2020 16th IEEE International Colloquium on Signal Processing & its Applications (CSPA 2020), 28-29 Feb. 2020, Langkawi, Malaysia

Virtual Markers based Facial Emotion Recognition using ELM and PNN Classifiers

M Murugappan*[1] Vasanthan Maruthapillai[2] Wan Khariunizam[3] A M Mutawa[4] Sai Sruthi[1] Chong Wen Yean[3]

[1]Department of Electronics and Communication Engineering, Intelligent Signal Processing Research Lab,
Kuwait College of Science and Technology, Doha, Kuwait
m.murugappan@kcst.edu.kw

[2]Faculty of Engineering and Information Technology, Southern University College, Johor Bharu, Malaysia

[3]School of Mechatronic Engineering, Universiti Malaysia Perlis, Ulu Pauh, Malaysia

[4]Computer Engineering Department, College of Engineering and Petroleum, Kuwait University, Kuwait

***Abstract*— Detecting different types of emotional expressions from the subject's face is important for developing intelligent systems for a variety of applications. This present work proposed virtual markers based on Facial emotion expression recognition using the Extreme Learning Machine (ELM) and Probabilistic Neural Network (PNN). A facial emotional expression database is developed with 55 undergraduate university students (male: 35, female: 20) of age range between 20 – 25 years with a mean age of 23.9 years. A HD webcam is used to capture the facial image and Haar Like features and Ada Boost classifier is used to detect the face and eyes through Open CV. A mathematical model based is used to place ten virtual markers called Action Units (AUs) on subjects face at a defined location. Later, Lucas-kanade optical flow algorithm is used to track the marker movement while the subject expressing different emotions and the distance between the center of the face to each marker is used as a feature for classifying emotions. One way Analysis of Variance (ANOVA) is used to test the statistical significance of the features and five fold cross-validation method is used to input the feature for classifiers. In this work, two non-linear classifiers namely, ELM and PNN are used for emotional expression classification. The experimental results give a maximum mean emotion classification rate of 88% and 92% in ELM and PNN classifiers, respectively. Maximum individual class accuracy of happiness – 96%, surprise – 94%, anger- 92%, sadness – 88%, disgust – 90% and fear 89% is achieved using PNN. The experimental results confirm that the proposed system is able to distinguish six different emotional expressions and could be used as a potential tool for a variety of applications which include, e-learning, pain assessment, psychological counseling, human-machine interaction-based applications, etc.**

Keywords—Face emotion recognition, Virtual Markers, Extreme Learning Machine, Probabilistic Neural Network

I. INTRODUCTION

Human to human communication is possible though exchange of emotions. One can perceive human emotion usually by studying the facial reactions of fellow human beings. In recent years, communication between humans and machines through emotions become active research in developing intelligent systems such as humanoids, nursing robots, etc. Hence, emotions play a significant role in human to human and human to machine communication for developing several intelligent systems. Besides, emotion recognition has been widely used in e-learning, psychological counseling, animation, pain assessment, smart environment designs, etc [1-3, 13]. Emotion recognition is one of the active fields of research in affective computing. Though emotions can be recognized through other modalities such as speech, biosignals (Electroencephalogram, facial Electromyogram), and gestures, facial emotion recognition is still popular due to the inexpensive system design and non-intrusive way of collecting user information [10-12]. Thereby, the researchers are started working on developing intelligent facial emotion recognition systems with more robustness. A comprehensive study on facial emotion recognition and different methods of assessment of 2D images is given in [5]. Most of the 2D images based facial expression recognition systems have several limitations in real-time and researchers started working on 3D facial images based facial expression recognition [9].

Facial Action Coding System (FACS) reports that a maximum of about 40 muscle locations in a human face responds to six basic emotional states (happiness, surprise, sadness, anger, disgust, and fear). Thereby, if it is possible to study the facial muscle responses, the emotional state of human can be perceived in real-time. The movement of facial muscle can be measured through Facial Action Units (FAU). Several researchers reported in the literature about the use of FAU based facial expression recognition, such as luminous stickers, facial masks, facial net, and others. In [2], the researchers have segmented the facial image into six segments and detected six basic emotions (happiness, sadness, fear, surprise, disgust, and anger) with a success rate of 83.3% using Lukas-Kanade Optical Flow Algorithm.

In [2], Neural Network based facial emotion classification is performed using angular features of 11 physical markers to classify six emotional states namely, anger, surprise, fear, sadness, happiness, and disgust) and achieved a maximum mean classification rate of 96%. Recently, Teng et.al have used Convolutional Neural Network to classify the four emotional expressions (joy, sadness, surprise and neutral) and achieved a maximum mean classification rate of 69.17% using facial image features [4]. In [5], the authors have proposed a dual feature (local region and weber local descriptor) fusion method to classify six emotional expressions (happiness, sadness, anger, fear, disgust, and surprise) and tested with three international standard databases. The maximum mean accuracy of 98.62%, 95.58% and 50% in Cohn-Kanade CK+, MMI and Static Face in the Wild (SFEW) databases, respectively. Convolutional features with Long Short Term Memory (LSTM) based deep learning network is used to classify the facial expressions in [6]. They have achieved a maximum mean classification rate of 99.73% in six classes.

Alphonse have used CK+ and MMI database for facial expression recognition using generalized supervised dimension reduction model to classify six basic emotions using Extreme Learning Machine (ELM) and Support Vector Machine (SVM) classifiers and achieved a maximum mean classification rate of 82.44%, 97.715%, 47% using CK+, MMI and SFEW databases, respectively. Conventionally, the facial images based emotion recognition method is widely used in the literature in offline and very limited works considered the real-time facial emotional expressions recognition. The researchers have analyzed the movement of lips, eyes and mouth region for extracting the features for emotion classification. However, processing the images with huge pixel information demands for higher computational memory and time. Hence, facial features based emotion recognition is highly needed for developing intelligent real-time systems.

Besides, the effects of background (static/dynamic), lighting, skin tone, facial hair, and neighboring environment significantly affect the accuracy of facial emotion recognition. Hence, developing an intelligent facial expression recognition system that works in real-time and has more robustness and with lesser computational complexity is highly challenging [14]. Hence, this present work mainly aims to classify the facial expressions using facial features called Virtual Marker distances using Optical Flow Algorithm (OFA). Here, the OFA effectively works in real-time and more reliable in handling facial images with background changes, lighting effects, and facial hairs. These facial features are fed into the machine learning algorithms (ELM and PNN) to classify the different facial expressions. The performance of the classifiers is assessed through the mean classification rate, sensitivity, and specificity.

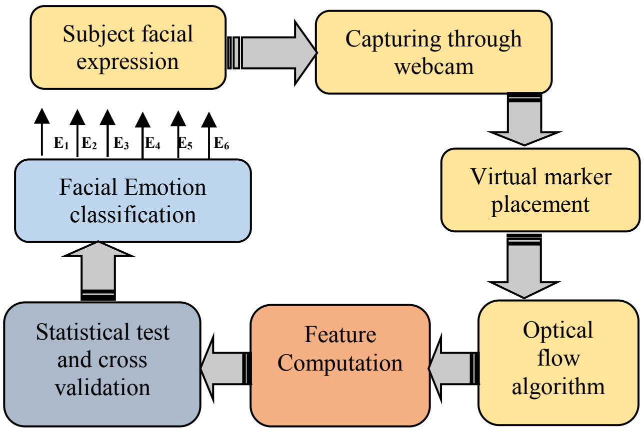

Fig. 1. Research methodology of facial emotion recognition using virtual markers.

II. METHODS AND MATERIALS

The complete research methodology of the present work is shown in Fig 1.

A. Database

In this work, the facial emotion recognition database of six basic emotions (happiness, surprise, anger, fear, disgust, and sadness) was developed with 55 undergraduate university students (35 male; 20 female) aged between 20 – 25 years with a mean age of 22.9 years. A built-in face time HD camera with a resolution of 2560 × 1600 at 227 pixels per inch is used to collect the facial images in a controlled environment (25^oC room temperature with 50 Lux lighting intensity). All the subjects are seated comfortably in a chair in front of the camera and the distance between the subject face to the camera is 0.95m.

A computerized PowerPoint slides are used to instruct the subjects to express the facial emotional expression by looking into the International Affective Picture System (IAPS) images of six different emotions. Fig 2 shows the flow of emotional expression sequences used in collecting the facial emotional expression samples for this work. Each emotion has 10 trials and each trial has a duration of 6 *sec*. In between the emotional expressions, 10 *sec* of break is given to the subjects to feel calm by showing natural scenes. The PowerPoint show starts with a set of instructions at the beginning of the experiment of 10 *sec* duration. Hence, the total time duration of the complete experiment is 7 min. The computing system continuously records the marker positions and saved it in comma-separated values (CSV) format for further processing.

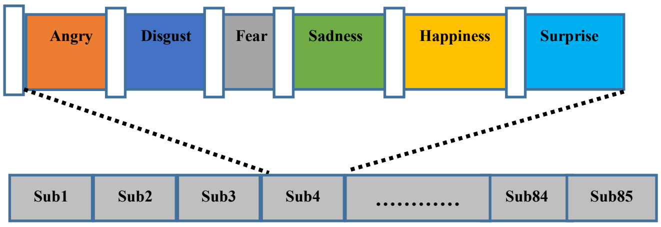

Fig. 2. Data acquisition protocol for emotion recognition

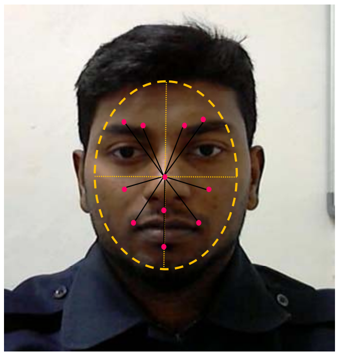

Fig. 3. Virtual markers placement

B. Virtual marker placement and distance measure

The first step in facial emotion recognition is face detection and there are different approaches have been proposed in the literature for detecting the faces [8]. In this work, the facial images were captured using the HD web camera and OpenCV library functions are used to implement the Viola and Jones face detection method to detect the faces. In Viola and Jones's method, Haar-like features are used to capture the face by using the Ada-Boost classification method. This method is a computationally inexpensive and simple approach to detect the faces and it works, based on pixel values than processing a whole image. Later, the captured RGB images are converted into greyscale images for the purpose of reducing the computational complexity of the proposed facial emotion recognition system. After the face detection, 10 virtual markers which are called Action Units (AUs) are placed on the acquired face through our mathematical model approach proposed in our earlier work (Fig 3). A detail description of virtual markers placement can be found in [12]. The initial position of the virtual markers was sent to Optical Flow Algorithm (OFA) to trace its positional changes. Lucas-Kanade algorithm is used for detecting future displacement of the marker. The distance between the center of the face to each marker is measured through the Euclidean distance method (Eqn 1).

$$\text{Distance} = \sqrt{(X_c - X_p)^2 + (Y_c - Y_p)^2} \quad (1)$$

where, (X_c, Y_c) is the coordinate of the center marker, and (X_p, Y_p) is the coordinate of a virtual marker.

C. Emotional expression classification

The extracted features were firstly fed into a k fold cross-validation method with a k value of 5 to split the training and testing set of features for classification. In 5 fold cross-validation method, the first fourfold of equal size of features were used for training and the 5th fold is used for testing. This procedure will be repeated five times with a different set of training and testing features and the average accuracy over five folds is reported in the results section. Later, these cross-validated features were used to classify six basic emotions using two non-linear classifiers namely Extreme Learning Machine (ELM) and Probabilistic Neural Network (PNN).

Extreme learning machine (ELM) is proposed by Huang et.al [15] and it is one of the powerful machine learning algorithms used in a variety of applications in signal and image processing domain. ELM is basically a feed-forward network with a single hidden layer compared to conventional neural network architecture. Hence, this classifier is computationally fast compared to other machine learning methods and it utilizes layered architecture for speeding up the computation. In this work, two different types of kernel functions used for emotional expression classification namely, Radial Basis Function (RBF) and Multi-Layer Perceptron (MLP). In MLP, we analyzed four different activation functions (sigmoid, tanh, hardlim and Gaussian) in the output layer for performance comparison. The grid search method is performed to find the optimal value of RBF width in the ranges of 0.01 to 0.1 with a step value of 0.01 and the hidden neurons of 1000 – 2500 with a step value of 100. The highest classification rate is achieved with tanh and Gaussian activation functions in MLP kernel compared to RBF kernel and other activation functions. In addition to ELM, the probabilistic neural network (PNN) is used for performance comparison on this work. In PNN, the value of standard deviation (sigma – σ) is varied with a step value of 0.01 in the ranges of 0.01 to 0.9 and the maximum mean classification rate is achieved at a value of σ = 0.01.

The performance of the classifier has been assessed using three performance measures, mean accuracy (Acc), sensitivity (Sen) and specificity (Spe). Mean accuracy is used to refer to the mean detection rate of all six classes for a given classifier. Sensitivity (Sen) and Specificity (Spe) refers to the efficiency of the classifier to detect a specific class of emotion as emotion and specific class of emotions as other emotion, respectively. The experimental results present the performance of the proposed emotional expression recognition system with the optimal value of hyper-parameters of ELM and PNN and it is reported in the following section. The following section also presents a detail experimental procedure, methodology, and experimental results. The proposed methodology has been implemented in MS Visual Studio and programmed using C++ language on a computing system with Intel *i7*, 8th generation processor with 16 GB RAM operating on Windows Operating System.

III. Results and Discussion

In this work, 55 subject's facial emotional expression data is used for developing the intelligent facial emotional expression recognition system. During the experiment, all the subjects were requested to look into the camera without moving or rotating their heads until the end of the experiment. The proposed system utilizes the virtual marker distance movement to distinguish different emotional expressions. Hence, the proposed work can be used to develop either a real-time (or) off-line based emotional expression recognition system. The distance between the center of the face to the marker is used as a feature to classify facial emotional expressions. The normalized feature values of each emotion with mean value are computed and presented through a Box plot and it's shown in Figure 4.

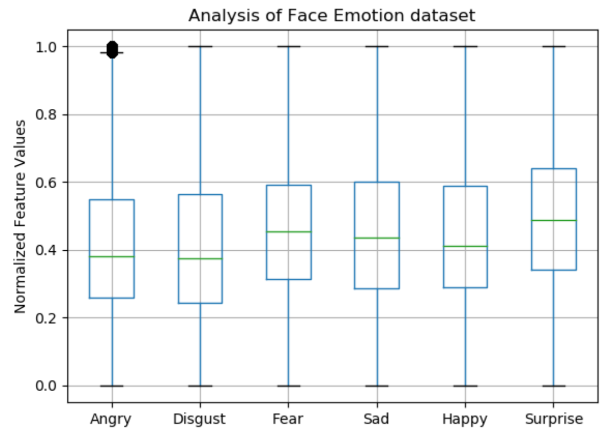

Fig. 4. Box plot analysis of facial emotional features

The box plot clearly shows that each emotional feature has a different mean value and shows larger variance between the emotions fear, sad, surprise and disgust and it could be easily distinguished with other emotions. One way analysis of variance (ANOVA) is performed to test the statistical significance of the distance feature in emotional expression classification. Table I presents the results of ANOVA on distance features. From the results, the high value of F and p indicates that the distance feature is highly efficient and significant in classifying six classes of emotions. The performance of two classifiers on emotional expression classification is presented in Table II. The maximum mean accuracy of the ELM classifier is 88% and PNN is 92% for six classes.

TABLE I. Statistical Analysis of Features through ANOVA

Feature	F Value	*p*-Value
Distance	4970.239	12.35×10^{-13}

TABLE II. CLASSIFICATION OF EMOTIONS USING ELM AND PNN CLASSIFIERS

Classifier	Hidden Layer / Types	Activation Function	Network parameters	Mean Accuracy ± Std Dev
ELM	RBF Kernel	Gaussian	Hidden Neurons = 2000; RBF width = 0.08	86±0.14
	MLP random layer	tanh	Hidden Neurons = 2000	88±0.20
		gaussian		88±0.24
		hardlim		87±0.16
		sigmoid		87±0.10
PNN	**σ = 0.01**			**92±0.15**

The individual class accuracy for two classifiers is given in Table III. The experimental results indicate that the PNN classifier outperforms ELM on giving higher individual class accuracy for facial expression recognition. The proposed system efficiently classifies six emotions with a maximum mean accuracy of 92%. Among the six different facial expressions, happiness and surprise emotions achieved a higher class accuracy than other emotions (sadness, disgust, fear, and anger). The major reason behind the lack of accuracy is due to the perception and expression of the subjects are different while expressing those emotions. Some of the subjects have not perceived the emotional feeling from the IAPS emotional pictures of the above emotions during the experiment. Since happiness and surprise are the most common emotional feelings that the subjects experience in their day to day life and it could easily perceive and expressed by them in the experiment.

The performance of ELM classifiers mainly depends on the three hyper-parameters namely, a number of hidden neurons, activation function at the output and kernel functions. Though ELM classifiers perform faster than other classifiers, ELM takes more computational time compared to PNN and it gives lower accuracy than PNN. Because the fine tuning of ELM classifier parameters through the grid-search method is computationally demanding compared to a single hyper-parameter (sigma) on PNN. Hence, the PNN classifier efficiently classifies the six emotional expressions with lesser computational time than ELM on this work. Though the accuracy of facial emotional expression recognition in the literature works achieved a high accuracy [2, 5-6] compared to the present work. But, this present work has the following features compared to earlier works: (a) utilizes the facial features in place of facial images, (b) requires lesser computational memory and time, (c) utilizes the virtual markers to detect the emotional states compared to physical markers, and (d) experimented with optical flow algorithm to detect the facial emotional expressions in real-time.

Though the proposed method achieves the higher emotional classification rate (92%), the system still has the following limitations: (a) the proposed system is only trained and tested with our facial emotion database and it has to be validated with international facial emotion databases for testing its robustness in emotion detection. (b) only a distance measure is used for emotion classification and the future work may focus on analyzing multiple features based on the marker movements. (c) the proposed system analyzes the performance of ELM and PNN classifiers only. Different types of machine learning algorithms such as regression tree, decision tree, neuro-fuzzy network, k nearest neighbor, and Random forest classifiers can also be used to analyze the performance of facial emotional expression recognition.

IV. CONCLUSION

In this work, facial features based on emotional expression recognition is performed through ELM and PNN classifiers. Ten virtual markers are placed on the subject facial images on defined locations and the position of the marker is continuously tracked by using the Lucas-Kanade optical flow algorithm. The virtual markers move while the subject expressing the facial expression and the distance between the center of the face to each marker is extracted through Euclidean distance measure for six emotional expressions. The one way ANOVA is used to test the statistical significance of the feature in distinguishing six different facial expressions. The maximum mean classification rate of 88% and 92% is achieved using ELM and PNN classifiers, respectively. Besides, happiness and surprise emotions give higher individual class accuracy compared to other emotions. The proposed system works in both real-time and offline and could be used as a potential tool for a variety of applications such as e-learning, pain assessment, human-machine interaction-based applications. Besides, the pixel-based approach significantly reduces the computational complexity (time and memory) of the system compared to the conventional facial emotion recognition system in the literature.

TABLE III. CLASSIFICATION OF EMOTIONS USING ELM AND PNN CLASSIFIERS

Classifier	Network parameter		Angry	Disgust	Fear	Sadness	Happiness	Surprise	Sensitivity	Specificity
ELM	RBF Kernel	Gaussian	91	80	80	82	90	90	97	85
	MLP random layer	tanh	90	88	85	84	92	91	97	86
		gaussian	85	86	88	87	92	90	97	84
		hardlim	89	88	82	80	94	91	97	86
		sigmoid	84	87	82	83	95	90	98	86
PNN	**σ= 0.01**		**92**	**90**	**89**	**88**	**96**	**94**	**98**	**90**

ACKNOWLEDGMENT

This work is financially supported by the Kuwait Foundation of Advancement of Sciences through an exploratory research grant. (Grant No: PR18-18EE-01).

REFERENCES

[1] S Macros, E Gonzalez, C Martin, L Flores, J Gomez, G Bermejo, E Zalama, "Virtual Avator for Emotion Recognition in Patients with Schizophrenia: A Pilot Study", Fron in Human Neuroscience, 2016. DOI: 10.3389/fnhum2016.00421.

[2] F Smith, S Rossit, “Identifying and detecting facial expressions of emotion in peripheral vision”, PlosOne, 13(5): e0197160. https://doi.org/10.1371/journal.pone.0197160

[3] Redferm, “Emotion Tracking for Remote Conferencing Applications using Neural Networks”, 21st National Conference on Artificial Intelligence and Cognitive Science, pp, 1-6, 2010.

[4] T Teng, X Yang, “Facial expressions recognition based on convolutional neural networks for mobile virtual reality”, 15th ACM SIGGRAPH Conference on Virtual-Reality Continuum and its Applications in Industry, Vol 1, 475-478, 2017.

[5] C Pelachaud, “Modelling multimodal expression of emotion in a virtual agent”, Phil. Trans. R. Soc, 364, pp, 3359-3348, 2009.

[6] A Mahmood, S Hussain, K Iqbal, W Elkilani, Recognition of Facial Expressions under Varying Conditions using Dual-Feature Fusion, Math. Prob. In Engg, 2019, 1-12. DOI: 10.1155/2019/9185481

[7] Z Yu, G Lu, Q Liu, J Deng, “Spatio-temporal convolutional features with nested LSTM for facial expression recognition”, Neurocomputing 317, 50-57, 2018.

[8] A Sherly, D Dharma, Novel directional patterns and Generalized Supervised Dimension Reduction System for facial emotion recongition, Mult. Tools. and Appl, 2017. DOI: 10.1007/s11042-017-5141-8.

[9] L Feldman, R Adolphs, S Marsella, Aleix M, S Pollak, “Emotional Expressions Reconsidered: Challenges to Inferrring Emotions from Human Facial Movements”, Physc. Sci. in the Pub Int, 20(1), 1-68, 2019.

[10] F Nonis, N Dagnes, F Marcolin, E Vezzetti, 3D Approaches and Challenges in Facial Expresion Recognition Algorithms – A Literature Review, Applied Sciences, 9 (3904), 2019. DOI: 10.3390/app9183904.

[11] M Murugappan, R Nagarajan, S Yaacob, “Combining spatial filtering and wavelet transform for classifying human emotions using EEG signals”, Journal of Med. and Bio Engg, 31(1), 45-51, 2011.

[12] R Yuvaraj, M Murugappan, M Norlinah, K Sundarj, M Khairiyah, “Review of emotion recognition in stroke patients”, Dementia and Geriatric Cognitive Disorders, 36(3-4), 179-196, 2013.

[13] V Maruthapillai, M Murugappan, “Optimal Geometric Set for Automated Marker Placement to Virtualized Real-Time Facial Emotions”, PlosOne, 11(2), e0149003, 2016. DOI: 10.1371/journal.pone.0149003.

[14] P Quelhas, J Stoyanovam Marker versus Markerless Augmented Reality. Which Has More Impact on Users?, Int. Jou, of Hum. Comp. Inter, 34(9), 819-833, 2017.

[15] J Tang, C Deng, G B Huang, “Extreme Learning Machine for Multilayer Perceptron”, IEEE Trans Neural Networks and Learning Systems, 27(4), 809-821, 2015.