

Since then, a lot of research has been done. A face recognition system is an identification and verification system of individuals that can check whether a person belongs to the system database, and identify if this is the case.

The first step in the facial recognition system is the face detection. Their effectiveness is directly influence on the face recognition system. Face authentication essentially depends on the method used to locate the face in the image. Even if many signs of progress in FR, there are several challenges among them: pose variations, partial occultation's, illumination variations, facial expressions.

The second important step of the FR system is the extraction of facial features. There are mainly three methods for face analysis: first is the local approach, second is global approach, and the third is hybrid approach, the local approach is based on the extraction of local feature points of the face (such as the nose, mouth and eyes and to model the relationship between these points, a topological graph is constructed for each face.

Local Binary Pattern (LBP) [2], Local Gabor Binary Pattern (LGBP) [3], is the most popular method in this approach. In the global approach, all pixels of the face are considered and then, dimensional reduction techniques are generally used to reduce pixel redundancy. Principal component analysis (PCA), Linear Discriminant Analysis (LDA) is the most popular methods in this approach.

Hybrid approaches are approaches that combine global and local features to improve the performance of face recognition. Indeed, local characteristics and global characteristics have quite different properties. We can hope to exploit their complementarity to improve classification. These techniques make it possible to obtain a reduced representation of the original data while preserving the most important feature for the treatments. But these approaches are sensitive to changes in illumination and facial expressions.

Alternative methods for the above mentioned are based on transforms such as Fourier transforms (FT), discrete wavelets transform (DWT) and discrete cosines transform (DCT) [4]. DWT based feature extraction methods are very useful for face recognition with very low computational cost, the wavelet transforms, considering both time and frequency resolutions at the same time which is an ideal tool in many areas such as image processing, and computer vision…etc.

The main contribution of the proposed methodology is to obtain a powerful recognition algorithm with high recognition rate, in this paper we proposed a model by applying Viola-Jones algorithm for two important reasons, firstly to detect faces, secondly to separate

the faces from the non-faces. Moreover, we employed a Discrete Wavelet Transform (DWT) on face images to extract features.
Finally, the classification process will be done by deep learning through convolution Neural Network. Convolutional neural networks are types of Artificial Neural Networks (ANN), which have wide uses in several areas such as image classification, decision-making etc.
The paper is organized as follows. In section II we briefly describe the three research areas related to our work: the steps of face recognition, Deep convolution Neural Network. Experimental results are presented in section III, section IV give the results comparison with other methods, and section V gives the conclusion of this paper.

2 METHODOLOGY

The automatic face recognition is performed in four main steps:

(1) Face detection.
(2) Facial image enhancement.
(3) Extraction of facial features.
(4) Identification and / or verification.

In what follows we will detail each step of the facial recognition system.

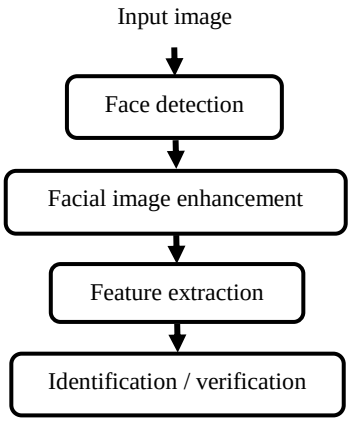

Figure 1: The steps of our approach.

2.1 Face detection using the technique of Viola-Jones

The effectiveness of biometric systems based on face authentication essentially depends on the method used to locate the face in the image. In this regard we use the Viola-Jones algorithm to efficiently detect various parts of the human faces such as mouth, eyes, nose, nostrils, eyebrows, mouth, lips, ears, etc [5].
Paul Viola and Michael Jones in 2001 who came up with the most effective algorithm to detect the human faces and its parts. This algorithm has been implemented in a software 'Matlab' using the method vision. CascadeObjectDetector.
There are 3 important techniques used by Viola - Jones for the detection of facial parts:

1. Haar-like features is of a rectangular type which is used for the feature extraction thus getting an Integral image.
2. Ada boost is artificial intelligence and machine learning method for face detection. The term 'boosted' determines a principle that brings together many algorithms that rely on sets of binary classifiers [2].
3. The third and last step is Cascade classifier that can efficiently combine many features. The term 'cascade' in a classifier determines the several filters on a resultant classifier.

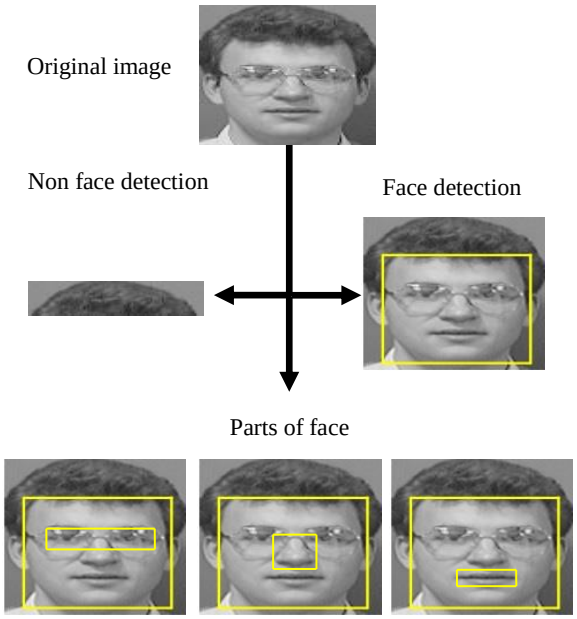

Figure 2: Detection features facial regions.

2.2 Facial image enhancement

Among the most common method used to enhance contrast in the digital image, there is histogram equalization (HE). This method is based on applying a transformation on each pixel of the image, and thus in obtaining a new image from an independent operation on each of the pixels. This transformation is constructed from the accumulated histogram of the original image.
The histogram equalization makes it possible to better distribute the intensities over the entire range of possible values by "spreading" the histogram. Equalization is interesting for images whose whole or only part is of low contrast (the set of pixels are of close intensity). The method is fast, easy to implement, and fully automatic.

2.3 Extraction of facial features by discrete wavelet transform (DWT)

The discrete wavelet transform, considering both time and frequency resolutions at the same time, is considered an extended version of the Discrete Fourier Transform (DFT). Information in different frequency bands plays an important role in the facial recognition system. For facial images, the overall characteristics of the facial images are located in the low-frequency band while all facial details are located in the high-frequency band. Facial expressions can be considered as a facial detail that can affect the frequency components of facial images.
The extraction of features such as eyes, nose, and mouth is a pretreatment step necessary for facial recognition: In this step, we applied the discrete wavelet transform [6].

The discrete wavelet transforms (DWT).consists.in decomposing the signal into different bands or frequency. The involve filters of DWT are known as 'wavelet filter' and 'scaling filter'. The wavelet. filter is high pass filter and other is low pass filter. DWT.performs on different mother wavelets such as Haar wavelet, Daubechies, Morlet, and others [7].

In image processing 2D-DWT is employed which perform operation throughout the rows of the original image by employing both low pass filter (LPF) and high pass filter (HPF) simultaneously [2]. Then down-sampled. by factor 2 and achieved a detailed part (high Frequency) and approximation part (low frequency).

A further operation is performed throughout the columns of the image. Four sub-bands are generated at each decomposition level. In our approach we considered 'db2' wavelet as mother wavelet.

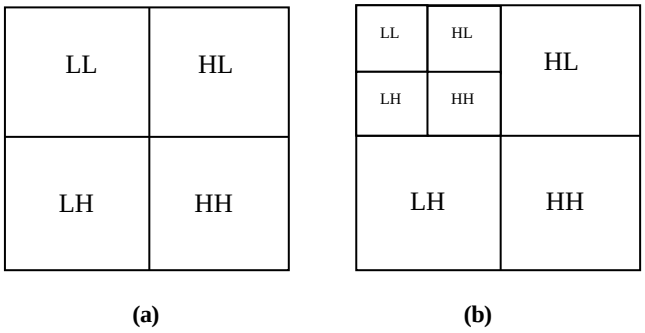

Figure 3: Multilevel wavelet decomposition Multilevel wavelet decomposition consists of the following: (a) single level components specification, (b) two-level components specification, LL—approximation; LH—vertical; HL—horizontal; HH—diagonal detail.

2.4 Classification using Deep Convolution Neural Network

The convolutional neural networks (CNN's) are a subclass of artificial neural networks, primarily are applied to classify images group them by similarity. CNN's are algorithms that can identify faces, character, human pose, tumors, street signs etc [8].

The extraction of local texture features of the human face was performed using DWT, the result is the input of the deep CNN. In this paper, we propose a network structure contains three convolutions, two pooled layers, and one fully connected layer,

The proposed CNN structure is shown in figure 4.

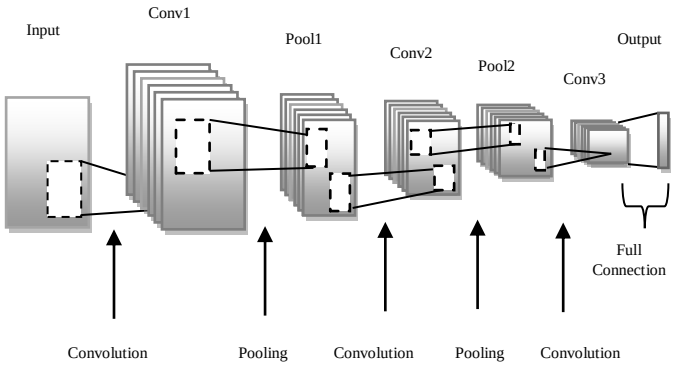

Figure 4: proposed Convolution Network Structure.

1) Convolution layer (ConvL)

The most important operation on CNN is the ConvL, the ConvL performs. the core building block of a Convolutional Network that does most of the computational heavy lifting [9].

Like the traditional neural network, the input of each ConvL is the output of the upper layer [10], on the one hand, in the ConvL, each of the feature graphs corresponds to a convolution kernel of the same size, on the other hand a convolution operation is done between each of the feature maps of the ConvoL is and the feature map of the previous layer [11], after that, we add a bias and then add the corresponding element obtained by function activation.

In this network the size of the convolution kernel C1 is 5x5, and the size of convolution kernel of the base layer C2 and C3 is 3x3, finally we define the size of the latter two convolutions at 3x3 in order to get the better results because two 3x3 increase the network's non-linear capabilities, That's what makes the decision function more discriminative, But if the size of the first layer is of 3x3, this will make the parameters of the entire network model too little, this means a reducing in performance.

The mathematical expression of the layer [12] is:

$$x_i^l = f\left(\sum_{i \in M_j^{l-1}} x_i^{l-1} k_{ij}^l + b_j^l\right) \qquad (1)$$

Where l represents the layer, f represents the activation function, k is the convolution kernel, b is the bias, and Mj represents the feature map.

2) Pooling

The output feature maps obtained after the calculation of the ConvL are generally not much reduced in dimension. If the dimension does not change, there will be a great amount of computation need to do, and the network learning process will become very difficult, more likely to get a reasonable result [12].

The pooling layer is another important concept of CNN's simplifies the output by performing non-linear down-sampling, reducing the number of parameters that the network needs to learn, and don't change the number of feature graphs. In this paper, the pooling layer is sampled with the maximum value. The sampling size is 2x2.

3) Rectified Linear Unit (RELU)

This is the most commonly used activation function in deep learning models, defined as the positive part of its argument, if the rectifier receives any negative input it will return to zero [13]; it is defined as follows:

$$f(x) = \max(0, x) \qquad (2)$$

4) Full-connected layer

For the network, after several convolutions and max-pooling layers, the high-level reasoning in the neural network is done via fully connected layers. Neurons in a fully connected layer have Connections to all activations in the previous layer, and these full-connected layers form a multi-layer perceptron (MLP), which plays the role of a classifier [14].

5) Output layer

The classifier layer is the output layer of the convolution neural network.

In this paper, we used softmax regression classifier [7]. Softmax is a multi-classifier which has a strong non-linear classify ability is used at the last layer of the network, and for a given training we enter the data x.

Where the output category y belongs to {1, 2 ...k}, in total there are k classes, in this network we have set at 7. It is assumed that the input data x is specified, the probability distribution of its class y = i is as follows, θi indicates the parameters to be fitted, the base of the natural logarithm is represented e, and the transpose represented by T, the probability P (y=i | x; θ) is the probability that the input data x corresponds to each class i can take 1 to k.

$$P(C_j = 1 | x) = \frac{e^{w_j^T x}}{\sum_{i=1}^{k} e^{w_j^T x}} \quad (3)$$

3 RESULTS AND DISCUSSION

In this paper, the tests were performed on a personal computer PC 64 bits system with I7 2.4 GHz processor and 8 GB of RAM using MATLAB R2017b.

3.1 The ORL database

In the ORL database (Our Database of Faces) there are 400 face images 10 photos for per person, a total of 40 people, for some subjects, the images. were taken at different times, varying the lighting, facial expressions. (open / closed eyes, smiling./.not smiling) and facial details (glasses./.no glasses) [15].

Figure 5: Partial image from the ORL database.

3.2 Experimental steps of the ORL database

- Firstly, process the pictures from the ORL database as follows: the size of all the images was reduced to 32x32 pixel size.
- After that, normalize the input image data from [0-255].
- Finally, 90% of the human faces in the database were selected as the training set, and the remaining 10% was used as the test set.

3.3 The AR Database

AR face database contains 100 people face images, including 50 men and 50 women, 26 photos for each person, a total of 2600 face images, And the images have face expression change, block change, light change in addition to other changes [16].

Figure 6: Partial image from the AR database.

3.4 Experimental steps of the AR database

- Firstly, process the pictures from the AR database as follows: the size of all the images was reduced to 32x32 pixel size.
- After that, normalize the input image data from [0-255].
- Finally, 90% of the human faces in the database were selected as the training set, and the remaining 10% was used as the test set.

4 COMPARISONS WITH OTHER METHODS

In this section, and in order to prove the effectiveness of our approach, the recognition accuracy is compared with other approaches for face recognition.

Tables 1 and 2 show the comparison of the recognition accuracy obtained with our approach and with other approaches for the ORL and the AR databases.

Table 1. Shows the comparison between different approach and our approach for the ORL face database.

Approach	Recognition rate (%)
Eigenface [17]	97.50
ICA [17]	93.75
2DPCA [18]	98.30
DWT + CNN in this paper	99.85

To check the results, we have obtained, we compared our method with different methods were employed the ORL database such as PCA based Eigenfaces [17], ICA [17], and 2DPCA [18]. Through comparison, we note that the recognition rate of our approach significantly higher than above-mentioned approaches, and very robust against the variation of the light, the facial expressions and the details of the face.

Table 2. Shows the comparison between different approach and our approach for the AR face database.

Approach	Recognition rate (%)
DWT [19]	90.80
PCA+GSRC [20]	97.14
LC-KSVD [21]	97.80
DWT + CNN in this paper	99.80

To check the results, we have obtained, we compared our method with different methods were employed AR database such as DWT [19], PCA+ GSRC (Gabor-feature based sparse representation based classification (SRC)) [20], LC-KSVD (A label consistent K-SVD) [21]. Through comparison, we note that the recognition rate of our approach significantly higher than above-mentioned approaches and again our approach proves that very robust against the variation of the expression, block, and light in addition to other changes.

5 CONCLUSIONS

A face recognition (FR) method based on the Viola-Jones face detection algorithm, discrete wavelet transforms (DWT), histogram equalization (HE) algorithm, and deep convolution neural network (Deep CNN) is presented in this paper. Extraction results of facial features using DWT are the input to training CNN network, this result we used to train the CNN network.

This network consists of three ConvL, two pooling layers, a full-connected layer, and one softmax regression layer to classify and complete face recognition. In experiments on the testing set of the ORL face database and the AR face database, the recognition rate is 99.85% and 99.80% respectively.

Moreover, the algorithm applied in this paper has a high recognition rate, it has great adaptability to the changes in lighting intensity, facial expression and variation of the pose, also the system works efficiently with a small and medium size database

From the above, we conclude that the algorithm used in this paper avoids a big data problem and reduces the cost and the time of training.

6 ACKNOWLEDGMENTS

This work is supported by research project about design, and implementation of a surveillance system based on biometric systems, for the detection and recognition of individuals, and abnormal behaviours (N° A25N01UN080120180002).

REFERENCES

[1] Lihong Zhao, Yu Cai, Xinhe Xu. "Face Recognition Based on Correlation of Wavelet Transform Images", 2006 6th World Congress on Intelligent Control and Automation, 2006.

[2] Rabul Saikia, Aditya Bihar Kandali. "DWT-ELBP based model for face recognition", 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), 2017

[3] H. Nguyen, L. Bai, and L. Shen, “Local gabor binary pattern whitened pca: A novel approach for face recognition from single image per person,” Advances in Biometrics, pp. 269–278, 2009.

[4] M. Wang, H. Jiang and Y. Li “ Face Recognition based on DWT/DCT and SVM” in International Conference on Computer Application and System Modeling (ICCASM 20ID) on, pp.507-510, IEEE, 2010.

[5] Hajar Filali, Jamal Riffi, Adnane Mohamed Mahraz, Hamid Tairi. "Multiple face detection based on machine learning", 2018 International Conference on Intelligent Systems and Computer Vision (ISCV), 2018.

[6] Vikram K and S.Padmavathi, “Facial parts detection using Viola Jones Algorithm,” in International Conference on Advanced Computing and Communication Systems (ICACCS-2015) on, Coimbatore, IEEE 2017.

[7] Bendjillali, R.I.; Beladgham, M.; Merit, K.; Taleb-Ahmed, A. Wavelet-Based Facial recognition. in 6th International Conference on Control Engineering & Information Technology (CEIT), IEEE 2018.

[8] Arel, D. C. Rose, and T. P. Karnowski, “Deep machine learning- a new frontier in artificial intelligence research [research frontier],”Computational Intelligence Magazine, IEEE, vol. 5, no. 4, pp. 13–18,2010

[9] Shijie Qiao, Jie Ma. "A Face Recognition System Based on Convolution Neural Network", 2018 Chinese Automation Congress (CAC), 2018.

[10] Minjun Wang, Zhihui Wang, Jinlin Li. "Deep convolutional neural network applies to face recognition in small and medium databases", 2017 4th International Conference on Systems and Informatics (ICSAI), 2017.

[11] Milla Ferro, Bruno Fernandes, Carmelo Bastos-Filho. "Non-negative Structured Pyramidal Neural Network for Pattern Recognition", 2018 International Joint Conference on Neural Networks (IJCNN), 2018.

[12] K. Yan, S. Huang, Y. Song, W. Liu, N. Fan “ Face Recognition Based on Convolution Neural Network” Proceedings of the 36th Chinese Control Conference, on, pp, 4077-4081, IEEE, 2017.

[13] Bendjillali, R.I.; Beladgham, M.; Merit, K.; Taleb-Ahmed, A. Improved Facial Expression Recognition Based on DWT Feature for Deep CNN. Electronics 2019, 8, 324.

[14] Bendjillali, R.I.; Beladgham, M.; Merit, K.; et al. Facial Expression Recognition Based on DWT Feature for Deep CNN. in 6th International Conference on Control, Decision and Information Technologies (CoDIT), IEEE 2019.

[15] https://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html

[16] A.M. Martinez and R. Benavente. The AR Face Database. CVC Technical Report #24, June 1998.

[17] Yang M.H. Kernel Eigenfaces versus Kernel Fisherfaces: Face Recognition using Kernel Methods. Proceedings of the IEEE International Conference on Face and Gesture Recognition (FG); Washington, WA, USA. 20–21 May 2002; pp. 215–220.

[18] Yang J, Zhang D, Frangi A F, et a1. Two-dimensional PCA: a new approach to appearance-based faces representation and recognition Pattern Analysis and Machine Intelligence.IEEE Transactions on, 2004, 26(1): 131-137.

[19] Priyanka Kumari,"Automated Human Identification and Obstacle Avoidance for Visually Impaired", International Journal of Engineering & Technology, 2018.

[20] Yang M, Zhang L. Gabor feature based sparse representation for face recognition with gabor occlusion dictionary [M].Computer Vision -ECCV 20lO. Berlin Heidelberg: Springer, 2010: 448-461.

[21] JiangG Z, Lin z, Davis L S. Learning a discriminative dictionary for sparse coding via labels consistent K-SVD [C] .2011 IEEE Conferences on Computer Vision and Pattern Recognition (CVPR). Washington-IEEE, 2011: 1697-1704.