Facial Emotion Recognition Based on Empirical Mode Decomposition and Discrete Wavelet Transform Analysis

H. Ali[1], M. Hariharan[1], H. Mansor[1], S.N. Adenan[1], M. Elshaikh[2] and K. Wan[1]
[1]*School of Mechatronic Engineering, University Malaysia Perlis, Kampus Pauh Putra, 02600 Arau, Perlis, Malaysia.*
[2]*School of Computer and Communication Engineering, Universiti Malaysia Perlis, Kampus Pauh Putra, 02600 Arau, Perlis, Malaysia*
hasimahali@unimap.edu.my

***Abstract*—This paper presents a new framework of using empirical mode decomposition (EMD) and discrete wavelet transform (DWT) with an application for facial emotion recognition. EMD is a multi-resolution technique used to decompose any complicated signal into a small set of intrinsic mode functions (IMFs) based on sifting process. In this framework, the EMD was applied on facial images to extract the informative features by decomposing the image into a set of IMFs and residue. The selected IMFs was then subjected to DWT in which it decomposes the instantaneous frequency of the IMFs into four sub band. The approximate coefficients (cA1) at first level decomposition are extracted and used as significant features to recognize the facial emotion. Since there are a large number of coefficients, hence the principal component analysis (PCA) is applied to the extracted features. The k-nearest neighbor classifier is adopted as a classifier to classify seven facial emotions (anger, disgust, fear, happiness, neutral, sadness and surprise). To evaluate the effectiveness of the proposed method, the JAFFE database has been employed. Based on the results obtained, the proposed method demonstrates the recognition rate of 80.28%, thus it is converging.**

***Index Terms*—Discrete Wavelet Transform; Empirical Mode Decomposition; Facial Emotion Recognition; K-Nearest Neighbour; PCA.**

I. INTRODUCTION

Facial expression is a versatile modality and the most cogent in conveying human expressions. Facial expression or simply facial emotion could impart an individual's emotional state, his/ her intensions and finally elicit different responses. Over the past decades, facial emotion recognition (FER) has received sizable impact in different applications such as in psychology, computer technology, medicine and security. Moreover, FER becomes a key point in the wave of research within human-computer interaction (HCI) that aims to communicate naturally between man and machines. Although the FER has achieved a level of maturity, however, development of robust FER is still a challenging task, largely due to unpredictable facial variations, subtle changes of facial features that may affect the performance of FER.

In the literature, approaches to facial feature extraction and recognition span a wide a wide range of methods. Some of them have used facial components [1], facial points [2], facial landmarks [3], pixel intensities [4], shape and texture [5], Gabor wavelet [6], local binary pattern [7]. For example, Lyons et al. [8] utilized Gabor wavelet on the labeled elastic graph of the facial image. The face image was firstly convolved with Gabor wavelet at different spatial frequencies and orientations. The two different grids which are rectangular grid of 7 x 7 and fiducial grid were then placed on the convolved face image separately at the located landmark of the face. The output features which are the amplitude of Gabor filter were extracted and ensembled into a single vector called a labeled graph vector. Donato et al. [9] compared several techniques of including Gabor wavelet, optical flow, PCA, independent component analysis and local feature representation for recognizing six single upper face action units and six lower face action units. Their results show that Gabor wavelet and independent component analysis achieved the best performances. However, the computational time of Gabor wavelet is very expensive. Rose [10] further compared the performance of Gabor, log-Gabor filters and image pixel in classifying the facial expression recognition. The facial points were manually extracted and convolved with Gabor and Log-Gabor to form feature vectors. PCA has been used to these feature vectors and classify them using LDA. The results show that classification performances of Gabor and Log-Gabor are comparable with 85% accuracy while image pixel shows the lowest accuracy of 77%. Gu et al.[11] proposed radial encoding schemes on local-Gabor features in classifying the facial expression. They divided the non-overlapping regions of facial image into a certain local block *so-called* patch and each patch was subjected to Gabor filters at specified scales and orientations. The outputs of Gabor filters on each local patch were subjected to radial encoding scheme to form local features. These local features were concatenated to form feature vectors before they fed to the classifiers. Even though a Gabor filter offers useful properties in term of tolerance against illumination and invariance to limit spatial transformation, the outputs of this filter are highly correlated with the redundancy of the information at neighboring pixels [12]. Moreover, the selection of facial points and the downsampling factor also may affect the final recognition performance in FER [7]. It is known that Gabor filters still consuming time and memory intensive for convolving face images with a bank of Gabor filters to extract multi-scale-multi-orientations coefficients.

Empirical Mode Decomposition (EMD) is a self-adaptive algorithm and is suitable for analyzing the non- stationary and nonlinear phenomena, since it can adaptively decompose the signal into Intrinsic Mode Functions (IMFs) [13]. The basic idea is to decompose the signal into multiple IMFs and then select the appropriate IMF to construct the envelope spectrum

using the Hilbert transform. Meanwhile, the DWT having advantages of providing good qualities in spatial and frequency domains/ information of textures which are useful for classification and have been considered in face recognition and facial expression [14]. This multi-resolution technique which based on human visual perception, intend to transform into a representation in which both spatial and frequency information are presented. These multi-scale features try to characterize textures by filter responses directly. Due to advantages of both techniques, therefore this paper presents a new framework of using EMD and DWT algorithm to recognize the seven facial emotions. This paper is organized as follows: Section II explains the materials and methods used in FER such as EMD, DWT and PCA. Section III presents and discusses obtained using the proposed method. Finally, Section IV concludes the results of the proposed method.

II. The Proposed System Description

Figure 1 shows the block diagram of the proposed method. It comprises image pre-processing, feature extraction, principal component analysis (PCA) and facial classification. The description of the working process is elaborated as follows. The original facial image of size 256 by 256 pixels is firstly cropped into 128 by 96 pixels to extract the face region from the background. Then, histogram equalization was applied to the cropped image to eliminate the illumination effects.

A. Facial Expression Database

The available public database, Japanese Female Facial Expression (JAFFE) was employed in this work. It contains 213 images consisting of 10 subjects (female) which performed six basic facial expressions (anger, disgust, fear, happiness, sadness, and surprise) and neutral. The examples of the subjects are illustrated in Figure 2.

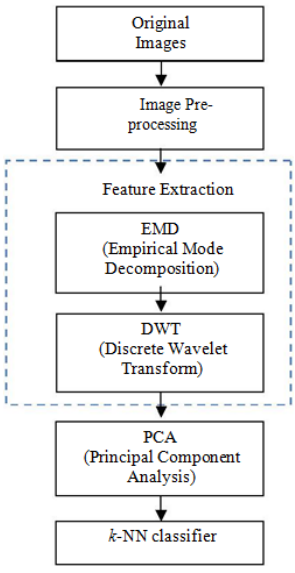

Figure 1: The block diagram of the proposed method

Figure 2: Example of JAFFE database

B. Empirical Mode Decomposition

Empirical mode decomposition (EMD) is a multi-resolution technique which used to adaptively decompose a non-stationary signal into a finite or a small set of functions [13]. The EMD is widely used in oceanography, radar satellite, signal processing and very recent in medical EEG signal, i.e. epilepsy classification. The EMD (one-dimensional) can be extended into a two-dimensional version for textural analysis[15]. It has the advantages of data-driven and no prior of basis function needed. In deriving a small set of functions of EMD so-called intrinsic mode functions there two conditions must be satisfied: (1) the number of extrema and number of zero crossing must equal or at most differ by one and (2) the mean value of the envelope must be zero or 'monotone'. The EMD based on Huang's algorithm can be summarized as follows:

Let the given signal, $x(t)$. The signal $x_{i,k}(t)$ defines a component of the sifting process, in which the first iteration is $x_{i,k}(t) = x(t)$:

1. Locate the local maxima and local minima of the $x_{i,k}$ (t)
2. Interpolate local maxima for upper envelope $e_u(t)$ and local minima for lower envelope $e_l(t)$
3. Calculate the mean of the envelope:

$$m_{i,k}(t) = \left(\frac{e_u(t) + e_l(t)}{2} \right) \quad (1)$$

4. Extract the details signal (original signal-mean signal) so that the next component sifting process is defined as, $x_{i,k\ 1}(t) = x_{i,k}(t) - m_{i,k}(t)$
5. Iterate step 1 to 4, with $k = k + 1$ if the criteria of IMF (the 2 conditions above) is not met.
6. The above steps are repeated until the resulting signal meets the IMF criteria and finally is IMF $c_i(t)$. To speed the process, the stopping criteria based on standard deviation SD is computed as:

$$SD = \sum_{t=0}^{T} \left[\frac{\left| x_{i,k-1}(t) - x_{i,k}(t) \right|^2}{s_{i,k-1}^2(t)} \right] < 0.3 \quad (2)$$

7. Repeat the next sifting process on resulting signal $r_i(t)$ of the extracted IMF $c_i(t)$ from the signal $x_{i,k}(t)$, $r_i(t)$ $x_{i,k}(t)$ $c_i(t)$,

$$r_i(t) = x_{i,k}(t) - c_i(t), \quad x_{i+1,k}(t) = r_i(t) \quad (3)$$

where: k = 1

Iterate the sifting process until all or the required numbers of IMFs are extracted from the signal. The original signal x(t) can be reconstructed by summing the extracted IMFs and the residual of the sifting process $r_N(t)$ by:

$$x(t) = \sum_{i=1}^{N} c_i(t) + r_N(t) \quad (4)$$

C. Discrete Wavelet Transform

The DWT analyzes the signal at different frequency bands with different resolutions by decomposing the signals into a coarse approximation and detail information. The DWT of 2D image $f(x, y)$ of size M x N can be defined as [16].

$$W_\varphi(j_0,m,n)=\frac{1}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y)\varphi_{j_0,m,n}(x,y)$$
$$W_\psi^i(j,m,n)=\frac{1}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y)\psi_{j,m,n}^i(x,y) \quad (5)$$
$$i=\{H, V, D\}$$

where: index I = Identifies the directional wavelets in horizontal (H), vertical (V) and diagonal (D)
j_0 = Arbitrary starting scale
$W_\varphi(j_0, m, n)$ = Coefficients define an approximation of $f(x, y)$ at scale j_0
$W^i_\psi(j_0, m, n)$ = Coefficients add horizontal, vertical and diagonal details for scale j j_0

The 2D-DWT can be implemented using a digital filter and downsamplers.

The single-scale filter bank can be iterated to produce a P scale transform in which scale j is equal to $J\text{-}1$, $J\text{-}2$, …, $J\text{-}P$. Image $f(x, y)$ is used as $W_\varphi(J, m, n)$ input[16]. Convolving its rows with $h_\varphi(-n)$ and $h_\psi(-n)$ and downsampling its column, results in two sub-images whose horizontal resolutions are reduced by a factor of 2. The high pass or detail component characterizes the image's high-frequency information with vertical orientation; the low pass, approximation component contains its low frequency, vertical information. Both sub-images are then filtered column wise and downsampled to yield four quarter-size output sub-images.

D. Principle Component Analysis

The goal of PCA[17] is to seek linear transformation from a higher dimensional space into a lower dimensional space. The PCA can be calculated by:

a. Extract the mean of the data
b. Find the covariance matrix, C
c. Find the eigenvalues λ_i and eigenvectors v_i
d. Sort the eigenvectors in descending order which corresponding to the largest eigenvalue. Choose the first k principal components [16].

III. Results and Discussion

In this section, we evaluated the proposed method using 213 images of JAFFE database. The pre-processed image is subjected to EMD framework. Then, the resultant of EMD features were further process using DWT technique. The informative features of low-frequency sub band were extracted and used as significant features. Hence, the coefficients of the respective sub- the band are high, thus PCA was used to reduce the data dimensionality. The reduced features were then classified using k-NN classifier.

A. Applied EMD on Facial Images

In extracting the features, the facial images are firstly decomposed using EMD technique to produce a small set of intrinsic mode functions (IMFs) which is IMF1, IMF2, IMF3 and residue. Figure 3 shows the EMD decomposition applied on facial images. As observed in Figure 3 that the IMFs exhibit the pattern structures from the finest to coarsest of the original image. Due to nature of EMD technique, as the order of the IMF increases, the relative mean of the data approaches to zero [18]. The first IMF (IMF1) effectively contains the largest magnitude extrema in the facial image which contribute to the highest local information that describes the characteristic of distinct facial emotions. Based on the above rationale, thus the first IMF is extracted and used as a feature for further analysis since the first IMF contains significant features and we discarded the second and third IMF.

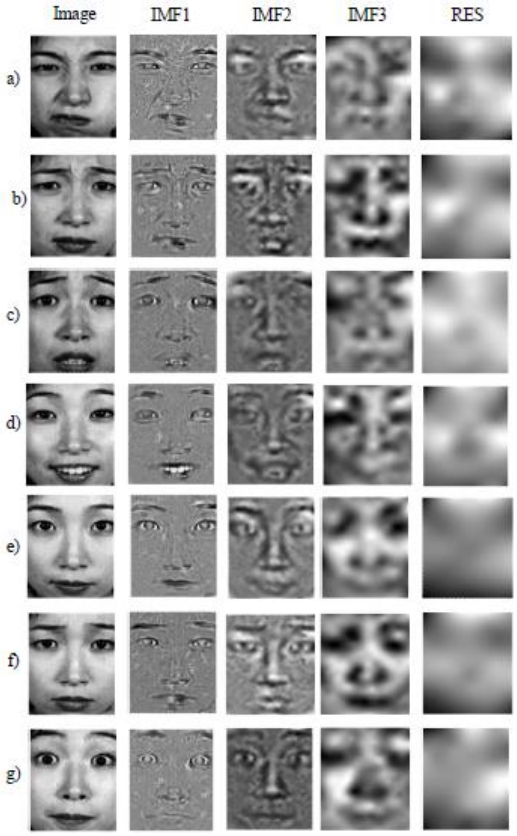

Figure 3: The original image and their corresponding IMFs (IMF1, IMF2 and IMF3) and residue using EMD technique for emotion: a) anger, b) disgust, c) fear, d) happiness, e) neutral, f) sadness and g) surprise

B. Applied DWT on the EMD Domain

In this framework, the first IMF (IMF1) is further analyzed using DWT technique. The main advantages of wavelets is that they have a varying window size, being wide for slow frequencies and narrow for the fast ones, thus leading to an optimal time-frequency resolution in all frequency ranges. Figure 4 illustrates the image (IMF1) is decomposed i.e. divided into four sub-bands or sub-sampled by applying DWT. These sub bands are approximate coefficients (cA) that represent the low-frequency level, horizontal coefficients (cH), vertical coefficients (cV) and diagonal coefficients (cD) correspond to the finest scale wavelet coefficients of detail images. The coefficients obtained from DWT of approximation images are basic features that are useful for classification.

Generally, low-frequency component represents the basic figure of an image which is less sensitive to varying image. These components are the most informative sub-images gearing with the highest discriminating power. In this study, sub-image of cA1 is extracted and used as features. The coefficients of this sub-images are transformed into 1-D

Feature vector form by concatenating each row of coefficients matrix. Then, the feature vectors are subjected to dimensionality reduction using principal component analysis (PCA). In this work, the first 200 principal components are used as features which correspond to 99% variability of eigenvectors.

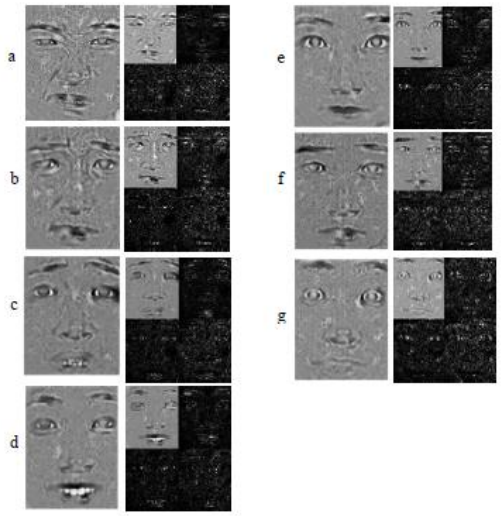

Figure 4: First level DWT decomposition on IMF1; a)anger, b) disgust, c) fear, d) happiness, e) neutral, f) sadness and surprise

C. *The Experimental Results on Extracted EMD and DWT features using k-NN*

To evaluate the proposed method, 10-fold cross validation strategy is used. In this procedure, the whole data (213) is randomly split into ten folds having equal size. At each process, nine subsets (folds) are used as a training and the remaining ones used for testing. The process is repeated for the ten folds, and finally, the average is computed.

Table 1 shows the confusion matrix of seven facial expression using reduced EMD and DWT features based on k-nn classifier. Based on Table 1 that the average recognition rate has achieved 80.28%. Emotion *fear* contributes the lowest recognition rate which is 69.7% or 10 out of 33 are misclassified. This emotion mostly confused with emotion *sad*, *surprise* and *neutral*. This may due to the energy of IMF1 of emotion *fear*, *sad* and *surprise* at wavelet sub-band resemble each other that observe high redundancies occur at feature distributions within the class scatter. Besides, emotion *disgust* shows the second lowest recognition rate which is 72.41%. Seven (7) out of 29 are wrongly classified with emotion *anger*. This can be interpreted that both *disgust* and *fear* may have similar trends of IMF1 that best describe the emotion characteristic.

On the other hand, emotion *surprise* achieves the highest recognition rate which is 93.33%. This implies that the IMF1 distribution which has the largest magnitude extrema able to characterize the behavior of the emotion. Furthermore, high energy consumption during eliciting this expression contributes to stable emotion features which lead better performance accuracy.

Table 1
The Confusion Matrix Based on Reduced EMD and DWT Features Using k-NN Classifier. Note: An (Anger), Dis (Disgust), Fe (Fear), Ha (Happiness), Ne (Neutral), Sad (Sadness) and Sur (Surprise)

	An	Dis	Fe	Ha	Ne	Sad	Sur	*Total*
An	**26**	1	3	0	0	0	0	30
Dis	7	**21**	0	0	0	1	0	29
Fe	1	0	**23**	1	2	3	3	33
Ha	0	0	0	**24**	0	2	4	30
Ne	0	0	1	1	**26**	2	0	30
Sad	1	1	5	1	0	**23**	0	31
Sur	0	1	0	1	0	0	**28**	30
Average							**80.28%**	**213**

IV. CONCLUSION

This paper has presented a new pattern recognition framework for recognizing facial emotion recognition using EMD and DWT as feature extraction technique. Based on the results obtained, the EMD decomposed the facial image into IMF1, IMF2, IMF3, and residue. It observed that the IMFs exhibit different pattern resolution is varying from coarsest to finer texture for each emotion. Due to having the largest magnitude of extreme in IMF1, IMF1 is extracted for further analysis. In DWT, the IMF1 space is decomposed into four sub band at first-level. The low frequency of cA1 is extracted since it contains most informative sub-images as well as with the highest discriminating power. The recognition rate of using reduced EMD and DWT has achieved 80.28% using k-NN classifier considerably a feasible way in extracting the facial features. However, further study should be conducted on raw EMD and DWT features for improving and enhancing the system performance as overall.

ACKNOWLEDGMENT

This research work is supported by Short Term Grant Scheme (STG: 9001-00560).

REFERENCES

[1] A M. Ilbeygi and H. Shah-Hosseini, "A novel fuzzy facial expression recognition system based on facial feature extraction from color face images," Engineering Applications of Artificial Intelligence, Vol. 25 (1), (2012), p. 130–146.

[2] M. Pantic, and L. J. M. Rothkrantz: Expert system for automatic analysis of facial expressions. Image and Vision Computing, 18 (2000), 881– 905.

[3] N. Aifanti, and A. Delopoulos, "Linear subspaces for facial expression recognition. Signal Processing: Image Communication, 29(1), (2014), p. 177–188.

[4] T. Danisman, I. M. Bilasco, J. Martinet, and C. Djeraba, "Intelligent pixels of interest selection with application to facial expression recognition using multilayer perceptron. Signal Processing, 93(6), (2013), p. 1547–1556.

[5] Kotsia, S. Zafeiriou, and I. Pitas, "Texture and shape information fusion for facial expression and facial action unit recognition. Pattern Recognition, 41(3), (2008), p. 833–851.

[6] Z. Zhang, M. Lyons, M. Schuster, and S.Akamatsu: Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron. In IEEE International Conference on Automatic Face and Gesture Recognition, 1998. (pp. 454–459.

[7] X. Feng, M. Pietikäinen, and A. Hadid, "Facial Expression Recognition with Local Binary Patterns and Linear Programming. Pattern Recognition and Image Analysis, 15(2), (2005), p. 546–548.

[8] M. Lyons, S. Akamatsu, M. Kamachi, & J. Gyoba, (1998). Coding facial expressions with Gabor wavelets. In *Proceedings - 3rd IEEE International Conference on Automatic Face and Gesture Recognition, FG 1998* (pp. 200–205).

[9] Donato, G., Bartlett, M. S., Hager, J. C., Ekman, P., & Sejnowski, T. J. (1999). Classifying facial actions. *IEEE Transactions on Pattern*

Analysis and Machine Intelligence, *21*.

[10] Rose, N. (2006). Facial Expression Classification using Gabor and Log-Gabor Filters. 7th International Conference on Automatic Face and Gesture Recognition (FGR06), pp 346-350.

[11] Gu, W., Xiang, C., Venkatesh, Y. V., Huang, D., & Lin, H. (2012). Facial expression recognition using radial encoding of local Gabor features and classifier synthesis. *Pattern Recognition*, *45*(1), 80–91.

[12] Owusu, E., Zhan, Y., & Mao, Q. R. (2014). A neural-AdaBoost based facial expression recognition system. *Expert Systems with Applications*, *41*(7), 3383–3390.

[13] N. Huang, Z. Shen, S. Long, M. Wu, H. Shih, Q. Zheng, H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *454*(1971), 1998.

[14] E. Magli, G. Olma, and L. Lo Presti, "Pattern recognition by means of the Radon transform and the continuous wavelet transform, Signal Processing, Vol. 73, (1999), p. 277-289.

[15] Nunes, J. ., Bouaoune, Y., Delechelle, E., Niang, O., & Bunel, P. (2003). Image analysis by bidimensional empirical mode decomposition. *Image and Vision Computing*, *21*(12), 1019–1026.

[16] R. C. Gonzalez, and R. E. Woods, "Digital Image Processing (3rd Edition). 3nd edition. Prentice Hall, New York, (2007).

[17] P. N. Belhumeur, J. P. Hespanha, & D. J.Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *19*(7), 1997, pp. 711–720.

[18] G. Rilling, P. Flandrin, & P.Goncalves, "On empirical mode decomposition and its algorithms," Proc. of the IEEE-Eurasip Workshop on Nonlinear Signaland Image Processing, 3, 2003, pp.8-11.