Illumination Effects on Facial Expression Recognition using Empirical Mode Decomposition

Hasimah Ali, Wan Khairunizam Wan Ahmad
*Centre of Excellence Unmanned Aerial Systems (*COE-UAS*), Faculty of Electrical Engineering & Technology, Universiti Malaysia Perlis, Arau, 02010, Malaysia*

Hariharan Muthusamy
Department of Electronics Engineering, National Institute of Technology Uttarakhand, India

Mohamed Elshaikh
*Department of Electronics Engineering & Technology (FTKEN), Universiti Malaysia Perlis, Arau, 02010, Malaysi*a
E-mail: hasimahali@unimap.edu.my, hariharan@nituk.ac.id, khairunizam@unimap.edu.my, elshaikh@unimap.edu.my

Abstract

Facial expression recognition (FER) has been acknowledged as a significant modality that could bring facial expression into human-machine interaction and make the interaction more efficient. However, the ability of FER to operate in a fully automated and robust manner is still challenging. Illumination effects, for example, make the facial expression images always contaminated with different levels of ambient noise (such as brightness variation) in a cluttered background. Thus, this paper aims to investigate the illumination effects (brightness variations) on facial expression recognition using empirical mode decomposition reconstruction techniques. In this framework, firstly, the noisy facial expression images were simulated with the illumination effects using different brightness levels of 30%, 40%, 50%, 60%, and 70%. Then, the EMD will decompose the noisy facial expression images into a small set of intrinsic mode functions (IMF), namely IMF1, IMF2, IMF3, and residue. Based on property held by EMD, the signals are decomposed into several IMF components, each with a different time scale. Because the last several IMFs represent the majority of illumination effects, various reconstruction techniques for IMFs have been investigated at various brightness levels. Feature reduction techniques Principal component analysis (PCA) and linear discriminant analysis (LDA) have been employed to reduce the high-dimensional space of IMF features into low-dimensional IMF features. The reduced IMF reconstructions were then used as input to the k-nearest neighbour classifier to recognise the seven facial expressions. A series of experiments have been conducted on the JAFEE database using various reconstruction IMFs together with PCA plus LDA. Based on the results obtained, the reconstruction of IMF1 + IMF2 + IMF3 shows the highest accuracy in high illumination conditions, which is 99.06%.

Keywords: Illumination, Facial Expression Recognition, Empirical Mode Decomposition, PCA plus LDA.

1 Introduction

Facial expression is the most natural and important tool people have to show their emotional-state or deliver information. For decades, facial expression has been regarded as an important modality for conveying feelings and attitudes, and it has had an impact on physical communication. Specifically, facial expression is formed by contracting facial muscles [1]. Facial expression recognition (FER) is a function that humans and computers can perform by locating the face from its sources. Then, facial features will be analysed to classify the emotion [2]. A lot of information about human behaviour can be effectively extracted from their facial expression. Facial expression recognition (FER) has been acknowledge as a significant modality that could bring facial expression into human machine interaction and makes the interaction more efficient. However, the ability of FER to operate in fully automated with robust is still challenging. Illumination effects for example, make the facial expression images always contaminated with different level of ambient noise (such as brightness variation) in clutter background. In previous work, limited study was conducted on illuminations effects for facial expression recognition. For example, [3] have combined the logarithm transform, discrete cosine transform, and illumination compensation as a normalised DCT to eliminate illumination variation contained in the facial images of the JAFFE database. To recognise the facial expression, the extracted features from the pre-processed image were combined with a neural network classifier. However, analysis of the effects of illumination was not discussed. In [4], the authors utilized the independence component analysis (ICA) and locality-preserving projection to compare the effectiveness of those methods towards the illumination variations on the face recognition. They found ICA outperformed the results on YALEB database with 64 illuminations types. In [5], the authors utilized deep learning approach for emotion recognition under different pose and illumination variations. The Convolution Neural Network were constructed using 15 layers (three layers per emotion) that learn from facial expression database (CMU Multi-PIE) using input size of 32 x 32 in recognizing the five basic emotions. Although they achieved 96.55%, selecting optimum the input size considered to be computational cost. In [6] the authors utilized a deep stacked convolutional autoencoder in attempting reconstruct new input images with better illumination for facial expression recognition. The model used as pre-training in greedy layer-wise by learning the image representation and then map it to approximately reconstruction of input image. The network model learns to encode the input images and create a feature vector form relatively similar illumination, ignoring the level of luminance of the facial images. Even though the used of deep learning is emerging, however handcraft features of the image provides significant impact in understanding the behaviour of spatial details of the images. Therefore, in this study, we propose to investigate the illumination effect on facial expression recognition (FER) by using empirical mode decomposition (EMD) approach. In this framework, the original facial expression database will go through face detection in a manipulated illumination condition, and then various illumination coefficients will be analysed. The EMD will be used to decomposed the illuminated images (in our case brightness variations) into a set of intrinsic mode functions (IMFs). These IMFs will be used to further analyse the illumination effects containing in facial expression images. The significant IMFs will be reconstructed by removing the one contains illuminations. In order to reduce dimensionality of reconstructed IMF, PCA plus LDA will be adopted. Finally, the reduced features of reconstructed IMFs will be classified using k-NN classifier to recognize the seven facial expression under the illumination effects.

2 Materials and Method

Fig. 1 shows the framework of the proposed system. It consists of five phases: facial expression image database, pre-processing, EMD-based feature extraction, IMF reconstruction, and classification. In this framework, firstly, the facial expression image was pre-processed to detect the local region of the face. Then, brightness variations were simulated on the detected face. By decomposing an image into a set of intrinsic mode functions via a sifting process on the simulated brightness of a facial expression database, the nonlinear technique, namely EMD, has been used as a feature extraction-method. To assess performance, an IMF reconstruction was performed on various IMFs. Later, feature reduction was applied to the reconstructed image to reduce its dimensionality before being subjected to the classifier. The detailed description of the process is highlighted as follows:

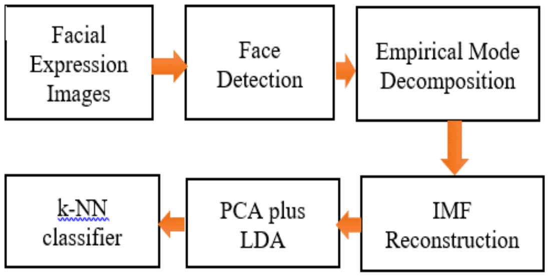

Fig. 1. Framework of the proposed method

2.1 *Facial Expression Database*

In this study, the Japanese Female Facial Expression (JAFFE) database has been used. It contains 213 grayscale facial expression images of ten subjects, all Japanese women, with seven facial expressions: neutral (30 images), angry (30 images), sad (30 images), happy (31 images), disgust (29 images), surprise (31 images), and fear (32 images).

2.2 Image Pre-Processing: Face Detection

Face detection is a technique used to detect and localize the face region from the unwanted background. The original size of JAFFE image is 256 x 256 pixel was cropped into 128 x 96 pixel that contains the face region. Fig. 2(a) and Fig. 2(b) show the example of the original and cropped image of seven facial expressions of a JAFFE subject, respectively.

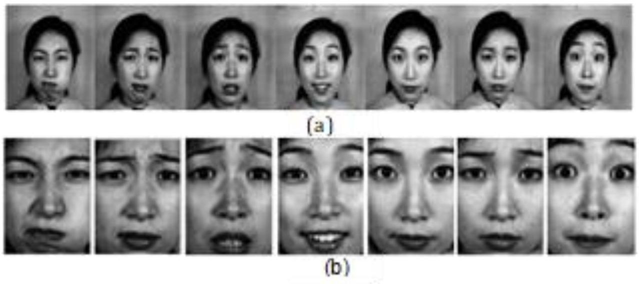

Fig. 2. Seven facial expression of JAFFE's subject: (a) original and (b) cropped image

2.3 Brightness Variation

In this study, illumination variation with different levels of brightness was used to simulate illumination effects that mimic real-world applications. The illumination variations were simulated by varying the brightness effect by controlling the intensity of each individual pixel inside the image. The percentage illumination level used in this project is 30%, 40%, 50%, 60%, and 70%. A percentage greater or less than those specified will result in either increased brightness or decreased darkness.

2.4 Feature Extraction: Empirical Mode Decomposition

Empirical mode decomposition (EMD) is a method of decomposing a natural signal without leaving the time domain. EMD decomposes any nonlinear signal into a small set of finite intrinsic mode functions (IMFs). Although EMD is focused on analysing one-dimensional signals, it can also be extended to bi-dimensional empirical mode decomposition (BEMD) to analyse two-dimensional signals while still having similar concepts and procedures [7], [8]. The main advantage of EMD is that it can extract data about local trends in the input signal by calculating and quantizing oscillations. Such oscillations can be quantized by local details and matching a local trend. The fundamental ideas of EMD are the implementation of the incorporation method such as cubic splines, sifting phase to extract the intrinsic mode function, and numerical convergence criteria to stop the sifting phase.

In the EMD technique, there are two steps that must be met to extract the intrinsic mode functions (IMFs). To begin, the number of extremes and zero crossings must be equal, or at most one. Second, the mean of its upper and lower envelopes must not be more or less than zero. Those two steps satisfy the physically compulsory process for defining a significant instantaneous frequency. The EMD technique's entire procedure is as follows: for input signal s(t):

1. Determine the local maxima and local minima of s(t). Let $d_0(t) = s(t)$.
2. Incorporate the local maxima and the local minima to develop the upper envelope eu(t) and lower envelope el(t), correspondingly, using cubic B splines.
3. Calculate the mean of envelope m(t): $m(t) = (e_u(t) + e_l(t))/2$.
4. Determine the details using $d_1(t)=d_0(t)-m(t)$ (sifting phase).
5. Repeat steps 1-4 on the residual signal until the details signal $d_k(t)$ is included in the IMF:$c_1(t)=d_k(t)$
6. Procedures 1–5 should be repeated on the residual $rn(t)=s(t).C_n(t)$ to obtain all of the signal's IMFs ($c_1(t)$, $c_2(t)$,..., $c_n(t)$).

When the residual signal is either unaltered, monotone, or a function with just one extrema, the procedure is terminated. The advantage of the EMD approach is that once these IMFs are acquired, they may easily switch back and forth from them to the original signal. Similarly, the original signal may be reconstructed by combining all of the IMFs together, which can be adjusted to account for slight variations owing to the implant discovered in the method. EMD can be expressed in mathematically as:

$$X(n) = \sum_{n=1}^{N} C_n(t) + R_N(t) \quad (1)$$

According to the technique, lower-order IMFs record rapid oscillation modes, whereas higher order IMFs record low oscillation modes. In this research, two-dimensional empirical mode decomposition (BEMD) is used to reduce illuminated face pictures into a limited collection of intrinsic mode functions (BIMFs). The BIMFs are features recovered at various scales or spatial frequencies using phase sifting.

2.5 Dimensionality reduction: PCA plus LDA

In this work, the PCA plus LDA has been employed as dimensionality reduction on the extracted IMFs features

2.5.1 PCA

Although the intrinsic mode functions effectively represent the different frequencies of oscillation for different expressions (local trends) in the image, the mode function responses can be large. A 128 x 96 image, for example, decomposes with the EMD resulting mode function into a 128 x 96 = 12,288-dimensional vector. To reduce the dimensions of the feature vector, we applied PCA to our mode function. Let consider, if $A=\{a_1, a_2, a_3 ... a_N\}$ is a vector of N images.

i. First, calculate the mean (μ) of the matrix, and covariance matrix (S) as:

$$S=\frac{1}{n}\sum_{i=1}^{n}(a_i-\mu)(a_i-\mu)^T \tag{2}$$

ii. Calculate both eigenvalues (λ_i) and eigenvectors (v_i) of S as:

$$Sv_i=\lambda_i v_i, \tag{3}$$

where i=1, 2, 3…n.

iii. Lastly, sort the eigenvectors in descending order that associated to the largest eigenvalues. The new projected space is given b:

$$y=W^T a, \tag{4}$$

which W is represent the transformation matrix [9].

2.5.2 LDA

One of the main problem in pattern recognition is the curse of dimensionality. LDA is one of the method used to solve the issue of dimensionality by reducing the features from higher-dimensional space to lower-dimensional space. The goal of LDA is to shape the class scatter by maximizing the between-class scatter and minimising the within-class scatter [9]. In this framework, firstly we need to compute the between-class variance S_B as:

$$S_B=\sum_{i=1}^{n}N_i(\mu_i-\mu)(\mu_i-\mu)^T \tag{5}$$

Second, within-class scatter, S_w need to be computed by $S_W=\sum_{i=1}^{c}\sum_{a_k\in A_i}(a_k-\mu_i)(a_k-\mu_i)^T$ .Finally, the lower-dimensional feature projection, W_{lda} can be determined via:

$$W_{lda}=\arg\max\left|\frac{W^T S_B W}{W^T S_W W}\right|=[w_1, w_2, w_3 ... w_m] \tag{6}$$

in which w_i (i = $1, 2,…m$) is the set of S_B and S_w generalized eigenvectors based on the largest generalized eigenvalues λ_i of $S_B w_i=\lambda_i S_W w_i$.

2.5.3 PCA plus LDA

To avoid singularity of the S_w, [10] proposed PCA plus LDA. This framework involves two stages. First, original feature vector having f-dimensional feature space is mapped onto t-dimensional intermediate feature space in which ($t < f$) via PCA. Then, project the intermediate (t-dimensional) to k-dimensional feature space via LDA resulting new projected space as:

$$z_i=W_{pca}^T W_{lda}^T x_i \tag{7}$$

where i = {1, 2, 3..N} is the number of input images.

3 Results and Discussion

In order to evaluate the effectiveness of the proposed system, the JAFFE database, which is publicly available, has been used in this experiment. The facial images obtained from successful face detection and applied with different levels of illumination will undergo feature extraction using the EMD technique in order to extract the features called intrinsic mode functions (IMFs) through sifting procedures. IMFs are extracted from high-dimensional data. As a result, feature reduction using PCA and LDA was used before being fed as input to the k-NN classifier to classify all seven emotions: anger, disgust, fear, neutral, happiness, sadness, and surprise.

3.1 Brightness Variation on Facial Expression Images

The facial images from JAFFE database are simulated with different level of illumination which is 30%, 40%, 50%, 60% and 70%. This is done by adding and subtracting the intensity of each pixel inside of the facial images. Fig. 3 shows the simulated illumination of facial expression images.

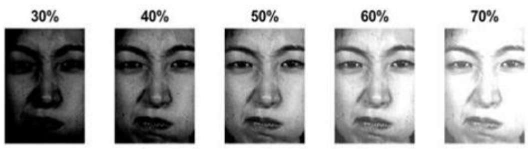

Fig. 3. Simulated illumination effects of facial expression images at different brightness level

3.2 Performance of applied EMD on expression images

The two-dimensional breakdown of a picture using the sifting process results in an interpretable display. Each IMF carries information on a certain scale that is efficiently segregated. Fig. 4 depicts the EMD decomposition of facial expression image into IMF1, IMF2, IMF3, and residual. Note that, the IMF1 contains local information about the smallest scales in the image, whereas the residue contains information about greater scales [7]. Specifically, IMF1 corresponds to the lowest time scale associated with the information's quickest time fluctuation. As the decomposition process progresses, the time scale increases, and the mean frequency of the mode decreases. As observed in Fig. 4, the set of IMF reveals the influent structure from the smallest to the largest relative to the input images. The IMF1 shows the most distinct facial features, including the boundary lines around the face region such as the mouth, nose, and eyes, which are crucial for expression recognition.

Fig. 4. EMD decomposition that associated with 30% brightness level

3.3 Performance Analysis on Facial Expression Recognition using Reconstructed Intrinsic Mode Function

According to [11], the majority of the illumination effects is represented by the last several IMFs. To evaluate how EMD can be employed to reduce the illumination effects in the proposed system. Fig. 5 shows the different combinations of the reconstructed mode functions. It should be noted that Figure 5(h) is a reconstruction of all IMF and a residue, and thus serves as a reference image. As can be seen in Fig. 5, the illumination effect was highly dominant in (b), (c), and (d). Whereas, (e), (f), and (g), the illumination effect appeared at a low degree. Meanwhile, in (a), which is IMF1+IMF2, the effects were almost gone. This observation agrees with [11] in which the several last mode functions are affected by illumination.

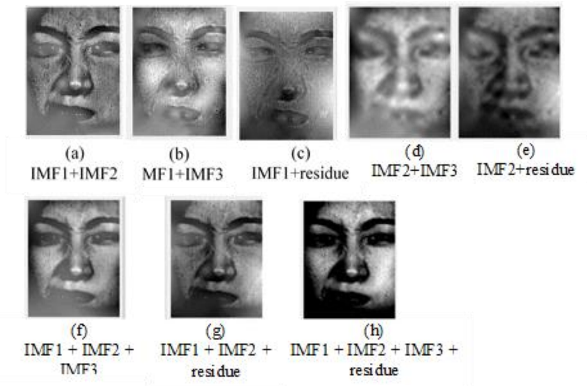

Fig. 5. Reconstruction of different mode functions (30% brightness level): (a) IMF1+IMF2, (b) IMF1+IMF3 and (c) IMF1+residue, (d) IMF2+IMF3, (e) IMF2+residue, (f) IMF1 + IMF2 + IMF3, (g) IMF1 + IMF2 + residue and (h) IMF1 + IMF2 + IMF3+residue (reference)

3.4 Performance Analysis of Applied PCA plus LDA on reconstructed IMFs

In this work, three reconstruction modes of IMFs have been investigated: IMF1+IMF2, IMF1+IMF2+IMF3, and IMF1+IMF2+IMF3+residue. Due to the higher dimensionality of feature space, the frameworks of PCA plus LDA were applied. The rationale for using PCA plus LDA is that, the feature vector space was reduced from a high-dimensional to a low-dimensional space by utilising the PCA plus LDA framework as mentioned in Section 2.5. Fig. 6 shows the distribution of reduced features due to IMF1+IMF2+IMF3+residue, whereas Fig. 7 shows the distribution of reduced features due to IMF1+IMF2+IMF3 at a 30% brightness level. Note that classes 1, 2, 3, 4, 5, 6, and 7 are denoted as *angry* (red), *disgust* (orange), *fear* (green), *happy* (cyan), *neutral* (blue), *sad* (purple), and *surprised* (pink), respectively. It can be seen in Fig. 6 that there is a low degree of overlap between the emotions of fear, anger, disgust, and neutral. Fig. 7 shows that the class emotions are well differentiated from one another.

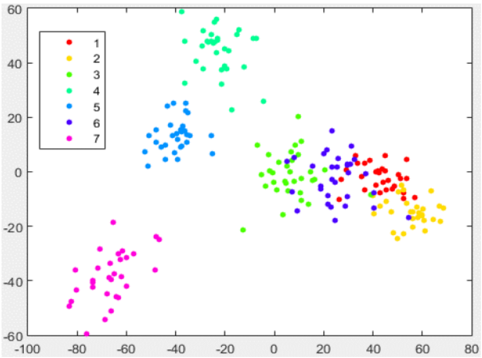

Fig. 6. Distribution of reduced features of **IMF1+IMF2+IMF3+residue** for seven facial expressions at 30% brightness level.

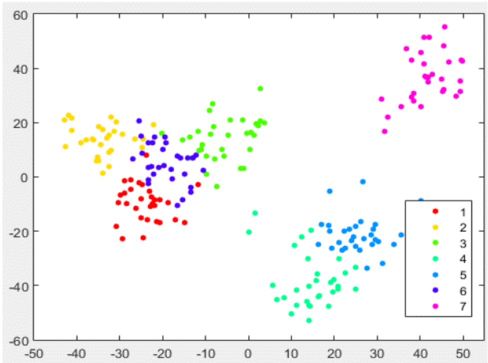

Fig. 7. Distribution of reduced features of **IMF1+IMF2+IMF3** for seven facial expressions at 30% brightness level.

3.5 *Experimental Results*

In this section, the reduced features using PCA plus LDA were applied to the reconstruction IMFs to recognize the seven facial expressions. The objective is to investigate the effectiveness of reconstruction IMFs in dealing with variations in illumination. Three different modes of IMF reconstruction were evaluated, namely IMF1+IMF2+IMF3+residue (used as reference), IMF1+IMF2+IMF3, and IMF1+IMF2 under different levels of illumination, which are 30%, 40%, 50%, 60%, and 70%. In this work, k-nearest neighbour was used as a classifier using a 10-fold cross validation strategy.

Fig. 8, Fig. 9, Fig. 10, Fig. 11, and Fig. 12 illustrate the recognition rate of IMF reconstruction of facial expression recognition at brightness levels of 30%, 40%, 50%, 60%, and 70%, respectively. As observed in Fig. 8, the reconstruction of IMF1+IMF2 shows the highest recognition rate, which is 99.06% at a 30% brightness level. It seems that the combined IMF1+IMF2 is enough and would achieve similar recognition as in the reference. We agree that, the illumination effect may lie inside the IMF3 and residue. When the brightness level increased to 40% (Fig. 9), the combined IMF1+IMF2 again showed the highest recognition rate of 99.53%, which is slightly higher by 0.53% than the reference.

In Fig. 10, at a brightness level of 50%, the combined IMF1+IMF2+IMF3 shows the highest recognition of 99.53%, which is the same as the reference. Similar trends appear for brightness levels of 60% (Fig. 11) and 70% (Fig.12), where the combined IMF1+IMF2+IMF3 outperformed the results, which are 99.53% and 98.59%, respectively. This can be inferred from the fact that as the brightness level increases, the combined IMF without residue shows significant results. Thus, we agree that the majority of illumination effects may be represented by the last IMFs and residue.

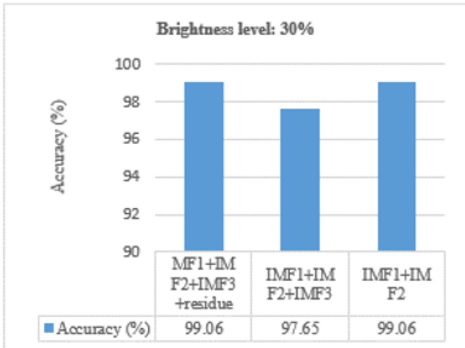

Fig. 8. Accuracy of FER based on reconstructed IMF at 30% brightness level.

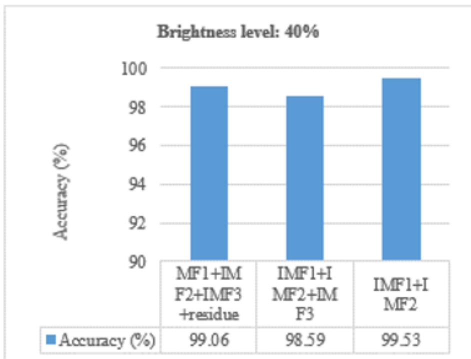

Fig. 9. Accuracy of FER based on reconstructed IMF at 40% brightness level.

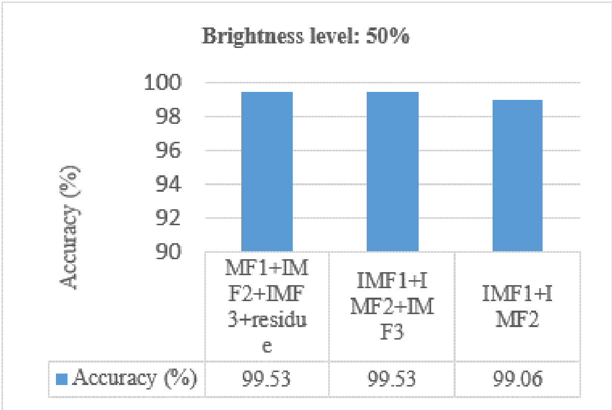

Fig. 10. Accuracy of FER based on reconstructed IMF at 50% brightness level.

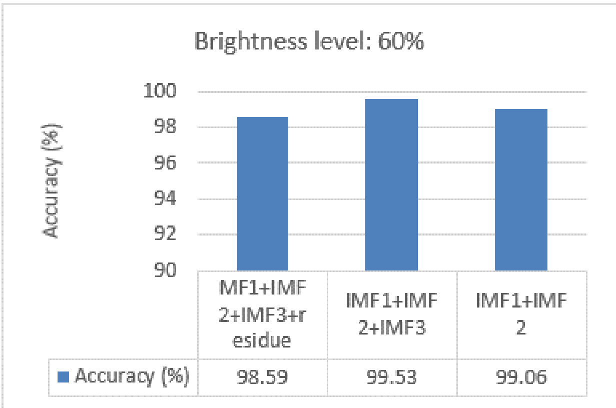

Fig. 11. Accuracy of FER based on reconstructed IMF at 60% brightness level.

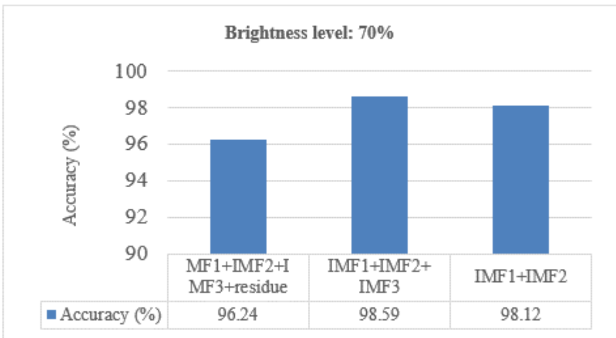

Fig. 12. Accuracy of FER based on reconstructed IMF at 70% brightness level.

4 Conclusion

This paper used an empirical mode decomposition approach to present the effects of illumination on facial expression recognition. Firstly, pre-processing of the facial image has been done via face detection to extract the local region of the face. Then, the face image was simulated to mimic the illumination effects by varying the brightness level between 30%, 40%, 50%, and 70%. Then, the simulated facial images were decomposed using empirical mode decomposition into a finite set of intrinsic mode functions. This work (IMF1, IMF2, IMF3, and residue) has been extracted via a sifting process. According to the findings, illumination effects may have been observed in the last few IMFs. Three reconstruction modes were considered for illumination analysis, which are the combined IMF1+IMF2, IMF1+IMF2+IMF3, and IMF1+IMF2+IMF3+residue (as reference). Based on the results obtained, at low brightness levels (30% and 40%), the combined IMF1+IMF2 gives the highest recognition rate of FER. However, as the brightness level was further increased to 50%, 60%, and 70%, the combined IMF1+IMF2+IMF3 gave the highest recognition rate. Therefore, we can conclude that as the brightness level increases, the combined IMF without residue shows significant results, and we agree that the majority of illumination effects may be represented by the last IMF and residue. However, more research on a larger dataset using different machine learning techniques is required to improve the robustness of FER to illumination effects.

Acknowledgment

The research is backed by the grant "IIRG003(b)-19IISS" from the University of Malaya, Malaysia.

References

1. Ali H, Hariharan M, Yaacob S, and Adom A H 2015 Facial emotion recognition using empirical mode decomposition *Expert Syst. Appl.* **42** p 1261.
2. Pantic M 2009 Facial Expression Recognition *Encycl. Biometrics* p 400.
3. Zankhana H S and Kaushik V 2015 Performance analysis of canny edge detection for illumination invariant Facial Expression recognition International Conference on Industrial Instrumentation and Control (ICIC).
4. Fernandes S L and Bala G J 2013 Comparative study on ICA and LPP based face recognition under varying illuminations and facial expressions International

Conference on Signal Processing , Image Processing & Pattern Recognition

5. Palaniswamy S and Suchitra 2019 A robust pose & illumination invariant emotion recognition from facial images using deep learning for human-machine interface 2019 4th International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS) p 1.
6. Ruiz-Garcia A, Palade V, Elshaw M and Almakky I 2018 Deep Learning for Illumination Invariant Facial Expression Recognition International Joint Conference on Neural Networks (IJCNN), p 1.
7. Nunes J C, Bouaoune Y, Delechelle E, Niang O, and Bunel P 2003 Image analysis by bidimensional empirical mode decomposition *Image Vis. Comput.* **21** p 1019
8. Bhuiyan S M A, Khan J F, and Adhami R R 2010 A bidimensional empirical mode decomposition method for color image processing Electrical Engineering Tuskegee *IEEE SiPS* p 272
9. Belhumeur P N, Hespanha J P and Kriegman D J 1997 Eigenfaces vs. Fisherfaces: Recognition using class special linear projection. *IEEE Trans. PAMI* **19**, p 711.
10. Pumlumchiak T. and Vittayakorn S. 2017 Facial expression recognition using local Gabor filters and PCA plus LDA 2017 9th International Conference on Information Technology and Electrical Engineering (ICITEE), pp. 1-6.
11. Qing, Jiang J, & Yang Z 2010 Empirical mode decomposition-based facial pose estimation inside video sequences *Optical Engineering*, *49*(3).

Authors Introduction

Ir. Dr. Hasimah Ali

She is senior lecturer at Faculty of Electrical Engineering Technology, UniMAP. She received her PhD in Mechatronic Engineering from UniMAP (2015), MSc in Mechatronic Engineering from IIUM (2008) and BEng in Mechatronic Engineering from IIUM (2004). Her research interests are signal and image processing, facial expression, ground penetrating radar, robotics gripper and IoT.

Prof Ir Dr. Wan Khairunizam Wan Ahmad

He received his B. Eng. degree in Electrical & Electronic Eng. from Yamaguchi University and Ph.D. in Mechatronic Eng. from Kagawa University, in 1999 and 2009 respectively. He is currently is a Prof. at Faculty of Electrical Engineering & Technology, University Malaysia Perlis and also serve as Director of Centre for Sustainable Academic Leadership Development (LEAD), UniMAP. He is member of Board of Engineer and Institute of Engineer, Malaysia. His research interest is in Human-Computer Interaction (HCI), Intelligent Transportation System, Artificial Intelligence and Robotics.

Assoc. Prof. Dr. Hariharan M.

He is currently working as an Associate Professor in the Department of Electronics Engineering, National Institute of Technology, Uttarakhand, India. He is a Senior Member of IEEE, USA, Member of IE, India and Member of IET, UK. He received Chartered Engineer status from Engineering Council, UK on February, 2016. His major research interests include Biomedical Signal and Image Processing, Speech Signal Processing, Signal Processing for Communications, AI/DL/ML Algorithms in VLSI:

Dr. Mohamed Elshaikh

He is a Senior Lecturer under Faculty of Electronic Engineering Technology (FTKEN), University Malaysia Perlis, Malaysia. He received his Ph.D. in Computer Engineering from University Malaysia Perlis, Malaysia, M.Sc Electrical and Electronics Engineering from University Technology Petronas (UTP), Malaysia and B.Sc Engineering Technology (Computer Engineering), University of Gezira, Sudan. His research mainly is on computer networking related.