Recognition of Human Facial Expressions Using DCT-DWT and Artificial Neural Network

Wasan. Maddah Alaluosi

Ministry of Education, Iraq

Abstract

   Facial expressions are a term that expresses a group of movements of the facial fore muscles that is related to one's own human emotions. Human–computer interaction (HCI) has been considered as one of the most attractive and fastest-growing fields. Adding emotional expression's recognition to expect the users' feelings and emotional state can drastically improves HCI. This paper aims to demonstrate the three most important facial expressions (happiness, sadness, and surprise). It contains three stages; first, the preprocessing stage was performed to enhance the facial images. Second, the feature extraction stage depended on Discrete Wavelet Transform (DWT) and Discrete Cosine Transform (DCT) methods. Third, the recognition stage was applied using an artificial neural network, known as Back Propagation Neural Network (BPNN), on database images from Cohen-Kanade. The method was shown to be very efficient, where the total rate of recognition of the three facial expressions was 92.9%.

Keywords: Facial expressions, Features extraction, DCT, DWT, Back Propagation Neural Network (BPNN).

تمييز تعابير الوجه باستخدام التحويل المويجي وتحويل الجيب تمام المنفصل والشبكة العصبية الذكية

وسن مداح حسن

وزارة التربية ، العراق

الخلاصه

   تعابير الوجه هو مصطلح يعبر عن حركات عضليه للوجه تتعلق بما يمتلكه من مشاعر البشرية. يعتبر التفاعل بين الإنسان والحاسوب (HCI) أحد أكثر الحقول جاذبية وأسرعها نموًا. يمكن أن تؤدي إضافة التعرف على التعبير العاطفي لتوقع مشاعر المستخدمين والحالة العاطفية إلى تحسين التفاعل بين الإنسان والكمبيوتر (HCI) بشكل كبير. تهدف هذه الورقة إلى التعرف على أهم ثلاث تعابير للوجه (السعادة، والحزن، والمفاجأة)، وتحتوي على ثلاث مراحل، وهي المعالجة المسبقة من أجل تحسين صور الوجه. تعتمد مرحلة استخراج الميزات على طريقة تحويل الموجة المنفصلة (DWT) وطريقة تحويل جيب التمام المنفصل (DCT). في مرحلة التعرف على تعابير الوجه تم استخدام الشبكة العصبية الاصطناعية المعروفة باسم الشبكة العصبية للانتشار الخلفي (BPNN) وباستخدام صور من قاعدة بيانات من CK. كانت الطريقة فعالة للغاية. بلغ معدل التعرف على تعابير الوجه الثلاث الإجمالي للبحث 92.9٪.

*Email: wamalousi@gmail.com

1. Introduction

It is well known that studying and recognizing human facial expressions is a powerful and challenging task in social communication since it focuses on studying human condition, emotions, and reactions, which in turn assists in deciding whether the information given was true or not. Facial expressions are useful in determining the conceded emotions which need to be verified. Also, they are used in the diagnosis of psychopathological stress level. Facial Expressions Recognition (FER) plays a crucial role in computer graphics where human faces have to be modelled and parameterized as computer animations by accurately characterizing face geometry and muscle motion. FER involves several technologies, such as virtual reality (VR) and augmented reality (AR) [1, 2].

Facial expressions transmit signals about the expresser's emotions, intentions, and environment and it plays a key role in successful social interaction [2, 3].

Facial expressions are signals of high biological value. It is important to consider how well these signals can be recognized or detected across different viewing conditions. Positive facial expressions, such as happiness and pleasure, demonstrate healthy emotion states, while unhealthy emotion states are represented by expressing negative facial expressions e.g., sadness and anger [4] .

2. Literatures work

Many previous studies have focused on investigating several aspects of face expressions.

M. A. Jaffar and E. A. Aleisa 2014, presented a method for the   classification of the facial expressions using the local features extracted, such as the DCT, wavelet, Haar transform, and Gabor wavelet -based functions. These functions were provided for an (SVM) Supports Vector Machine learning classification algorithms. The three datasets used in this case were the Multimedai Understanding Group (MUG) dataset , MMI facial expressions dataset , and Japanese Female Facial Expression (JAFFE) . Facial expression accuracies were shown to have the values of 69.02 %, 66.60 %, and 85.32 % for these data bases, respectively [5].

Mayya *et al,* 2016, used a method for automatically recognizing facial expressions using Deep Convolutional Neural Network (DCNN). The proposed model recognized expressions for 209 images with an accuracy values of 98.12 % for JAFFE database and 96.02 % for CK database. [6].

Revina and Emmanuel in 2018,introduced a review paper which discussed the analysis of various FER techniques. This survey compared algorithms based on preprocessing, feature extraction, classification, and major contributions. The highest recognition accuracy of 99 % was provided by the SVM classifier which recognized several expressions. JAFFE and CK were shown to have higher efficient performance than the other databases[7].

Yanga *et al.* in 2018, proposed an emotion recognition model. This method used Haar Cascades to detect the input image. This proposed model introduced emotion recognition based on facial recognition in virtual learning environments. The results showed accuracy of 95.25 % for happy expression and 78.54 % for sad expression[8].

Abbas, A. R., & Kareem, A. R in 2018, applied a machine learning technique for intelligent age estimation from facial images using J48 classifier on FG_NET dataset. The research involved three stages; first, the preprocessing which included noise removal, face detection, and image size normalization. Second, feature extraction which depended on Linear Discriminant Analysis (LDA). Third, the classification which depended on the C4.5 decision tree. The results showed that J48 classifier had an accuracy of 89.13 % [9].

In addition, FER has been used in applications such as driver state monitoring, medicine and forensic medicine, robot industry, and deception detection [10]. The types of emotions were classified by Ekman and Friesen [2]. into six basic expressions (happiness, sadness, fear, disgust, anger, and surprise). This research discusses and focuses on three expressions: happy, Surprise, and sad,. The emotional facial expressions are numerous, even with some expressions that are composed of a combination of more than one [11, 12]. They are affected by several factors, such as age. The main challenge that the recognition systems encounter is the type of images and the distortion they have. This research contains five sections. Section 1 includes the introduction. Section 2 presents the related works. The details of the proposed system and the structure of the project are introduced in section 3. The experimental results and their discussion, as well as the details of the database, are presented in section 4. Finally, the conclusions are introduced in section 5.

3. The proposed system

The components of the proposed system are demonstrated inn Figure-1. The system is organized to perform the following steps:

1. Reading images from database .
2. Preprocessing of images.
3. Feature extraction: this step implements the Discrete Wavelet Transform, which is a technique used to transform image pixels into wavelets. It divides the image into four parts, LL, LH, HL, and HH . The Discrete Cosine Transform (DCT) was implemented with a block of (8×8) to create feature vector for facial expressions images.
4. Facial expression recognition: The highest 30 values representing the features inserted as input to the neural network were taken and, thus, the results were obtained.
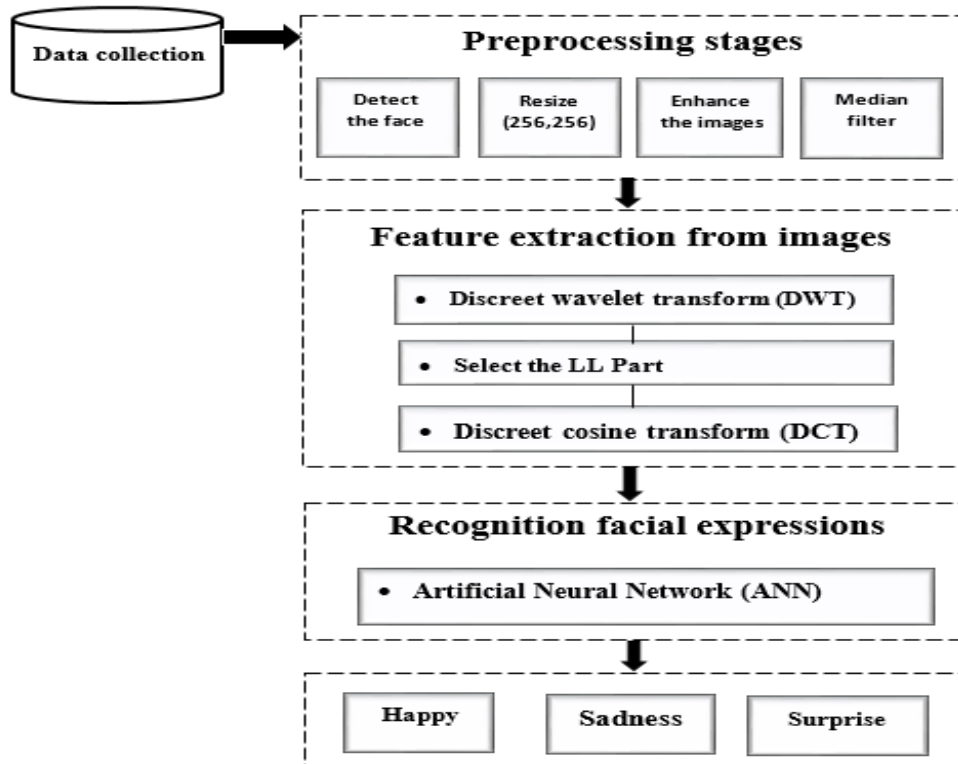

Figure 1- Facial Expression Recognition system block diagram

3.1 Preprocessing

Preprocessing is a multistep process to extract the face expressions from the images, reduce the computational time, and enhance the images in terms of noise. It includes enhancement and edge detection, which is used to identify the properties of the edges contained in the image, thus enabling its analysis improvement. It is needed to remove all noise from the image, leaving the perceptual fidelity of the visual information [13]. Accordingly, the image was improved by reducing the noise (salt & pepper) and converting it to a binary system image in order to focus only on the main data in the image. This process involves several steps;

Sharpening can improve images quality and reduces its size. Sharpening is done using unclear copy from the original image then unsharp masking was Implemented whith radius 10 and the amount1. An unsharp filter is an operator used to sharpen an image.

Cropping and clipping includes the elimination of the unwanted white space around the face in the background. This process was performed by calculating the number of white pixels from every direction (top, down, left, right) and implementing the median and mean filters on the cropped images. The images were then resized to (250 , 250).The image was enhanced from salt and pepper noise using the median filter.

3.2 Feature extraction

Feature extraction is a very important process that helps avoiding the overlapping and redundancy of

the features and information. Extraction of the features from images of facial expressions is a very difficult process that needs to be very accurate. DWT can divide an image into four areas; that is, after the first scale of decomposition in the forward DWT, coefficients from four frequency sub-bands are obtained (LL, LH, HL, HH). LL component has enough energy that is used to recognize a face image [14]. For an image A, the size is A (m, n). In the horizontal filtering, the size of the image is reduced to [n×(m/2)]. After further filtering and down-sampling in the vertical direction, four subimages are obtained, each being of size [(N/2) × (M/2)]. H represents the high-pass filter and L represent the low-pass filter. The LL part contains the most useful and necessary information. The equation is presented below[14]:

$$WL(n, k) = \sum_p WL(p, k + 1)h'(n - 2p) + \sum_l WH(l, k + 1)g'(n - 2l) \qquad \ldots (1)$$

where the frequency index is denoted by k, the scaling coefficient is W_L (p,q)m while the wavelet coefficient is W_H (p,q). Also, h' (n) is the low-pass filter and g'(n) is the high-pass filter.
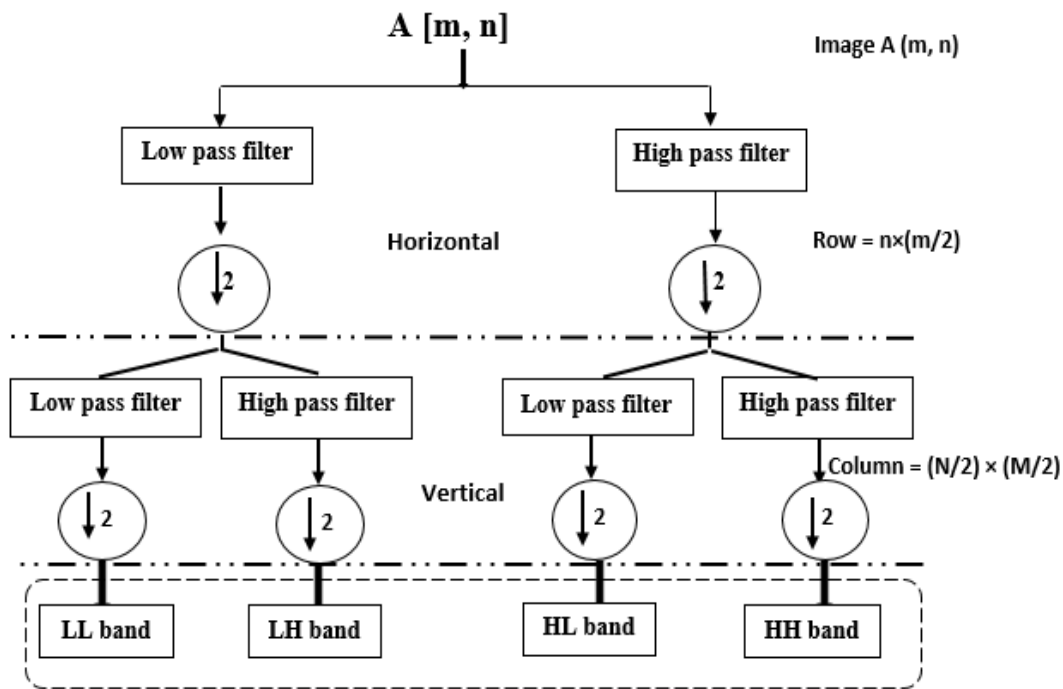

Figure 2-Block diagram to extract the features using Discreet Wavelet Transforms

The DCT converts images from time (spatial) domain to frequency domain to de-correlate pixels [15].In 2D-DCT N_2 pixels represent wide while N_1 pixels high. A (i, j) is the intensity of the pixel in a row i and column j; B (k_1, k_2) is the DCT coefficient in a row k_1 and column k_2 of the DCT. Each image is divided into sub blocks with an (N x N) size matrix [15]. Each element in the DCT matrix represents a frequency in the image. DCT is characterized by having the greatest value in the upper left corner, while reducing the values of the coefficients to the right side or the bottom. The coefficients near the upper-left corner of the conversion matrix represent low-frequency compounds (smoothed image), while the remaining coefficients are edges. The low components have more useful information about the image. The high components can be neglected [14]. In DCT the block size of 8 x 8 it obtain 64 basis images thruogh computing equation (2).

$$_B(K_1, K_2) = \sum_{i=0}^{N1-1} \sum_{j=0}^{N2-1} 4.A(i,j).cos \left| \frac{x.k1}{2.N1} (2i + 1) \right| cos \left| \frac{x.k2}{2.N2} (2.j + 1) \right| _{---(2)}$$

The operation of the DCT includes [15]:

- The input image is ($N_1 \times N_2$) size of matrix; each image is divided to sub blocks.
- A (*i* , j) is the pixel in a row *i* and column j;
- B (K_1, K_2) is the DCT coefficient in a row k_1 and column k_2 of the DCT matrix. For the images, much of the signal energy lies at low frequencies; these appear in the upper left corner of the DCT.

The DCT input is an 8 by 8 array of integers. This array contains each pixel's gray scale level. The numeric values of the features represent the values of pixels (x,y) in each image that is as an input in a neural network.

1.2Recognition of facial expressions using train neural network Backpropagation neural network (BPNN) is used at each layer to achieve the minimum error possible between the layer's response and the actual data [16]. Back propagation algorithm is used to enhance the neural network. The process of learning includes building a neural network with three layers, namely the input, the hidden, and the output layers.

Neural network training is about finding weights that reduce the prediction error. The training starts with a set of randomly generated weights. Then, backpropagation is used to update the weights in an attempt to correctly map the arbitrary inputs to outputs. Backpropagation means the 'backward propagation of errors', which is used to update the weights using a gradient descent. It calculates the gradient of the error function with respect to the neural network's weights.

The calculation proceeds backwards through the network. To update weight, we take the current weight and subtract the partial derivative of error function with respect to weight. Optionally, we multiply the derivative of the error function by a selected number to make sure that the new updated weight is minimizing the error function; this number is called (learning rate) [19].

Backpropagation is organized in the following steps.

- Initial weight is calculated as: *Wi = w1, w2…wn.*
- Weights and inputs are used to predict the output; output= $wi * x$. The results are then passed forward to the next layer.
- Error =*1/2* (production- actual output)**2.**
- The actual outputs of the neurons working under the hidden layer are calculated as follows:

$$actual\ output = a\left(\frac{\partial Error}{\partial wi*x}\right)\ldots\ldots(3)$$

$$Wx = wi.x - a\left(\frac{\partial Error}{\partial wx}\right)\ldots\ldots\ldots(4)$$

$W = weight,\ x = input,\ a = learning\ rate.$

4. Experimental Results

The proposed system consists of the following stages: data collection, preprocessing, feature extraction, and recognition of facial expressions images. Each stage is subdivided into a number of processes. In the preprocessing stage, the median filter is used to improve the image and remove noise. After that, the images will be resized to (250 ×250). The region of the face, which will be represented by a box, is cropped from the image without affecting the background. Features of facial expressions can be obtained through the implementation of the wavelet method, which converts the frequency domain to the spatial domain. This is possible to implement with high-resolution. In this research, DCT was implemented after DWT wavelet. All these processes are illustrated in Figure-3.
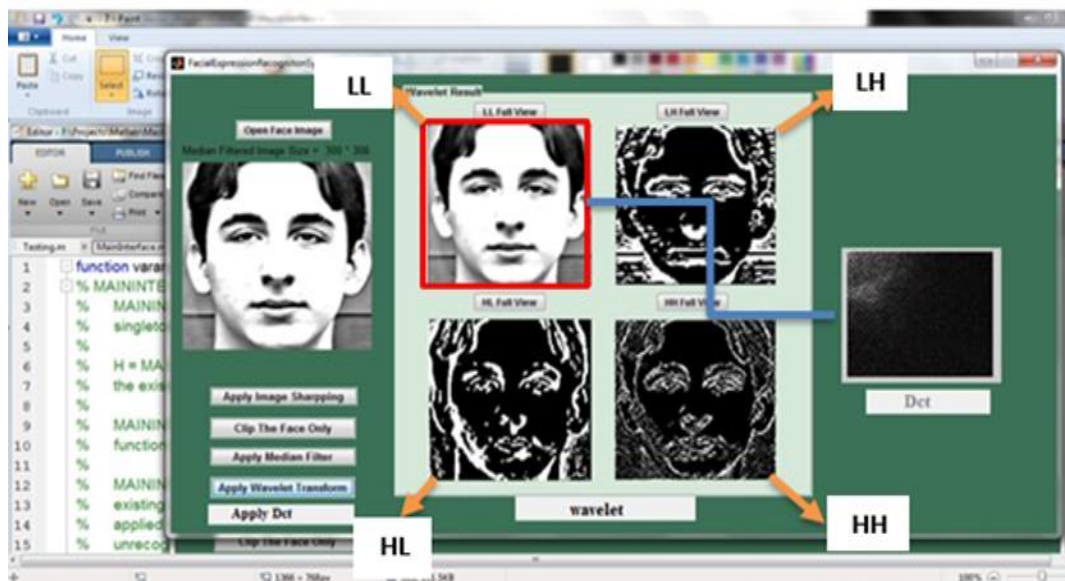

Figure 3- Examples of implemenation DWT and DCT

The numeric values of the features in the figure are listed in Table-1, which shows an Excel database that contains 60 values representing the [x, y]. The top 30 values were selected, being obtained from the application of wavelet transformation. Through the use of the wavelet transform method to extract the features, 30 features were extracted and selected to reflect the best features, because the usage of 20 and 60 features resulted in confusion in the training of the network.

Table 1-illstratate of Numaric Value of Feature (x,y).

	x	y	x	y	x	y	x	y	x	y
image1Happy 1.png	108	56	147	109	99	42	47	44	40	60
image1 Happy 2.png	108	40	103	41	47	43	100	41	141	107
image1 Happy 3.png	50	44	52	57	106	41	87	90	42	45
image1Happy 4.png	50	44	52	57	106	41	87	90	42	45
image1 Happy 5.png	91	89	98	42	48	44	138	106	105	41
image1 Sad 1.png	90	91	107	41	109	41	49	44	145	106
image1 Sad 2.png	46	44	106	57	53	58	86	94	97	41
image1 Sad 3.png	41	60	107	41	61	94	86	95	49	44
image1 Sad 4.png	41	60	107	41	61	94	86	95	49	44
image1 Sad 5.png	103	41	146	106	39	45	43	45	41	56
image1 Surprise 1.png	108	41	43	45	66	93	52	44	44	55
image1 Surprise 2.png	104	41	49	44	87	91	108	41	85	93
image1 Surprise 3.png	138	107	140	107	97	42	52	44	45	57
image1 Surprise 4.png	138	107	140	107	97	42	52	44	45	57
image1 Surprise 5.png	103	51	48	44	88	91	98	41	74	122

The highest 30 value which represent the highest 30 pixels were selected and stored in an Excel database, where they are called as the input of the network. The initial network has 2 hidden layers with 3 neurons, and was trained three times using a different initial weight in each time. The best results of the 3 training sessions were recorded. The outputs of these networks were matched with the person images using three expressions (happy, sad, and surprised). The best training performance was at epoch 25, which represents the performance at the 2.40-08 and the gradient at 5.22e-07. The results of the recognition system using the neural network are shown as a figure. A total of three hidden layers was used, the first layer consists of 30 neurons, the second layer consists of 15 neurons, while the output layer produced 3 neurons that represent the three facial expressions.

The number of images was 90 images for 6 people and 3 expressions, with each expression being represented by 5 images. The training phase contained 75 images, while the testing phase contained 15 images, with each expression being represented by 5 images. Table-2 represents the output results of the neural network for each expression.

Table 2- illustrates the Output of Neural Network

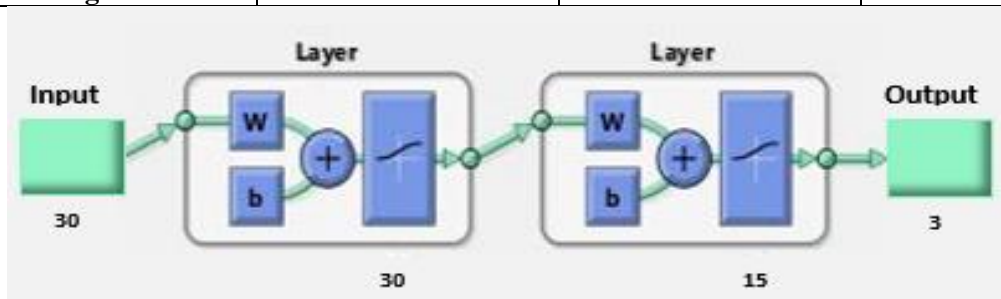| Excpressions	Node of Surprised	Node of Sad	Node of Happy
Happy	0	0	1
Sad	0	1	0
Surprises	1	0	0
No recognition	0	0	0

Figure 4- a) Results of the Recognition using Neural Network with two hidden layers , b) The best training performance at epoch 25.

To calculate the results of recognition accuracy, the following equation was applied:

$$\text{Accuracy of Recognition } (\%) = \frac{no.of\ corrct\ Recognition}{Total\ number\ of\ images} \cdot \qquad \dots (5)$$

Table 3- illustrates the Results of Facial Expressions Recognition with Comparison between the above works and the proposed method.

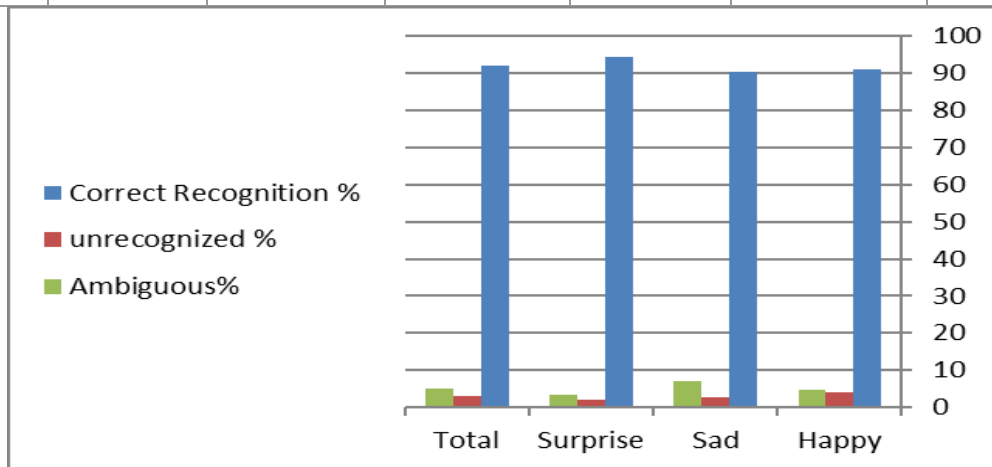| Facial expression	DWT-DCT- Backpropagation			DWT- Backpropagation		
	Correct Recognition	Unrecognized	Ambiguous %	Correct Recognition	Unrecognized	Ambiguous %
Happy	91.12%	4.18%	4.7%	88.20%	4.2%	7.6
Sad	90.31%	2.63%	7.06%	85.32%	5.3	9.38
Surprised	94.5%	2.03%	3.47%	89.18%	6.7	4.3
Total	91.976%	2.946%	5.067%	87.5%	5.4%	7.1%

Figure 6-Represent the accuracy of each expressions

By using only the discrete wavelet transforms for extracting the features, the percentage of

recognition was 87.5 %. The average facial expression recognition of the proposed method (discrete wavelet transforms – discrete cosine transforms and backpropagation) was 91.9 %. Figure-7 explains the comparison between recognition rates.
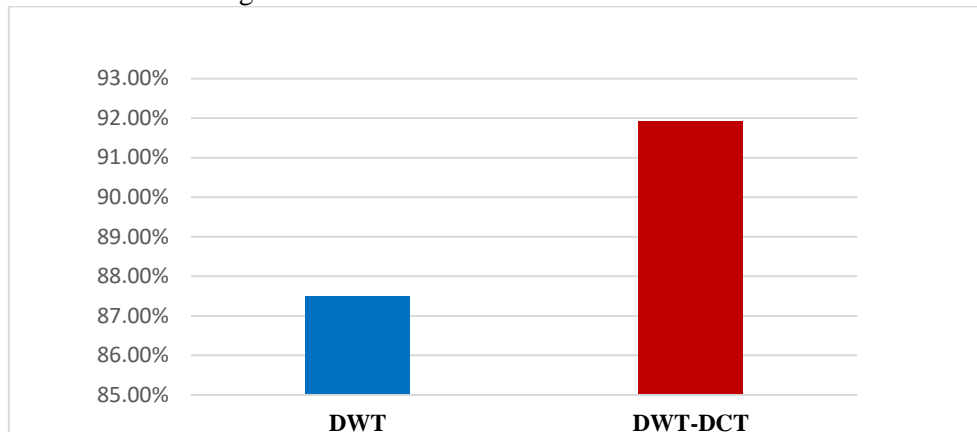

Figure 7-Represent comparison of recognition rates

The database

The recognition of facial expressions was implemented by Matlab (R2016 a) program based on Cohn- Kanade CK database. Images of 10 persons were saved in an PNG format. 90 images were taken, with each person being proposed by 9 different photos. Each image had the size of (640 × 490) pixels with 8 gray levels. The database was divided into training (70 images) and test (20 images) sets.

Figure 1-Database from Cohn- kanade, a:- surprised , b:-sad, c:-happy

5. Conclusions

In this paper, system recognition of facial expressions was applied depending on Cohn –Kanade dataset which consists of 90 images taken for a group of 10 persons, with three expressions (surprised, happy, and sad). Wavelet transformation methods were used to extract features from images under consideration, which are represented by 30 features, rom which the top 30 values were extracted for each image. Adding a DCT led to improving the extraction of the important features and, therefore, led to increase the proportion of recognition of facial expressions. The numeric value of the features represents the value of pixels (x,y) in each image, which is as an input in a neural network for recognition of facial expressions by using BPNN. This approach is very efficient. The values of accuracy of recognition for each expression were as follows: 91.12% for happy, 90.31% for sad, and 94.5 % for surprised. As a result, the total rate of recognition was 91.7%.

References

1. Al-Rawi, Salah Sleibi[1], Sadiq, Ahmed T[2], Alaluosi Wasan M[3]., **2016**. Feature Extraction of Human Facial Expressions Using Haar Wavelet and Neural Network, *Iraqi Journal of Science*, **57**(2C):1558-1565, ISSN: 0067-2904.
2. Najmeh Samadiani [1], Guangyan Huang [1], Borui Cai [1], Wei Luo [1], Chi-Hung Chi [2], Yong Xiang [1]and Jing He 3, **2019**. A Review on Automatic Facial Expression Recognition Systems Assisted

by Multimodal Sensor Data, Sensors, 19, 1863; doi:10.3390/s19081863 www.mdpi. com/ journal/sensors

3. Giacomo Mancini[1*], Roberta Biolcati[1], Sergio Agnoli[2], Federica Andrei[3] and Elena Trombini[3], **2018**. Expressions Among Italian Pre-adolescents, and Their Affective Reactions, Front. Psychol., https://doi.org/10.3389/fpsyg.2018.01303.

4. Fraser.W. Smith1, Stephanie Rossit2, **2018**. Identifying and detecting facial expressions of emotion in peripheral vision, PLOS ONE, School of Psychology, University of East Anglia, Norwich, United Kingdom, https://doi.org/10.1371/journal.pone.0197160.

5. M. Arfan Jaffar and Eisa Al Eisa, **2014**. "Classification of Facial Expression Using Transformed Features", *International Journal of Information and Electronics Engineering*", **4**(4) July.

6. Veena Mayya, Radhika M. Pai, Manohara Pai M. M, ,**2016.** "Automatic Facial Expression Recognition Using DCNN", 6th International Conference On Advances In Computing & Communications, ICACC 6-8 September, Cochin, India, Published by Elsevier B.V. (http://creativecommons.org/licenses/by-nc-nd/4.0/).

7. Michael Revina and W.R. Sam Emmanuel, **2018**. "A Survey on Human Face Expression Recognition Techniques", *Journal of King Saud University –Computer and Information Sciences*. https://doi.org/10.1016/j.jksuci.2018.09.002.

8. D. Yangaو Abeer Alsadoona, P.W.C. Prasad, A. K. Singhb and A. Elchouemi, **2018**. " An Emotion Recognition Model Based on Facial Recognition in Virtual Learning Environment, Procedia Computer Science 125 ,2–10, Published by Elsevier B.V.

9. Ayad R. Abbas and Asaad R. Kareem Iraqi Journal of Science, **2018**," Intelligent Age Estimation from Facial Images Using Machine Learning Techniques". *Iraqi Journal of Science*, **59**(2A): 724-732 DOI:10.24996/ijs.2018.59.2A.10.

10. Jyoti Kumaria, R. Rajesha, KM. Poojaa **2015.**" Face Expression Recognition: A survey, Second International Symposium on Computer Vision and the Internet1877 -0509 © 2015 The Authors. Published by Elsevier B.V.

11. Pushpaja V. Saudagare, D.S. Chaudhari, **2012.**" Facial Expression Recognition using Neural Network –An Overview " *International Journal of Soft Computing and Engineering* (IJSCE),**2**(1), March.

12. Galateia Iatraki, **2011**. "Emotional Facial Expressions Recognition & Classification", Master's thesis of Media and Knowledge ,Engineering Man Machine Interaction Group, Delft University of Technology, Holland.

13. N. P. Galatsanos, C. A. Segall, and A. K. Katsaggelos, **2003**. "Digital image enhancement", In Encyclopedia of Optical Engineering, R. Drigger Ed. Taylor &Francis, pp. 388 - 402.

14. Ming Yu, Gang Yan, Qing-Wen, **2006**. "New face recognition method based on DWT/DCT combine feature selection". School of Computer Science, Hebei University of Technology, Tianjin 300130, China Proceedings of the Fifth International Conference on Machine Learning and Cybernetics, Dalian, 13-16 August

15. Muzhir Shaban Al-Ani & Wasan M. Al-Aloosi, **2013.**"Biometrics Fingerprint Recognition using Discrete Cosine Transform DCT", *International Journal of Computer Applications* (0975 – 8887), **69**(6), May.

16. Ongalo P. N. Fedha, Huang Dong Jun & Richard Rimiru, **2012.** "A Neural Network Based Classifier for a Segmented Facial Expression Recognition System Based on Haar Wavelet Transform". *Global Journal of Computer Science and Technology*, **12**(7) Version 1.0.

17. Anissa Bouzalmat[1], Naouar Belghini[2], Arsalane Zarghili[3], and Jamal Kharroubi[4], **2011.**" Face Detection and Recognition Using Back Propagation Neural Network and Fourier Gabor Filters", Sidi Mohamed Ben Abdellah University, Morocco, .(*Signal & Image Processing: An International Journal* (SIPIJ) **2**(3).

18. HMKCODE, Backpropagation Step by Step https/, 03 NOV **2019**, ://hmkcode.com /ai/ backpropagation-step-by-step.

19. Dinesh Chandra Jain, V. P. Pawarm, **2012**. " A Different Scenario of Neural Network Algorithms towards Face Expression in Human Computer Interaction" *International Journal of Advanced Research in Computer Science and Software Engineering*, **2**.