A Facial Region Segmentation Based Approach to Recognize Human Emotion Using Fusion of HOG & LBP Features and Artificial Neural Network

Bayezid Islam, Firoz Mahmud, Arafat Hossain, Pushpen Bikash Goala, Md. Sumon Mia
Department of Computer Science & Engineering
Rajshahi University of Engineering & Technology
Rajshahi-6204, Bangladesh
Emails: (bayezid.shouvik, fmahmud.ruet, arafat.ruet.bd, pushpen.bikash, sumon.ruetbd)@gmail.com

Abstract—**Mental condition and sentiment of a person can be analyzed through facial expressions. An emotion recognition system is proposed by recognizing facial expressions. Input images are preprocessed and then proposed image segmentation method is applied to segment a facial image into four parts that contribute highly in representing facial expressions. Features are extracted from the segmented parts using a fusion of Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP). The dimension of the feature vector is reduced using Principal Component Analysis (PCA). Finally, Artificial Neural Network (ANN) is used to classify the facial expressions properly. The proposed system is tested using three widely used facial expression datasets (JAFFE, CK+, RaFD). At last, the achieved performance is compared with other facial expression recognition systems to justify that the proposed method succeeds in achieving state-of-the-art performance.**

Keywords—Emotion Recognition, Facial Expression Recognition (FER), Image Preprocessing, Image Segmentation, Fusion of HOG and LBP, Principal Component Analysis (PCA), Artificial Neural Network (ANN), Multilayer Perceptrons (MLP), Backpropagation

I. INTRODUCTION

There are different modes of expressing the feelings of a person. Among them, facial expression is the highest (55%) contributory factor [1]. So if facial expressions can be recognized successfully, then there is a high probability of determining the feelings of a person successfully. Facial expressions are also useful for nonverbal communication and social communication. They can be used for security purposes as well. Mental disorders can also be analyzed using facial expressions.

The emotion of a person can be analyzed to determine whether a person is fit or not for a sensitive task. Human-like robots can be given Facial Expression Recognition (FER) ability to get better performance from them. Social networks can add such features which will suggest user post status depending on the expression of their uploaded photo. Some smartphone cameras already have the feature of capturing the photo by detecting the smiling face of the user. As everything is getting automated, so the ultimate goal would be to give machines the ability to recognize facial expressions spontaneously as a human can do.

Usually, seven basic facial expressions are considered in FER problems and they are angry, sad, surprise, happy, fear, disgust, neutral [2]. Due to its growing applications in various fields, FER has been an active research topic in computer vision and human-computer interaction track for the last few decades. Basic steps of a simple FER system includes preprocessing of images, analyzing regions of interest, extraction of features and classification. Different FER systems perform the first two steps differently. But face detection is a common part of image preprocessing. Proposed method segments the facial image into four parts to use them as regions of interest.

Popular feature extraction techniques include Principal Component Analysis (PCA) [3-7], Linear Discriminant Analysis (LDA) [3], Gabor wavelets [8], moments, gray level co-occurrence matrix, Speeded-Up Robust Features (SURF), Scale Invariant Feature Transform (SIFT) and many others. Gabor features are robust and effective against photometric disturbances but produce feature vector with gigantic dimension. SURF features only concentrate on the few strongest features. To avoid such problems, both Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP) are used to extract features. Combination of HOG and LBP has been successfully used in many FER systems such as in [9].

Different classification techniques have been used in FER systems as well. Support Vector Machines (SVM) [10], nearest neighbor classification [11], AdaBoost classification [12], random forests [10], Extreme Learning Machine (ELM) [13], Artificial Neural Network (ANN) [14-16] are among the most used classifiers. Overfitting, proper value of regularization term, proper kernel function are issues with SVM. Number of neighbors to be considered and proper distance metric are challenges with nearest neighbor classification. Noisy data can be problematic if AdaBoost classification is used. Random forests are tough to visualize. ELM is prone to overfitting. To overcome the issues with these classifiers, ANN is used for the classification task by the proposed method. The proposed FER system is briefly mentioned in the next section. Each step of the proposed method is elaborately described in the subsequent sections. Last few sections are reserved for result analysis, state-of-the-art comparison and discussing the flaw, scopes of the work.

II. PROPOSED METHOD

As the first step, the input image is converted to grayscale if it is a color image. Otherwise, this step is ignored. Then only the face region is detected from the grayscale image using Viola-Jones face detection method [17]. The facial image is then converted to a fixed size. Combination of these steps results in the preprocessed image. The preprocessed image is then segmented into four facial expression regions (right eye, left eye, nose, mouth) using the proposed segmentation method. Features from these segmented parts are extracted using both HOG and LBP. Length of the feature vector is reduced by applying PCA on it. Finally, for training the system ANN is used with backpropagation. Performance of the system is then tested with the test samples. The flow diagram of the proposed FER system is illustrated in Fig. 1.

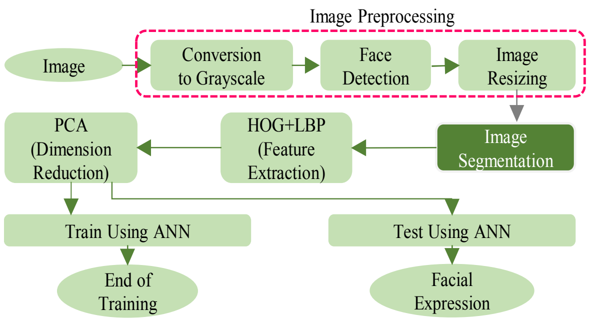

Fig. 1. The flow diagram of the proposed FER system.

III. Image Preprocessing

If the input image is a color image then it is converted to grayscale. If it is not a color image but a grayscale image then there is no need to perform this step. If the image contains any face in it then that facial region is detected from the whole image using Viola-Jones face detection method [17]. The facial region is then converted to a fixed size of 150×150 pixels for using it conveniently in the image segmentation step. The steps of image preprocessing are shown on a sample image from RaFD dataset [18] in Fig. 2.

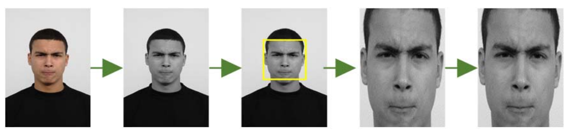

Fig. 2. Steps of image preprocessing.

IV. Image Segmentation

Choosing appropriate regions of interest that contribute most in representing facial expression is a crucial step. Right eye, left eye, nose and mouth are four facial parts that contribute highly in facial expressions and so these four parts are considered in the proposed FER system. Viola-Jones object detection method [17] and Active Appearance Models (AAM) [19] are two popular techniques for doing such tasks. But when the eye is closed or almost closed then Viola-Jones method fails to detect the eye region. Dimension mismatch of same parts of different images is another concern with Viola-Jones method. Wide variations of objects in training set and dependency on proper initialization are challenges with AAM. To overcome these issues a manual segmentation method is proposed in this paper. This method works by finding a defined coordinate point from an image of size 150×150 pixels and using a region of defined width and height from that coordinate point. For understanding the process, let us consider the example of segmenting mouth from a facial image of size 150×150 pixels. At first, the coordinate point 50.24,114.0 is found from the facial image. Then a region with a width value of 57 and a height value of 34.18 is selected and used as the mouth region of that image. The same procedure is followed to segment other three parts by using different coordinate, width, height values. These 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

These values are defined by analyzing many facial images and the position of four parts in those images. If the values of TABLE I are applied to an image of size 150×150 then how that image gets segmented into four parts is shown in Fig. 3.

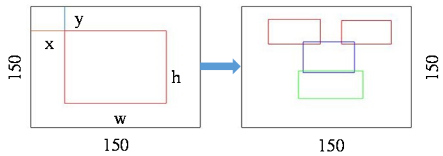

Fig. 3. Proposed image segmentation method on a block of size 150 × 150.

To understand the proposed image segmentation method properly, the segmentation of all four parts are step by step shown in Fig. 4 on sample image from RaFD dataset [18].

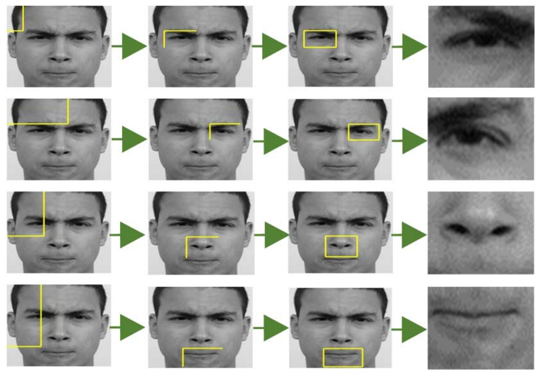

Fig. 4. Proposed image segmentation method (step by step) on a sample image.

V. Feature Extraction

Useful features from the four segmented facial parts were extracted using a fusion of HOG and LBP features.

A. Histogram of Oriented Gradients (HOG)

HOG features are robust against photometric and geometric transformations. These are just a few among the benefits that compelled to select HOG as one of the feature extraction techniques. HOG has been extensively used in many applications successfully. A popular work is to detect objects [20] using HOG features. Necessary steps [20] to calculate HOG features from an image are mentioned below:

- Divide the whole image to create small adjacent cells
- Calculate gradient magnitude, direction for each pixel in each cell
- Calculate the corresponding bin for each gradient magnitude, direction to represent is using a histogram of gradients
- Create blocks from adjacent cells, perform block normalization and calculate feature vector from it

Using proper cell size, block size, number of bins are problem specific task. During the implementation, 9 bins in the histogram, cells of 8×8 pixels and blocks of 2×2 pixels and unsigned orientation of gradients were used. Fig. 5 represents the HOG feature extraction step on the nose of a sample image from RaFD dataset [18].

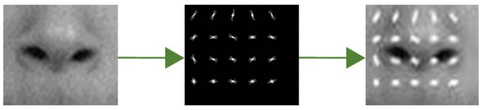

Fig. 5. HOG feature extraction on the nose of a sample image.

B. Local Binary Patterns (LBP)

LBP features are robust against illumination changes. They are comparatively faster to compute as well. Facial analysis, motion analysis, texture analysis, image analysis are some areas where LBP features have been used successfully. A widely accepted work with LBP to analyze texture is shown in [21]. Required steps to calculate LBP features from an image are mentioned below:

- Create small cells with a specified radius, number of neighbors from the whole image
- Perform thresholding considering the center pixel and its neighboring pixels
- A binary number (LBP) is found as the result of thresholding, later it is converted to a decimal number (LBP)
- Store the ‘count’ of each LBP
- Calculate the histogram over the cell, considering the frequency of each ‘count’ occurring
- Concatenate histograms of all the cells to compute the feature vector

Using the appropriate value of radius, number of neighbors are problem dependent. While implementing, a radius of 1 with 8 neighbors surrounding the center pixel was considered.

C. Fusion

From the four segmented facial parts at first HOG features are extracted and then from the same four segmented parts LBP features are extracted. Then features calculated from the same parts using HOG and LBP are concatenated to form the final feature vector. Final feature vector had in total 1892 features and among them, 1656 features came from HOG descriptor, while rest came from LBP.

VI. Dimension Reduction

For reducing the dimension of the feature vector, PCA [22] is used for its ability to represent data in terms of principal components. Steps [22] of PCA are as follows:

- Normalize the data, calculate the covariance matrix
- Calculate eigenvectors, eigenvalues of the covariance matrix
- Analyze principal components, translate the data in terms of the components

While implementing, PCA was used retaining 99% variance ratio which resulted in 89.85% dimension reduction. 1892 features were reduced to 192 features by PCA. The dimension of four segmented parts are different and so is the length of their corresponding feature vector. But for any particular part, the dimension of the feature vector is same for all sample images. So when the features are given to the classifier, same parts have feature vector of the same length and as a result, there was no mismatch of dimensions.

VII. Classification

To classify the expressions through their corresponding features, ANN is used by the proposed method. Multilayer perceptrons with backpropagation of information are used for training the system. ANN can model real-world complex problems and can imitate the learning process of a human. Better generalization, graceful degradation, fault tolerance are among the other properties that made choose ANN as the classifier. Limited memory BFGS (L-BFGS) method is used for optimizing weights. At time t, given a set of inputs $d^{(t)}$, activation function outputs $g^{(t)}$, weights $w^{(t)}$ and tolerance ϵ, then the algorithm [23-24] for multilayer perceptrons with L-BFGS weight optimization is mentioned below:

1. Initialize all weights $w^{(0)}$ randomly with small values, set the value of ϵ, set $t = t + 1$
2. Present all training sets to the network and calculate error function $E(w^{(t)})$ for current weights
3. If $t = 0$, then $d^{(t)} = -\nabla E^{(t)}$
   Else $d^{(t)} = -\nabla E^{(t-1)} + Ap + Bv$
   where $p = w^{(t+1)} - w^{(t)}$, $v = g^{(t+1)} - g^{(t)}$
4. Calculate $w^{(t+1)} = w^{(t)} - \alpha d^{(t)}$ using $\alpha^{(t)} = -\frac{{d^{(t)}}^T g^{(t)}}{{d^{(t)}}^T H d^{(t)}}$
5. Calculate A, B for the next iterations as
   $A = -\left(1 + \frac{v^T v}{p^T v}\right)\frac{p^T g^{(t+1)}}{p^T v} + \frac{v^T g^{(t+1)}}{p^T v}$, $B = \frac{p^T g^{(t+1)}}{p^T v}$
6. If $E\left(w^{(t+1)}\right) > \varepsilon$, then $t = t + 1$ and go to 2, else stop.

As the stopping criteria ε=1e-8, logistic sigmoid function as the activation function, two hidden layers with 50 nodes in each of it were used during the implementation. The number of nodes in each layer was found by trial and error. The number of hidden layers was chosen according to Kolmogorov’s theorem [25].

VIII. Result

A computer with 64-bit operating system, 4GB memory and core-i5 processor was used to implement the proposed FER system. JAFFE [26], CK+ [27] and RaFD [18] datasets were used for evaluating the performance of the proposed method. All 213 expression images from JAFFE dataset, 1219 expression images from 22 people from CK+ dataset, 1407 front facing expression images from 67 people from RaFD dataset were used. Seven basic expressions were considered. Contemptuous expression images from RaFD dataset were excluded from the experiment as the other two datasets did not have images of that expression. K-fold cross-validation was used to get unbiased results. Splitting of training and testing set was done randomly. TABLE II represents the average (Avg.) accuracy achieved on different datasets by the proposed method. Here accuracy means Correct Recognition Rate (CRR). Feature length (F.L.(no D.R.)) without any dimension reduction is also mentioned in the table. Traditional method refers to extracting features from the full face without considering any regions of interest. From the table, it is apparent that the performance of the proposed method is almost identical to traditional method but there is a huge difference in the length of the feature vector. Yes, the dimension of the feature vector can be reduced by dimension reduction techniques. But reducing 10463 features to a few and reducing 1892 features to a few are not computationally equivalent. The former would be more time consuming and be a barrier to creating a real-time system.

TABLE II. ACCURACY OF DIFFERENT DATASETS

Dataset Used	No. of Folds	Proposed Method		Traditional Method	
		Avg.	*F.L.(no D.R.)*	*Avg.*	*F.L.(no D.R.)*
JAFFE	5	90.60	1892	92.98	10463
JAFFE	10	93.51		94.33	
CK+	5	99.67		99.75	
CK+	10	99.67		99.84	
RaFD	5	99.01		99.50	
RaFD	10	99.29		99.50	

Performance comparison of the three used datasets is shown in Fig. 6. As mentioned in TABLE II, JAFFE is the dataset with the lowest correct recognition rate and can be seen from Fig. 6 as well.

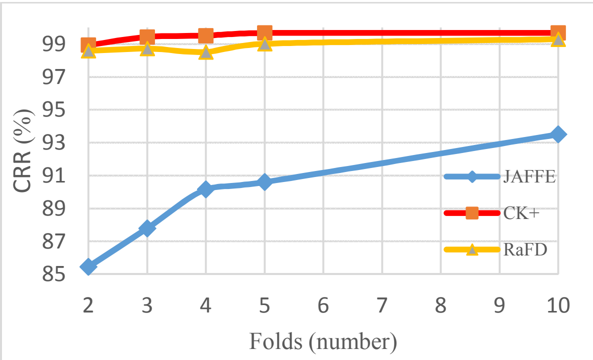

Fig. 6. Performance comparison on JAFFE, CK+ and RaFD dataset.

Along with accuracy, confusion matrices are also analyzed for FER problems. TABLE III, TABLE IV and TABLE IV represent confusion matrices on JAFFE, CK+ and RaFD datasets respectively on random test cases. Ne, Ha, An, Su, Fe, Di, Sa represents neutral, happy, angry, surprise, fear, disgust, sad facial expressions respectively and Acc. Represents the accuracy in these confusion matrices. For JAFFE dataset happy and surprise expression images had some misclassifications for that particular test case. Disgust and sad expression images had the lowest CRR for CK+ and RaFD datasets respectively on random test cases.

TABLE III. CONFUSION MATRIX OF CRR ON JAFFE DATASET

	Ne	Ha	An	Su	Fe	Di	Sa	Acc.
Ne	100	0	0	0	0	0	0	100
Ha	11.11	88.89	0	0	0	0	0	88.89
An	0	0	100	0	0	0	0	100
Su	25.00	0	0	75.00	0	0	0	75.00
Fe	0	0	0	0	100	0	0	100
Di	0	0	0	0	0	100	0	100
Sa	0	0	0	0	0	0	100	100

TABLE IV. CONFUSION MATRIX OF CRR ON CK+ DATASET

	Ne	Ha	An	Su	Fe	Di	Sa	Acc.
Ne	100	0	0	0	0	0	0	100
Ha	0	100	0	0	0	0	0	100
An	0	0	100	0	0	0	0	100
Su	0	0	0	100	0	0	0	100
Fe	0	0	0	0	100	0	0	100
Di	3.03	0	0	0	0	96.97	0	96.97
Sa	0	0	0	0	0	0	100	100

TABLE V. CONFUSION MATRIX OF CRR ON RAFD DATASET

	Ne	Ha	An	Su	Fe	Di	Sa	Acc.
Ne	100	0	0	0	0	0	0	100
Ha	0	100	0	0	0	0	0	100
An	0	0	100	0	0	0	0	100
Su	0	0	0	100	0	0	0	100
Fe	0	0	0	0	100	0	0	100
Di	0	0	0	0	0	100	0	100
Sa	0	0	2.44	0	0	0	97.56	97.56

IX. STATE-OF-THE-ART

TABLE VI. STATE-OF-THE-ART COMPARISON

Study	Technique	Dataset	Acc.
Proposed Method	Viola-Jones face detection, **facial region segmentation**, HOG+LBP, PCA, ANN, MLP, backpropagation	JAFFE CK+ RaFD	**93.51** **99.67** **99.29**
2017 [14]	Viola-Jones face detection, Gabor filter, ANN	JAFFE	85.70
2017 [15]	Appearance-based models, shape signature, grid, MLP	JAFFE CK+	92.00 97.90
2015 [16]	Gabor filter, Radial Bases Functions Networks (RBFN), ANN	JAFFE CK+	92.88 92.40
2015 [28]	SURF, gentle AdaBoost	RaFD	90.64
2017 [29]	Visual saliency,deep learning	RaFD	95.70

X. CONCLUSION

Other than the facial region segmentation method, the proposed FER system just used a combination of existing methods such as Viola-Jones face detection, HOG, LBP, PCA, ANN. But the existing methods were combined in such a way that the combination worked effectively. Initially, the proposed image segmentation method might look ineffective. But the system was tested with almost 3000 front facing images which included images of people from different race, ethnic groups, geographical location. Images with unusual facial structure, unusual position of four parts might not get segmented properly by the proposed segmentation method. Some expressions were even tough to recognize for humans, let alone machines. Such expressions affected the overall performance of the system.

The system is capable of handling only front-facing images due to its image segmentation method. But it is not a desirable property in creating a robust system. Developing a robust system capable of handling images rotated at any angle in any direction would be a future work. The ultimate goal was to recognize emotions and emotions are not only dependent on facial information. So developing a system to recognize emotion combining both facial and vocal information would be another interesting work in near future.

REFERENCES

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

[2] G. Sandbach, S. Zafeiriou, M. Pantic and L. Yin, “Static and Dynamic 3D Facial Expression Recognition, A Comprehensive Survey”, *Image Vision Compute*, Vol. 30, No. 10, pp. 683-697, 2012.

[3] 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, pp. 1-4, 2015.

[4] F. Mahmud, M. E. Haque, S. T. Zuhori and B. Pal, "Human face recognition using PCA based Genetic Algorithm", *2014 International Conference on Electrical Engineering and Information & Communication Technology*, Dhaka, pp. 1-5, 2014.

[5] 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, pp. 582-587, 2015.

[6] A. Matin, F. Mahmud and M. T. B. Shawkat, "Recognition of an individual using the unique features of human face", *2016 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE)*, Pune, India, pp. 57-60, 2016.

[7] 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, pp. 1-4, 2017.

[8] 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, pp. 1-4, 2016.

[9] S. An and Q. Ruan, "3D facial expression recognition algorithm using local threshold binary pattern and histogram of oriented gradient," *2016 IEEE 13th International Conference on Signal Processing (ICSP)*, Chengdu, pp. 265-270, 2016.

[10] J. Jayalekshmi and T. Mathew, "Facial expression recognition and emotion classification system for sentiment analysis," *2017 International Conference on Networks & Advances in Computational Technologies (NetACT)*, Thiruvanthapuram, pp. 1-8, 2017.

[11] S. Agrawal and S. Yadav. "Approach Based on HOG Descriptor and K nearest Neighbour for facial Expressions Recognition", *International Journal of Innovations & Advancement in Computer Science*, Vol. 7, No. 5, pp.239-249, 2018.

[12] R. Verma and M. Y. Dabbagh, "Fast facial expression recognition based on local binary patterns," *2013 26th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, Regina, SK, pp. 1-4, 2013.

[13] Z. T. Liu, G. T. Sui, D. Y. Li and G. Z. Tan, "A novel facial expression recognition method based on extreme learning machine," *2015 34th Chinese Control Conference (CCC)*, Hangzhou, pp. 3852-3857, 2015.

[14] K. Verma and A. Khunteta, "Facial expression recognition using Gabor filter and multi-layer artificial neural network," *2017 International Conference on Information, Communication, Instrumentation and Control (ICICIC)*, Indore, pp. 1-5, 2017.

[15] A. Barman and P. Dutta, "Facial expression recognition using shape signature feature," *2017 Third International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*, Kolkata, pp. 174-179, 2017.

[16] R. Saabni, "Facial expression recognition using multi Radial Bases Function Networks and 2-D Gabor filters," *2015 Fifth International Conference on Digital Information Processing and Communications (ICDIPC)*, Sierre, pp. 225-230, 2015.

[17] 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.

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

[19] 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.

[20] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, San Diego, CA, USA, pp. 886-893, Vol. 1, 2005.

[21] T. Ojala, M. Pietikainen and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 7, pp. 971-987, Jul 2002.

[22] L. I. Smith, "A tutorial on principal components analysis," Cornell University, USA, 2002.

[23] M. Brescia, *Multi Layer Perceptron Trained by Quasi Newton Learning Rule*, Naples, 2012 [Online]. Available: http://dame.dsf.unina.it/documents/DAME_MLPQNA_Model_Mathematics.pdf. [Accessed: 07- Jun- 2018].

[24] Brescia et al., *Publications of the Astronomical Society of the Pacific*, Vol. 126, No. 942, pp. 783-797, 2014.

[25] V. Kůrková, "Kolmogorov's theorem and multilayer neural networks", *Neural Networks*, Vol. 5, No. 3, pp. 501-506, 1992.

[26] M. J. Lyons, S. Akemastu, M. Kamachi and J. Gyoba, "Coding Facial Expressions with Gabor Wavelets", *3rd IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 200-205, 1998.

[27] 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.

[28] 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, pp. 630-635, 2015.

[29] V. Mavani, S. Raman and K. P. Miyapuram, "Facial Expression Recognition Using Visual Saliency and Deep Learning," *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, Venice, pp. 2783-2788, 2017.