Facial Expression Recognition Based on DWT Feature for Deep CNN

Bendjillali Ridha Ilyas\textsuperscript{1}, Beladgham Mohammed\textsuperscript{2}, Merit Khaled\textsuperscript{3}, Abdelmalik Taleb Ahmed\textsuperscript{4}, ALOUANI Ihsen\textsuperscript{5}

\textsuperscript{1,2,3} Laboratory of TIT, Department of Electrical Engineering, Tahri Mohammed University, Bechar, Algeria,

\textsuperscript{4,5} Laboratory of IEMN DOAE. UMR CNRS 8520, University of Valenciennes, Valenciennes, France

1 rayben43@gmail.com, 2 beladgham.tlm@gmail.com, 3 merit1984@gmail.com, 4 Abdelmalik.Taleb-Ahmed@univ-valenciennes.fr, 5 ihsen.alouani@univ-valenciennes.fr

Abstract— Facial expressions recognition have become one of the most important fields of research in pattern recognition, in this paper, we propose a method to identify the facial expressions of the people through their emotions, this method combining Viola-Jones face detection algorithm, Facial image enhancement using histogram equalization, discrete wavelet transform (DWT) and deep convolution neural network. Extraction results of facial features using DWT are the input of CNN, which are used directly to train the CNN network. Our experimental were performed on CK+ database and JAFFE face database, the obtained results based on this network is 96.46 \% and 98.43\% respectively.

Keywords— Facial Expression Recognition (FER); Deep Convolutional Neural Network (deep CNN); Discrete Wavelet Transform (DWT); Histogram Equalization (HE).

I. INTRODUCTION

Facial recognition is currently the most important biometric identification technology; this technique has many advantages, such as low cost, high reliability etc. Facial recognition has been used in several areas such as pattern recognition, Computer vision, security and cognitive science.

In recent years, facial expressions recognition (FER) technique arouses more and more the interest of the scientific community \cite{1}. Facial expressions are an effective way in human and computer machine interaction and a non-verbal interpersonal communications; it has many different applications in various fields such as security-surveillance, artificial intelligence, military and police services, and psychology science etc.

Facial expressions are classified into six basic expressions as follows: Anger, Disgust, Fear, Sad, Happiness, and Surprise, a Neutral expression was added of this group to be the seventh expression.

Facial expression recognition consists of three main steps. The first step is face detection in the image. Its effectiveness has a direct influence on the performance of the FER system. The second important step of a FER system is facial features extraction; Alternative methods are based on transformation such as Fourier transform (FT), short time FT (ST-FT) and discrete wavelet transform (DWT) \cite{8}. Feature extraction based on DWT method is very useful for FER with very low computational cost, which is an ideal tool in image processing and computer vision. The main contributions of the proposed methodology are: first to developing a robust approach of feature extraction, second to improve the performance and speed of FER system and to obtain a high recognition rate. In this paper we proposed a model by applying Viola-Jones face detection algorithm, firstly to detect faces, secondly to separate the faces from the rest of the parts, which are considered non-faces. Moreover we employed Discrete Wavelet Transform (DWT) on face images to extract features.

Finally, the classification process will be done by deep learning through the Convolutional 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 facial expressions recognition, Deep convolution Neural Network. And we present our experimental results in section III and finally, section IV presents the conclusion of the work. Extraction; the third and last step is the expression classification.

In recent years, a wide variety of facial expression recognition methods have been proposed, such as elastic bunch graph matching (EBGM) \cite{2}, Independent Component Analysis (ICA) \cite{3}, Linear Discriminant Analysis (LDA) \cite{4}, and Principal Component Analysis (PCA) \cite{5}, Scale Invariant Feature Transform (SIFT)\cite{6}, embedded hidden Markov models (EHMM) \cite{7}. These methods have improved the recognition rate and speed of facial expressions recognition, but it faces several challenges with regard to variation of pose, changes in illumination, etc.

II. PROPOSED METHODOLOGY

Automatic facial expressions recognition system is performed in three main steps:

(1) Face and facial parts detection.
(2) Facial image enhancement and features extraction.
(3) Expression Classification.

In what follows we will detail each steps of the FER system.
The effectiveness of biometric systems based on face authentication essentially depends on the method used to locate the face in the image. In our method we use the Viola-Jones algorithm to detect various parts of the human faces such as mouth, eyes, nose, nostrils, eyebrows, mouth, lips, ears, etc [9].

While many researchers are trying to reach an algorithm to detect the face, the most effective algorithm to detect the human faces and its parts came up by Paul Viola and Michael Jones in 2001. This algorithm has been implemented in ‘Matlab’ using the vision.CascadeObjectDetector.

There are 3 important techniques used by Viola - Jones for the detection of facial parts:

1- Haar-like features are digital image features of a rectangular type which is used in object recognition.
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 [10].
3- The third and last step is Cascade classifier that can efficiently combine many features and determines the several filters on a resultant classifier.

Among the most common method to enhance contrast in digital image, there is histogram equalization (HE). It consists in 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.

The wavelet is a famous tool in data and image processing, pattern recognition, and has several applications, such as compression, communication, analysis...etc. The discrete wavelets transform, and have the ability to localize a signal in both time and frequency resolutions at the same time, is considered a new generation of the Discrete Fourier Transform (DFT) [11].

The discrete wavelets transform (DWT) decompose the signal at several bands or frequency. The wavelets have two filters are: high pass filter (HPF) and low pass filter (LPF). It is also called the ‘involve filters’ and ‘scaling filter’. The DWT performs on different mother wavelets such as Haar, Symlet, Coiflet, and Daubichi, etc.

The 2D Discrete Wavelet Transform (2D-DWT) is a very important tool in image processing; it works by scanning operation throughout the ranks of the original image by employing both (HPF) and (LPF) at the same time [12]. Then down sampled by factor 2, this process produces detailed part (high frequency) and approximation part (low frequency). After that further operation is performed throughout the columns of image.

Four sub-bands are generated at each decomposition level: an 'approximation' subband (LL), and three 'detail' subbands are vertical (LH), horizontal (HL) and diagonal detail (HH). We considered ‘db2’ wavelet as mother wavelet in our approach.
D. Classification using Deep Convolution Neural Network

The convolutional neural networks (CNNs) are a subclass of artificial neural networks, primarily applied to classify images grouped by similarity. CNNs are algorithms that can identify faces, characters, human pose, tumors, street signs, etc [13].

Through the discrete wavelet transform, the local texture of human face was extracted; this result is the input to the deep convolutional neural network.

In this paper we propose a network structure contains three convolutions, two pooled layers, and one fully-connected layer, the proposed structure is shown in Fig 4.

![Figure 4. Our convolution Network Structure.](image)

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 [10].

Like the traditional neural network, the input of each ConvL is the output of the upper layer [11], 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 ConvL is and the feature map of the previous layer [12], 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 [16] is:

\[
x'_j = f \left( \sum_{i \in M_j} x_{i-1}' k_i + b'_j \right)
\]

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 [16].

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 [14]; it is defined as follows:

\[
f(x) = \max(0, x)
\]

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.

5) Output layer

The classifier layer is the output layer of the convolution neural network; in this paper, we used softmax regression classifier [17]. 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 \( \{1, 2 \ldots k\} \), 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, \( \theta \) indicates the parameters to be fitted, the base of the natural logarithm is represented \( e \), and the transpose represented by T, the probability \( P(y=i|x; \theta) \) is the probability that the input data x corresponds to each class i can take 1 to k.

\[
P(C_j = 1 | x) = \frac{e^{\theta^T x}}{\sum_{i=1}^{k} e^{\theta^T x}}
\]

III. 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 R2018b.

1) JAFFE database

The Japanese female facial expression database is abbreviated as (JAFFE), consists of 213 grayscale images of 10 japanese female models, these images are almost frontal pose and including 7 facial expression images, each image is
of size 256 x 256 [18]. An illustration of the database is shown in Fig. 5.

![Partial image from the ORL database.](image)

- Firstly, process the pictures from the JAFFE database as follows: the size of all the images was reduced to 64x64 pixel size.
- After that, for the normalization of the illumination histogram equalization is used.
- Finally, we used 149 images for Training (about 70% of the total), and 64 images for testing (about 30% of the total).

In TABLE I and TABLE II, N, A, D, F, H, Sa, and Su is used to represent seven basic expressions as Neutral, Anger, Disgust, Fear, Happiness, Sadness, and Surprise respectively.

### TABLE I. SHOW THE RESULTS OBTAINED BY THE JAFFE DATABASE.

|    | N  | A  | D  | F  | Sa | H  | Su |
|----|----|----|----|----|----|----|----|
| N  | 98.3 | 0.2 | 1.0 | 0.2 | 0.0 | 0.1 | 0.0 |
| A  | 0.6 | 98.3 | 0.7 | 0.1 | 0.7 | 0.2 | 0.0 |
| D  | 0.0 | 1.1 | 99.0 | 0.0 | 0.1 | 0.1 | 0.0 |
| F  | 0.2 | 0.5 | 0.5 | 97.3 | 0.5 | 0.2 | 0.8 |
| S  | 0.8 | 0.2 | 0.0 | 0.9 | 98.7 | 0.0 | 0.3 |
| H  | 1.4 | 0.0 | 0.0 | 1.4 | 0.0 | 98.7 | 1.3 |
| S  | 0.0 | 0.0 | 0.0 | 0.3 | 0.1 | 1.0 | 98.7 |

The proposed method provided high recognition accuracy 99% for disgust, 98.7% for surprise happiness and sadness, while the other three expressions Angry, Neutral and Fear had a high accuracy level but less than the previous facial expressions, recognition accuracy ranging between 97.3% – 98.3%. The JAFFE database achieved a recognition accuracy of 98.43%.

2) CK+ Database

The Extended Cohn-Kanade is abbreviated as (CK+) database consists of 593 images in total from 123 subjects that had a human facial emotion based on the subject’s impression of each of the 7 basic emotions [19].

![Partial image from the CK+ database.](image)

- Firstly, process the pictures from the CK+ database as follows: the size of all the images was reduced to 64x64 pixel size.
- After that, for the normalization of the illumination histogram equalization is used.
- Finally, we used 415 images for Training (about 70% of the total), and 178 images for testing (about 30% of the total).

### TABLE II. SHOW THE RESULTS OBTAINED BY THE CK+ DATABASE.

|    | N  | A  | D  | F  | Sa | H  | Su |
|----|----|----|----|----|----|----|----|
| N  | 99.4 | 1.0 | 0.9 | 0.2 | 0.0 | 0.3 | 0.2 |
| A  | 1.1 | 94.4 | 0.3 | 0.0 | 3.4 | 0.3 | 0.0 |
| D  | 0.0 | 2.1 | 97.9 | 0.0 | 0.1 | 0.1 | 0.0 |
| F  | 0.2 | 2.1 | 3.3 | 93.1 | 0.5 | 0.3 | 0.8 |
| S  | 0.1 | 3.8 | 0.0 | 3.7 | 92.5 | 0.0 | 0.0 |
| H  | 0.0 | 0.0 | 0.0 | 1.4 | 0.0 | 98.8 | 2.1 |
| S  | 0.0 | 0.3 | 0.0 | 0.2 | 0.0 | 0.7 | 99.1 |

The proposed method provided high recognition accuracy 99.4% for Neutral and 99.1 for Surprise, 98.8% for happiness, while Angry Disgust sadness and Fear had lower accuracy between 93.1% – 97.9%. The CK+ database gives a recognition accuracy of 96.46%.

This result is good but less than that given by using JAFFE database, this can be explained by: in CK+ database the images were captured in more difficult pose and lighting conditions.
3) Comparison to other methods

In order to prove the effectiveness of our approach the average recognition accuracy is compared to other approaches for FER. The comparison results of the recognition accuracy obtained by our approach and the other approaches for both databases are shown in Table III and Table IV.

| TABLE III. SHOWS THE COMPARISON BETWEEN DIFFERENT APPROACH AND OUR APPROACH FOR JAFFE FACE DATABASE. |
|-----------------|-----------------|
| Approach        | Recognition rate % |
| SVM [6]         | 95.60            |
| Gabor [6]       | 93.30            |
| 2-Channel CNN [6] | 94.40            |
| Deep CNN [20]   | 97.71            |
| Proposed Method | 98.43            |

| TABLE IV. SHOWS THE COMPARISON BETWEEN DIFFERENT APPROACH AND OUR APPROACH FOR CK+ FACE DATABASE. |
|-----------------|-----------------|
| Approach        | Recognition rate % |
| SVM [6]         | 95.10            |
| Gabor [6]       | 90.62            |
| 3DCNN [6]       | 95.00            |
| Deep CNN [20]   | 95.72            |
| Proposed Method | 96.46            |

IV. CONCLUSIONS

This paper presents facial expressions recognition (FER) system based on Viola-Jones face detection algorithm, facial image enhancement using histogram equalization (HE), discrete wavelet transforms (DWT) and deep CNN. Features extraction results of face using DWT is the input to training CNN network, the trained network are used for facial expressions recognition.

This network consists of three convolution layers, two pooling layers, a full-connected layers and one softmax regression layer to classify and complete facial expressions recognition. The results achieved on the JAFFEE and CK+ database, clearly confirms the performance effectiveness and robustness of our method.

In experiments on the testing set of JAFFEE database and CK+ database, the expression recognition rate is 98.43% and 96.46% respectively.

ACKNOWLEDGMENTS

This work is supported by a research project about design and implementation of a surveillance system based on biometric systems for the detection and recognition of individuals and abnormal behaviors (No. A25N01UN080120180002).

REFERENCES

[1] Xiaoguang Chen, Xuan Yang, Maosen Wang, Jiancheng Zou. "Convolution neural network for automatic facial expression recognition", 2017 International Conference on Applied System Innovation (ICASI), 2017.

[2] L. Wiskott, J. M. Fellous, N. Kuiger, and C. V. Malsburg, "Face Recognition by Elastic Bunch Graph Matching", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 775-779, 1997.

[3] M. S. Bartlett, J. R. Movellan and S. Sejnowski, “Face Recognition by Independent Component Analysis”, IEEE Transactions on Neural Networks, vol. 13, no. 6, pp. 1450-1464, 2002.

[4] P. N. Belhumeur, J. P. Hespanha and D. J. Kriegman, "Eigenfaces vs. Fisherfaces", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 711-720, 1997.

[5] M. Turk and A. Pentland, "Eigenfaces for Recognition", Cognitive Neuroscience, vol. 3, no. 1, pp. 71-86, 1991.

[6] Lucy Nwosu, Hui Wang, Jiang Lu, Ishag Unwala, Xiaokun Yang, Ting Zhang. "Deep Convolutional Neural Network for Facial Expression Recognition Using Facial Parts", 2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, pp.1318-1321, IEEE 2017.

[7] D. J. Kim, K. W. Chung and K. S. Hong, "Person authentication using face, teeth, and voice modalities for mobile device security", IEEE Transactions on Consumer Electronics, vol. 56, no. 4, pp. 2678-2685, 2010.

[8] 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 2010) on, pp.507-510, IEEE, 2010.

[9] Paul Viola and Michael Jones. "Robust real-time object detection". In Second international work shop on statistical and computation atheories of vision, Vancouver, Canada, July 13 2001.

[10] 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.

[11] Priyanka Dharani, Dr. A S Vibhute “ Face Recognition Using Wavelet Neural Network” Volume 7, Issue 5, May 2017 International Journal of Research in Computer Science and Software Engineering.

[12] A. Benzouzi, A. Boukrouche, H. Doghmane, and H. Bourouba,”Face recognition using 1dlbp, dwt and svm,” in Control, Engineering & Information Technology (CEIT), 2015 3rd International Conference on, pp. 1–6, IEEE, 2015.

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

[14] 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.

[15] M. Wang, Z. Wang, J. Li “Deep Convolutional Neural Network Applies to Face Recognition in Small and Medium Databases” in 4th International Conference on Systems and Informatics, on, pp. 1368-1378, IEEE, 2017.

[16] 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.

[17] 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

[18] Coding Facial Expressions with Gabor Wavelets Michael J. Lyons, Shigeru Akamatsu, Miyuki Kamachi & Jiro Gyoba Proceedings, Third IEEE International Conference on Automatic Face and Gesture Recognition, April 14-16 1998, Nara Japan, IEEE Computer Society, pp. 1378, IEEE, 2017.

[19] 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

[20] L. Nwosu, H. Wang, J. Lu, I. Unwala X. Yang and T. Zhang “Deep Convolutional Neural Network for Facial Expression Recognition using Facial Parts” 15th International Conference on Dependable, Autonomic and Secure Computing, on, pp.1318-1321, IEEE 2017.