Facial Region Segmentation Based Emotion Recognition Using K-Nearest Neighbors

Firoz Mahmud, Bayezid Islam, Arfat Hossain, Pushpen Bikash Goala
Department of Computer Science & Engineering
Rajshahi University of Engineering & Technology
Rajshahi-6204, Bangladesh
fmahmud.ruet@gmail.com, bayezid.shouvik@gmail.com, arafat.ruet.bd@gmail.com, pushpen.bikash@gmail.com

***Abstract*—As part of human emotion recognition, a facial expression recognition method is proposed where segmentation of facial regions are done manually in a unique yet effective way by analyzing many human faces and the position of the right eye, left eye, nose and mouth in those faces. Feature extraction from the segmented parts are done using 2D Gabor filter, redundant features are eliminated using downsampling of the extracted features and finally, classification of the expressions are done using K-Nearest Neighbors (KNN) classification technique. To evaluate the performance of the proposed method CK+, RaFD and KDEF datasets are used. For the CK+, RaFD and KDEF dataset the recognition rates are 99.75 %, 94.60 % and 86.02 % respectively which indicates the effectiveness of the proposed method.**

Keywords—Emotion Recognition, Facial Expression Recognition (FER), Image Preprocessing, Image Segmentation, Gabor Filter, K-Nearest Neighbors (KNN)

I. Introduction

Human emotions play a vital role in determining the mental condition of a human. So the ability to detect the emotion has a great importance in many sectors of our life. This is the main reason why there have been many types of research to find out the most effective way of determining human emotions. But according to A. Mehrabian facial expression contributes the most (55%) to express emotions of a speaker comparing with voice (38%) and words (7%) [1], so if the facial expression of a human can be recognized properly then the possibility of correctly determining the emotion is really high.

Successfully recognized facial expressions can be used in social communications, determining whether a driver can be allowed to drive a vehicle or not or in any field where the disturbed emotion of a human can cause harm if the person is given to do any critical job. It can also be used in many other sectors to make our everyday life more safe and comfortable. These are the reasons why facial expression recognition have been a prime research topic in the field of computer science for several years and a lot of works have been done on this topic that apply different techniques to recognize facial expressions more accurately and in the least possible time.

Different classification techniques have been applied to recognize facial expressions among them LDA [2], AdaBoost classification [3-4] with neural network classifiers [5], multi-class Support Vector Machine (SVM) [6], multi-layer perceptron with backpropagation [7], Extreme Learning Machine (ELM) [8], K-Nearest Neighbors (KNN) [9-14] are some worth mentioning. Gabor filter [15] was also used for facial analysis. But all of them have some issues to deal with and as a result, the performance is not up to the desired level. Neural network classifier and multi-layer perceptron with backpropagation technique requires a large training sample to work well and long training time so these are not suited for real-time application perfectly. SVM requires a huge resource for computation so it is not applicable where the computational resource is not sufficient. ELM solves the issues mentioned above but it is tough to find optimal solution using ELM and it might be affected with local minima and overfitting issue. The works involving KNN have issues regarding the extraction of useful features for recognizing the test expressions.

So to solve the above mentioned issues, in this paper a facial expression recognition model is proposed where Viola-Jones object detection method along with a newly proposed method was used for segmentation of images, then Gabor filter was used for the extraction of informative features from the segmented images and finally KNN was used for classifying the test images.

II. Proposed Method

The proposed model has a few steps which are illustrated in the Fig. 1. From the figure, it is observable that the first step of the model is to convert the image into grayscale if it is a color image. The next step is to detect only the facial region from the whole image for using it in the subsequent stages. Viola-Jones object detection method [16] is used for it. Later the facial region of the input image is resized into a fixed size for using in the later stages of segmentation. Then the resized image is manually segmented to get the four facial region parts right eye, left eye, nose and mouth from it. Features are extracted from these four segmented parts and Gabor filter is used to do this portion of the overall task. The features of an image are given a label depending on the expression of the image and in this way all of the images except the testing images are labeled. Using these features and labels KNN is used to find out the expression of the test images by finding the distance from members of different classes depending on the value of K. Finally, the performance of the proposed model is evaluated by comparing the expected expression and found expression of the test images. Fig. 1 illustrates the whole process step by step.

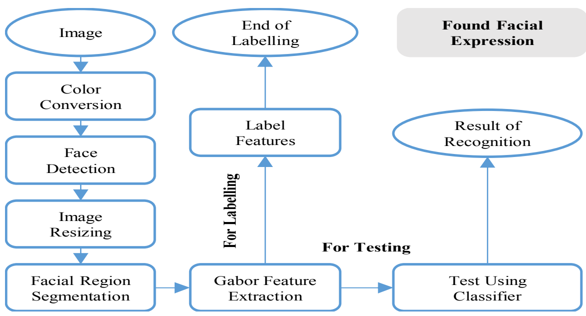

Fig. 1. Proposed model of facial expression recognition.

978-1-5386-5229-9/18/$31.00 ©2018 IEEE

III. IMAGE PREPROCESSING

An input image is checked whether it is a color image or a grayscale image. If it is a color image then it is converted to grayscale otherwise this step is omitted. Then Viola-Jones face detection method [16] is used to effectively detect the facial region from an image. Then the facial image is resized to a fixed size of 150 x 150 pixels. The process is shown in Fig. 2 on a sample image from the KDEF dataset [17].

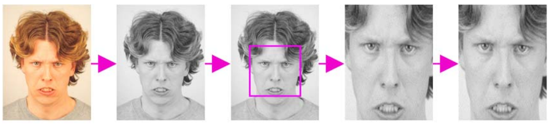

Fig. 2. Steps of image preprocessing.

IV. IMAGE SEGMENTATION

This is one of the most significant parts of this paper. There are different methods available to do this task such as Active Appearance Models (AAM) [18], Viola-Jones objects detection [16] and others but each has some issues. For example when the eye is almost closed or the mouth is highly skewed at a direction then Viola-Jones method find it difficult to detect that part properly. Besides the Viola-Jones method detect the four parts of different sizes for different images (mouth, eyes, nose with different sizes for different images). High dependency on the training set and proper initialization are issues with AAM. As the eyes, nose, mouth are usually placed at any certain position of the human face so in normal cases with frontal images it is not impossible to segment these four parts manually according to our method and our experimental result also indicates that. At first, four separate points for four regions are defined and then the four parts are segmented by using four different width and height values for the four parts. These values are selected by analyzing many facial images and the position of eye, nose, and mouth in these images. For the right eye, the specific point is at 23.55, 46.58 and its width and height values are 46 and 31 respectively. So the point 23.55, 46.58 on any 150 x 150 image is found and then a portion of the whole image with width 46 and height 31 is cropped for segmenting the right eye. Coordinate values and corresponding width, height values are mentioned in TABLE I.

TABLE I. VALUES FOR SEGMENTATION

Facial Part	Coordinate (x,y)	Width (w)	Height (h)
Right Eye	23.55,46.58	46.01	30.67
Left Eye	88.29,46.58	44.00	29.15
Nose	54.33,81.84	45.43	38.00
Mouth	50.24,114.0	57.00	34.18

If the above-mentioned values are applied to an image of size 150×150 then the image segmentation process is shown in Fig. 3.

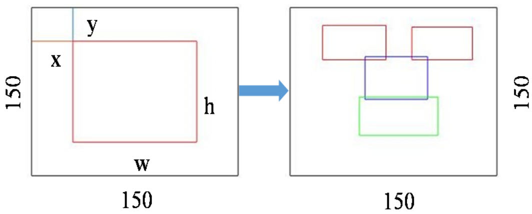

Fig. 3. Proposed image segmentation method on a block of size 150 × 150.

The process is illustrated in a brief in Fig. 4 on a sample image from the KDEF dataset [17].

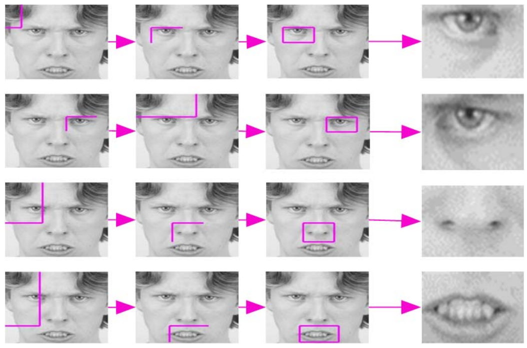

Fig. 4. Step by step segmentation of a facial region into four parts.

V. FEATURE EXTRACTION

Feature extraction is an important part of the whole process as the extracted features drive the classifier to take any decision when it comes to classification. Linear Discriminant Analysis (LDA) [19], Principal Component Analysis (PCA) [19-20] are some of the most popular feature extraction techniques. To encode the features one-dimensional Log-Gabor wavelets are used effectively in [21]. But in the proposed method two-dimensional Gabor filter is used to extract the features from the segmented parts of the images because of its advantages. Gabor filters are robust against photometric disturbances, such as illumination changes and image noise which is one of its advantages. Other advantages include their invariance to rotation, scale and translation [22]. A two-dimensional Gabor filter is a Gaussian kernel function modulated by a complex sinusoidal plane wave, defined as

$$G(x,y) = \frac{f^2}{\pi\gamma\eta} e^{\left(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right)} e^{(j2\pi f x' + \o)}$$

$$x' = x\cos\theta + y\sin\theta \qquad (1)$$

$$y' = -x\sin\theta + y\cos\theta$$

Where f is the frequency of the sinusoid, θ is the orientation of the normal to the parallel stripes of a Gabor function, ø is the phase offset, σ is the standard deviation of the Gaussian envelope and γ is the spatial aspect ratio which specifies the ellipticity of the support of the Gabor function [22].

While implementing, a Gabor filter bank with five scales and eight orientations was generated using equation (1) with $\gamma = \sqrt{2}$,$\eta = \sqrt{2}$ and the highest value of frequency was set to 0.25 which are empirical value. The magnitude and real part of the generated Gabor filter bank are shown in Fig. 5.

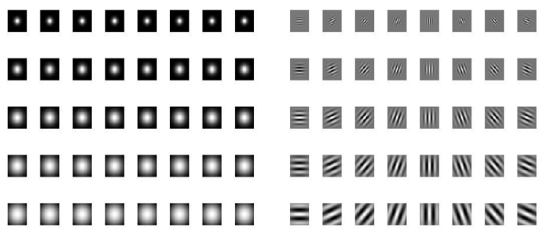

Fig. 5. The magnitude and real part of a generated Gabor filter bank.

The generated Gabor filter bank is then used to filter the segmented parts of images but by doing so it would generate a feature vector which would be of size 40 x m x n if the image to be filtered is of size m x n as there are 40 Gabor filters in the bank. To overcome this problem and reduce the size of the filtered image we downsampling [22] is performed to eliminate the redundancy of information and it is done by a factor of 4 and finally feature vector is formed from the filtered image and size of the feature vector is (40 x m x n)/(4 x 4). For right eye it is (40 x 46 x 31)/(4 x 4)=3565, for left eye it is (40 x 45 x 30)/(4 x 4)=3375, for nose it is (40 x 47 x 39)/(4 x 4)=4582 and for mouth it is (40 x 58 x 35)/(4 x 4)=5074.

VI. CLASSIFICATION

KNN is one of the simplest algorithms that can be used for classification. Despite its simplicity, it is quite effective as a classifier. It was first proposed [23] by T. M. Cover and P. E. Hart in 1967 but later modified by different researchers to improve the performance of KNN. It performs well when the number of sample is large. The basic concept of KNN as a classifier is that it will have several training samples in the feature space and whenever a new test sample appears, the distance from other trained samples are measured depending on the value of K and the test sample is determined a member of that class whose sample has minimum distance with the test sample. If K=1 then the test sample is assigned to the class of single nearest neighbor. However, finding the appropriate value of K for a specified problem is an issue which affects the performance of KNN. Different techniques are used for measuring the distance such as Euclidean, cityblock, cosine, correlation, hamming etc. In our work Euclidean distance is used which is defined as,

$$d(p,q) = d(q,p) = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2} \qquad (2)$$

The feature vectors collected previously from the feature extraction step are used in this step to measure the distance and find the class of any test image. The process is shown in Fig. 6.

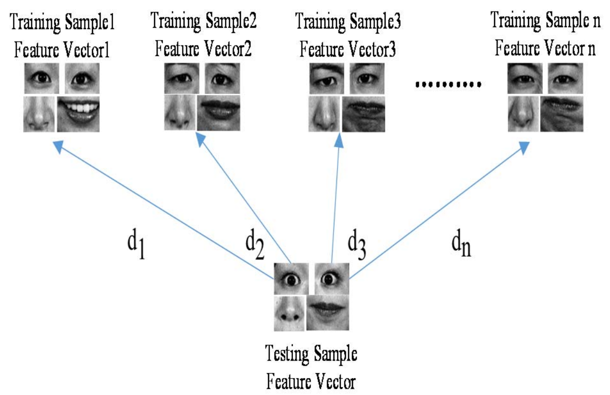

Fig. 6. Distance among the feature vectors.

From the figure, it can be seen that when a new testing sample arrives then the distance between its feature vector and other K training samples' feature vectors are measured and the minimum distance is found out. The testing sample is assigned the class of that corresponding sample's class who has the minimum distance or which class has the maximum vote depending on the distances.

VII. EXPERIMENTAL RESULT

A computer with Intel Core i5 processor and 64-bit system was used to implement the proposed method. Matlab-2017a was used for the task. CK+ (Extended Cohn-Kanade) [24], RaFD (Radbound Faces Database) [25], KDEF (Karolinska Directed Emotional Faces) [17] datasets were used to evaluate the performance of the proposed method. From CK+ dataset 1219 facial expression images of 22 different males and females were used for the experiment. 1407 front facing images from 67 persons and 980 images from 70 persons were used from RaFD and KDEF datasets respectively. So in total 3606 images from 159 different male, female and children were used. Basic facial expressions of these images were neutral, happy, anger, surprise, fear, disgust, sadness. To avoid biased results 10 folds cross-validation were used and the average Correct Recognition Rate (CRR) of all these folds are recorded as the CRR of the system. The average and maximum accuracy achieved by the proposed system using a different number of folds and different datasets are mentioned in TABLE II.

TABLE II. THE ACCURACY OF DIFFERENT DATASETS

No. of Folds	Accuracy Achieved (%)					
	CK+		*RaFD*		*KDEF*	
	Avg.	*Max.*	*Avg.*	*Max.*	*Avg.*	*Max.*
2	98.11	98.20	85.43	85.78	85.92	87.55
3	99.67	99.75	89.27	90.41	85.61	89.30
4	99.76	100	90.90	92.61	85.82	88.57
5	99.75	100	92.82	94.33	85.92	89.80
10	99.75	100	94.60	97.16	86.02	89.80

The above-mentioned information is graphically represented in Fig. 7 which is a plot of CRR versus percentage of training samples on different datasets.

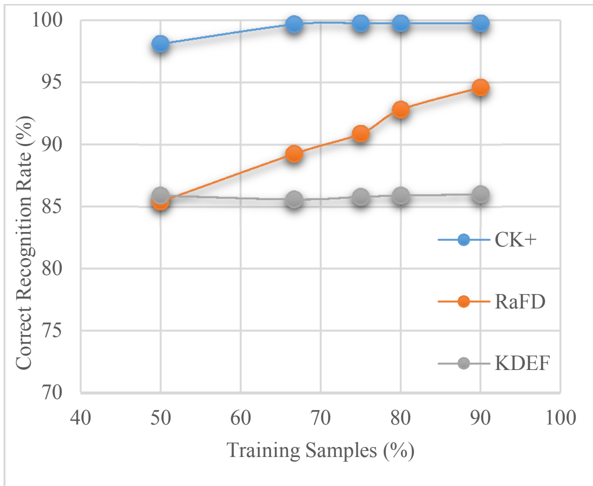

Fig. 7. Performance comparison on CK+, RaFD and KDEF dataset.

To analyze which type of expressions are classified properly and which are not, confusion matrices are analyzed in FER problems. TABLE III, TABLE IV and TABLE V are confusion matrices on CK+, RaFD and KDEF datasets with random test cases from 5 folds. As these are randomly generated test cases so the accuracy calculated from the confusion matrices might be slightly different from TABLE II. Entries of these confusion matrices represent the values in percentage. Ne, Ha, An, Su, Fe, Di, Sa represents neutral, happy, angry, surprise, fear and disgust facial expressions respectively. Exp. means one of the seven facial expressions used.

TABLE III. CONFUSION MATRIX OF CRR ON CK+ DATASET

Exp.	Ne	Ha	An	Su	Fe	Di	Sa
Ne	100	0	0	0	0	0	0
Ha	0	100	0	0	0	0	0
An	0	0	100	0	0	0	0
Su	2.86	0	0	97.14	0	0	0
Fe	0	0	0	0	100	0	0
Di	0	0	0	0	0	100	0
Sa	0	0	0	0	0	0	100

TABLE IV. CONFUSION MATRIX OF CRR ON RAFD DATASET

Exp.	Ne	Ha	An	Su	Fe	Di	Sa
Ne	95.45	0	2.27	0	0	0	2.27
Ha	0	100	0	0	0	0	0
An	4.65	0	88.37	0	0	0	6.98
Su	0	0	0	88.57	11.43	0	0
Fe	6.25	0	0	3.12	90.63	0	0
Di	1.96	0	0	0	0	98.04	0
Sa	2.33	0	0	0	0	0	97.67

TABLE V. CONFUSION MATRIX OF CRR ON KDEF DATASET

Exp.	Ne	Ha	An	Su	Fe	Di	Sa
Ne	96.43	0	0	0	3.57	0	0
Ha	0	100	0	0	0	0	0
An	3.57	0	85.72	0	3.57	0	7.14
Su	3.57	0	0	89.29	7.14	0	0
Fe	0	3.57	0	14.29	64.29	0	17.85
Di	0	3.57	3.57	0	0	92.86	0
Sa	17.86	0	0	0	3.57	0	78.57

From the confusion matrix of the KDEF dataset, it is evident that fear expressions had the lowest CRR. From these confusion matrices and the information of TABLE II, it is apparent that the proposed method struggles to classify the expressions from the KDEF dataset. In KDEF dataset for any specific model's any specific expression there are only two images and as the train and test sets are randomly generated for performance evaluation so there is a high possibility that some expressions were placed only in the test set without any corresponding expression image in the train set. It could degrade the performance of the system.

VIII. STATE-OF-THE-ART

In this section, the performance of the proposed method is compared with some other emotion recognition (facial expression recognition) methods. While choosing other methods for comparison, the dataset used by those other methods were considered and only few promising methods with CK+, RaFD, KDEF datasets were considered for comparison.

TABLE VI. STATE-OF-THE-ART COMPARISON

Study	Technique	Dataset	Acc.
Proposed Method	Viola-Jones face detection, **facial region segmentation**, 2D Gabor filter, downsampling, KNN	CK+ RaFD KDEF	**99.75** **94.60** **86.02**
2015 [26]	Speeded Up Robust Features (SURF), gentle AdaBoost	RaFD KDEF	90.64 74.05
2017 [27]	Visual geometry group model, CNN architecture network, deep learning	CK+ RaFD	99.33 93.33
2017 [28]	Discriminative Feature Dictionary (DFD), Vertical 2DLDA (V-2DLDA)	CK+ KDEF	91.87 82.24
2018 [29]	Histogram equalization, deep PCA, ELM	KDEF	83.00

IX. CONCLUSION

The effectiveness of the proposed model can be evaluated from the experimental results. High facial expression recognition rate indicates the effectiveness of the proposed model. Effective segmentation procedure along with an effective feature extraction technique helped to achieve high accuracy. But as the training and testing images were randomly selected so every time different training and testing images were selected which resulted in different recognition rate. Some facial expressions were too confusing and difficult for humans to recognize, let alone machines. It affected the overall performance of the proposed model.

A flaw of the proposed model is that if the image is not a frontal one and rotated at any angle in any direction then it might not be segmented properly if it is not manually rotated again to make it a frontal one. Besides images with abnormal facial structure might also face difficulties to get segmented properly according to the proposed model. Future work is to overcome the above-mentioned problems and to develop a more robust model to handle such scenarios.

As facial expression is not the only criteria to define emotion so a multimodal emotion recognition model as a fusion of both facial expression recognition and speech recognition could be an interesting and challenging work in future and it could be used to recognize human emotions effectively which is also the main goal of facial expression recognition.

REFERENCES

[1] A. Mehrabian, "Communication without Words", *Psychology Today*, Vol. 2, No. 4, pp.53-56, 1968.

[2] P. N. Belhumeur, J. P. Hespanha, D. J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linearprojection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, pp. 711-720, 1997.

[3] K. -H. An, M. –J. Chung, "Learning Discriminative MspLBP Features Based on Ada-LDA for Multi-Class Pattern Classification", *IEEE International Conference on Robotics and Automation Anchorage Convention District*, Anchorage, AK, May 3-7, pp. 4803-4808, 2010.

[4] C. –S. Fahn, M. –H. Wu, C. –Y. Kao, "Real-time Facial Expression Recognition in Image Sequences Using an AdaBoost-based Multi-classifier", *Proceedings of 2009 APSIPA Annual Summit and Conference*, Sapporo, Japan, October 4-7, 2009.

[5] K. Takahashi, S. Takahashi, Y. Cui, M. Hashimoto, "Remarks on Computational Facial Expression Recognition from HOG Features Using Quaternion", *Engineering Applications of Neural Networks*, Vol. 459, pp. 15-24, 2014.

[6] X. Zhang, M. H. Mahoor, R. M. Voyles, "Facial Expression Recognition using HessianMKL Based Multiclass-SVM", *the 10th IEEE International Conference on Automatic Face and Gesture Recognition*, Shanghai, China, April 22-26, pp. 1-6, 2013.

[7] M. Owayjan, R. Achkar, M. Iskandar, "Face Detection with Expression Recognition using Artificial Neural Networks", *3rd Middle East Conference on Biomedical Engineering (MECBME)*, Beirut, Lebanon, pp. 115-119.

[8] F. Mahmud and M. A. Mamun, "Facial Expression Recognition System Using Extreme Learning Machine", *International Journal of Scientific & Engineering Research*, Vol. 8, No. 3, pp. 26-30, 2017.

[9] J. Ou, "Classification Algorithms Research on Facial Expression Recognition", International Conference on Solid State Devices and Materials Science, Vol. 25, pp. 1241-1244, 2012.

[10] A. S. Md. Sohail, P. Bhattacharya, "Classification of facial expressions using K-nearest neighbor classifier", *MIRAGE'07 Proceedings of the 3rd international conference on Computer vision/computer graphics collaboration techniques*, Rocquencourt, France, March 28-30, pp. 555-566 ,2007.

[11] R. A. Nugrahaeni, K. Mutijarsa, “Comparative analysis of machine learning KNN, SVM, and random forests algorithm for facial expression classification”, *2016 International Seminar on Application for Technology of Information and Communication (ISemantic)*, pp. 163-168, 2016.

[12] Q. Wang, K. Jia, P. Liu, “Design and Implementation of Remote Facial Expression Recognition Surveillance System Based on PCA and KNN Algorithms”, *2015 International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)*, pp. 314-317, 2015.

[13] H. Alshamsi, H. Meng, M. Li, “Real time facial expression recognition app development on mobile phones”, *2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, pp. 1750-1755, 2016.

[14] S. Dongcheng, C. Fang, D. Guangyi, “Facial Expression Recognition Based on Gabor Wavelet Phase Features”, *2013 Seventh International Conference on Image and Graphics*, pp. 520-523, 2013.

[15] A. Matin, F. Mahmud, T. Ahmed and M. S. Ejaz, "Weighted score level fusion of iris and face to identify an individual," *2017 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, Cox's Bazar, 2017, pp. 1-4.

[16] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features", *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, pp. I-511-I-518, Vol.1, 2001.

[17] D. Lundqvist, A. Flykt, A.Öhman, "The karolinska directed emotional faces (KDEF)", *CD ROM from Department of Clinical Neuroscience Psychology section Karolinska Institutet*, pp. 91-630, 1998.

[18] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active Appearance Models", *IEEE Trans on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 6, pp. 681-685, 2001.

[19] F. Mahmud, M. T. Khatun, S. T. Zuhori, S. Afroge, M. Aktar and B. Pal, "Face recognition using Principle Component Analysis and Linear Discriminant Analysis", *2015 International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)*, Dhaka, 2015, pp. 1-4.

[20] F. Mahmud, S. Afroge, M. A. Mamun and A. Matin, "PCA and back-propagation neural network based face recognition system," *2015 18th International Conference on Computer and Information Technology (ICCIT)*, Dhaka, 2015, pp. 582-587.

[21] A. Matin, F. Mahmud, S. T. Zuhori and B. Sen, "Human iris as a biometric for identity verification", *2016 2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE)*, Rajshahi, 2016, pp. 1-4.

[22] M. Haghighat, S. Zonouz and M. A. Mottaleb, "CloudID: Trustworthy cloud-based and cross-enterprise biometric identification", *Expert Systems with Applications*, Vol. 42, No. 21, pp. 7905-7916, 2015.

[23] T. M. Cover and P. E. Hart, “Nearest Neighbor Pattern Classification”, *IEEE Transactions on Information Theory*, Vol. 13, No. 1, pp. 21-27, January,1967.

[24] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar and I. Matthews, “The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression”, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, San Francisco, CA, June 13-18, pp. 94-101, 2010.

[25] O. Langner, R. Dotsch, G. Bijlstra, D.H. Wigboldus, S.T. Hawk, A. van Knippenberg, "Presentation and validation of the Radboud Faces Database", *Cognition and Emotion*, Vol. 24, pp. 1377-1388, 2010.

[26] Q. Rao, X. Qu, Q. Mao and Y. Zhan, "Multi-pose facial expression recognition based on SURF boosting", *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, Xi'an, 2015, pp. 630-635.

[27] A. Fathallah, L. Abdi and A. Douik, "Facial Expression Recognition via Deep Learning", *2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA)*, Hammamet, 2017, pp. 745-750.

[28] Z. Sun, Z. Hu, M. Wang and S. Zhao, "Discriminative feature learning-based pixel difference representation for facial expression recognition", in *IET Computer Vision*, Vol. 11, No. 8, pp. 675-682, 12 2017.

[29] K. Rujirakul and C. So-In, "Histogram equalized deep PCA with ELM classification for expressive face recognition", *2018 International Workshop on Advanced Image Technology (IWAIT)*, Chiang Mai, 2018, pp. 1-4.