Facial Expression Recognition System Using Extreme Learning Machine

Firoz Mahmud, Dr. Md. Al Mamun

Abstract— Interest is growing in improving all aspect of the interaction between human and computer including human emotions. It is a crucial task for a computer to understand human emotions. A very meaningful way of expressing human emotions is facial expression. In this paper, a model facial expression recognition based on Extreme Learning Machine is proposed. Salient facial feature segments like eyebrows, eyes, mouth, and nose are detected from a face image and then these feature segments are extracted by using morphological image processing operation and edge detection technique to form feature vectors. Extreme Learning Machine, a feed-forward neural network classifier with a single layer of hidden nodes is used for recognizing expressions of the input faces into six basic categories like happy, sad, surprise, angry, disgust, and fear. The experiments of facial expression recognition system are carried out on JAFFE facial expression database and performances of experimental results are analysed.

Index Terms— Facial Expression Recognition, Extreme Learning Machine, Facial Feature Segments, Morphological Image Processing, Edge Detection, Feature Vector, Feed-forward Neural Network.

1. Introduction

INTERACTION between human and computer is increasing day by day. The most challenging task for a computer is to understand the human behaviors and emotions. The most expressive way to express human emotions is facial expressions. Nowadays facial expressions recognition has become a great concern in modern scientific research. Facial expression is a non-verbal way of communication. Physical movement of muscles beneath the skin of your face refers to facial expressions. These movements convey the emotional states of an individual to observers. The expressions are fully innate and sometimes people do not know that they are showing their expressions. Facial expression recognition has been spread out in various fields in our day to day life including behavioral science, clinical practice, business negotiation, safe driving, remote education, intelligent family robot, virtual games, effective biometric identification, video surveillance camera, human computer interaction (HCI), and analysis of human emotions.

Facial expressions recognition technique can be stepped out into three parts, acquisition of set of images of face, extracting facial features, and classification of facial features as a particular facial expression. There are six prototypic facial expressions, happy, anger, sadness, fear, surprise, and disgust [1].

Computer scientists and researchers have great interest in the field of computer vision. As a consequence, so many classification techniques have already been proposed and implemented. Among these classification techniques some of them are included here. Cohen et al. [2] proposed a method for recognizing facial expressions from a video sequence by using Tree-Augmented-Naïve Bayes (TAN) as classifier. Ma & Khorasani [3] proposed a method which was combination of two-dimensional discrete cosine transform (2D-DCT) and constructive one-hidden layer feed forward neural network. Buciu et al. [4] used cosine similarity measure (CSM) or SVM as classifier. Fuzzy support vector machine (FSVM) with KNN was proposed to classify multiclass problems and reducing computational complexity [5]. Discrete wavelet transformation (DWT) was proposed to locate the facial features automatically and describe deformations between neutral and non-neutral expressions by dividing image into various frequency bands [6]. Another approach used Gabor Wavelets and its fusion with local binary pattern for feature extraction, then reduced the dimension of feature vectors by principle component analysis (PCA) and finally classified the facial images using nearest neighbor (NN), multilayer feed forward neural network (MFNN) and extreme learning machine (ELM) [7]. Khandait et al. [8] proposed morphological image processing operations for face portion segmentation and localization from a frontal face image and feed forward back propagation neural network for classifying the facial expressions. Huang et al. [9] proposed Extreme Learning Machine, which in an improved feed-forward neural network that can randomly generate weights and thresholds, and only the number of neurons in the hidden layer are arranged. It doesn't need to set the other parameters, and it can obtain optimal solution after learning. Jia, Bo, et al. [10] describe a technique to make matrices directly as input 2DELM can be used instead of commonly used vector. 2DELM take the matrices as input features without vectorization [10]. Uçar et al. [11] was proposed a novel algorithm for facial expression recognition by integrating curvelet transforma and online sequential extreme learning machine (OSELM) with radial basis function (RBF) hidden node, where spherical clustering was employed to the features set to

- *Firoz Mahmud is an Assistant Professor in the department of CSE, Rajshahi University of Engineering & Technology, Bangladesh. His research interests are Computer Vision, Image Processing, Robotics, Optimization, and Cloud Computing. PH-+8801912851490. E-mail: fmahmud.ruet@gmail.com*
- *Dr. Md. Al Mamun is an Associate Professor in the department of CSE, Rajshahi University of Engineering & Technology, Bangladesh. His research interests are Computer Vision, Satellite Image Mining. PH-+8801712622764. E-mail: cse_mamun@yahoo.com*

automatically determine optimal hidden node number and radial basis function (RBF) hidden node parameters of OSELM. Claude et al. [12] presented a high-level overview of automatic expression recognition by highlighting the main system components and some research challenges. Le et al. [13] an approach based on ELM is proposed which can detect the face images that contain a smile or non-smile facial expression. First of all, the faces from original images are detected using Viola Jones face detector and then the detected faces are registered using a fully automated flow-based registration system. Facial features are extracted from these registered faces using Local Binary pattern (LBP), Local Phase Quantization (LPQ), and Histogram of Oriented Gradient (HOG) descriptors. Finally ELM is used to the classifier. The proposed system reveals that, ELM has less computational cost, high accuracy, efficiency, and better generalization performance the other benchmarking classifier such as Support Vector Machine (SVM) and Linear Discriminant Analysis (LDA). Dewan et al. [14] survey paper discusses the importance of facial expression in terms of communication and provides a system for facial expression recognition that can be divided into three modules- i. Face detection and tracking, ii. Feature extraction, iii. Expression classification.

The author describes some methods for each module briefly. KanadeLucas-Tomasi tracker was one of the most popular methods for face detection and tracking but its false detection rate is pretty high. The method developed by Viola and Jones for frontal face detection is very fast and accurate. CANDIDE is a face mask that can be used for coding human faces which is based on global and local action units.

For feature extraction, Tian et al. used both permanent and transient features for automatic face analysis. Cohen et a.l used local deformation of facial features unit for feature extraction. Bartlett et al. used AdaBoost with Gabor filter for feature extraction but Gabor wavelet representations of images are both time and memory intensive. Mahmud et al. [19-21] sued PCA for feature extraction.

For expression classification, Cohen et al. noted that the performance of tree augmented naive bayes classifier is better for static image analysis and multilevel Hidden Markov Model is better for dynamic classifier. Bourel et al. presented a data fusion approach which can recognize facial expression in presence of occlusion.

To greatly reduce the large amount of data processing time and gain better efficiency for facial expression recognition, an approach of facial expression recognition based on Extreme Learning Machine is proposed. The procedure contains three sub-divisions. First of all, facial features are detected and segmented from a set of still frontal posed images. Secondly, feature segments are extracted by using some image processing operations and feature vectors are formed. Finally, ELM is used as classifier to classify the facial expressions using the extracted feature vectors. This approach is evaluated by the experiment on a well-known Japanese Female Facial Expression (JAFFE) dataset [15].

The rest of the paper is organized as follows. The model facial expression recognition is described in section 2. The facial features detection and segmentation are briefly discussed in section 3. Facial features extraction and formation of feature vectors are presented in section 4. A brief discussion of ELM algorithm is presented in section 5. Experimental results analysis is given in section 6.

2. The Model of Facial Expression Recognition

Facial expression recognition based on extreme learning machine is proposed, which includes detection and segmentation of facial features, feature extraction and formation of feature vector and finally expression recognition using ELM classifier. At first, facial features are detected using Viola-Jones object detection framework and these features are segmented by cropping these features. After that, applying some morphological image processing operations and edge detection techniques, feature segments are extracted and feature vectors are formed. Finally, these feature vectors are used in ELM classifier to classify the expressions.

The model of facial expression recognition is shown in fig.1. Images are obtained from JAFFE standard image database. Facial features detection and segmentation are performed in pre-processing steps. Then facial features are extracted and feature vectors are formed from pre-processed feature segments. These feature vectors are then used for training in ELM neural network. After completion of training, some of these feature vectors are tested against the model and finally facial expressions are classified by ELM classifier.

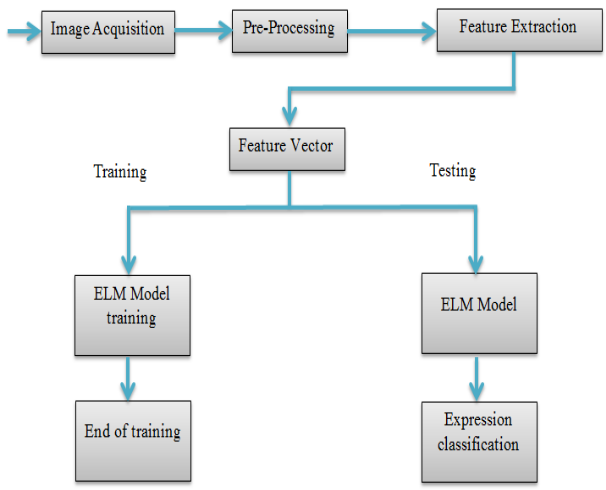
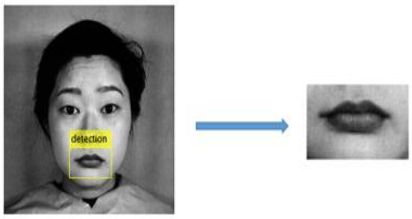


Fig.1 The model of facial expression recognition

3. Facial Features Detection and Segmentation

The role face area and facial features is very important in facial expression recognition. Salient facial features such as left eye, right eye, left eye brow, right eye brow, nose, and mouth are detected using Viola-Jones object detection framework from a face image. This object detection framework has to tell whether of an image in defined size contain required shape or object or not, if contain where it is [17]. In this work we detect and segment 6 different regions of a face. Different facial features require different thresholding value for detection. So a variable thresholding value is used for facial features detection. Then the detected features are segmented by cropping properly. An example of facial feature detection and segmentation is shown in figure 2.



a. Detection of facial feature (mouth) b. Segmentation of facial feature (mouth)

c. Segmentation of facial feature (nose) d. Segmentation of facial feature (eye)

Fig.2 Detection and Segmentation of facial features

Detection and segmentation of nose, left eye, right eye, left eye brow, and right eye brow are done by the same procedure as the mouth.

4. Facial Features Extraction and Formation of Feature Vectors

For extracting facial features, some image processing operations are performed on the segmented facial features. At first, Otsu's optimum global thresholding method is used for obtaining binary image. Black portion of this binary image is the feature segment and white portion of this image is the background. Then image complementation operation is performed on that binary image for applying morphological image processing operation. For removing unwanted noises and smoothing the boundary of the feature segment, erosion morphological operation within an appropriate structuring element (diamond) is performed.

(a) (b) (c)

(d) (e)

Fig.3 (a) Mouth segment, (b) Binary image of mouth segment, (c) Complemented image of binary mouth segment, (d) Eroded mouth segment, and (e) Edge of mouth segment.

Then Laplacian of Gaussian (LoG) method within appropriate threshold value is applied on that feature segments for detecting the edges. An example of feature extraction process is shown in figure 3.

After that, the height and width are calculated from the detected edge segment. For extracting features from every feature segments of images, the same procedure was applied. Figure 4 is shown the width and height of mouth. Same procedure is followed for calculation the height and width for left eye, right eye, nose, left eye brow and right eye brow.

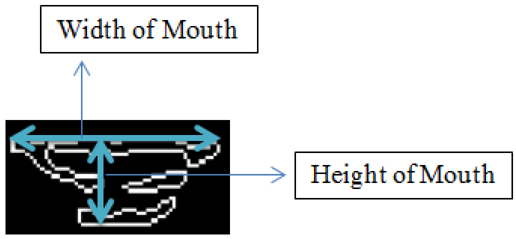


Fig.4 Calculating the height and width of facial features (mouth)

After extracting features from the images, feature vectors are formed. Each feature vector contains 12 features. These are:

$$f_v = \{h_1, w_1, h_2, w_2, h_3, w_3, h_4, w_4, d_1, d_2, w_5, w_6\} \quad (1)$$

Where, h_1 = height of mouth, w_1 = width of mouth, h_2 = height of nose, w_2= width of nose, h_3= height of left eye, w_3= width of left eye, h_4= height of right eye, w_4= width of right eye, d_1= distance between left eye and eyebrow, d_2= distance between right eye and eyebrow, w_5= width of left eye, w_6= width of right eye.

5. Extreme Learning Machine

Extreme Learning Machine (ELM) is a feed-forward neural network for classification with a single layer of hidden nodes where the weights connecting inputs to hidden nodes are randomly assigned and never updated. These weights between hidden nodes and outputs are learned in a single step. Unlike traditional neural networks, there is no need of tuning for weight adjustment. That's why ELM is extremely fast and greatly reduces the data processing time. The training time of ELM is ten times faster than traditional neural networks [18]. ELM contains fewer parameters than other single hidden layer feed-forward neural networks [16].The output of ELM is

$$f(x) = \sum_{i=1}^{L} \beta_i G(a_i, b_i, x) = \beta \cdot h(x) \quad (2)$$

Where β_i is the output weight from i-th hidden node to the output node. $G(a_i, b_i, x)$ is the output of i-th hidden node. $h(x) = [\, G(a_1, b_1, x), \dots, G(a_L, b_L, x)]^{\mathrm{T}}$ is the output vector of hidden layer with respect to the input **x**. For binary classification, the decision function of ELM is

$$f(x) = sign(\sum_{i=1}^{L} G(a_i, b_i, x)) = sign(\beta \cdot h(x)) \quad (3)$$

The ELM aims to minimize the training error and the norm of the output weights.

$$\text{Minimize} \sum_{i=1}^{L} \| \beta \cdot h(x_i) - t_i \|, \text{ minimize } \| \beta \| \quad (4)$$

where t_i the desired output.

In summary, ELM algorithm [11]**:** Given a training set $\{(x_i, t_i) \mid x_i \in R^m, i = 1, 2, \ldots, N\}$, hidden node output function $g(w, b, x)$, and number of hidden nodes L.

1) Assign randomly hidden node parameters (w_i, b_i), $i = 1,2, \ldots L$.
2) Calculate the hidden layer output matrix $\boldsymbol{H}$.
3) Calculate the output weights β: $\beta = \boldsymbol{H}^{+}\boldsymbol{T}$.

Where, $\mathbf{H}^{+}$ is the Moore-Penrose generalized inverse of hidden layer output matrix **H**.

6. Experiments and Results Analysis

The proposed model is evaluated in well-known JAFFE dataset. As the feature vectors contain 12 values, number of nodes in input layer is 12. With the proper approximation of number of hidden nodes in hidden layers, these values are given as input in input layer and number of nodes in output layer is six as the proposed model deals with six types of expressions. System performance is analysed when the system is trained properly. A set of 42 images consisting of 6 basic types of facial expressions and one neutral image is used for training. Then several images are tested. Extreme learning machine algorithm is extremely fast. It requires .0936 seconds to train 42 images in the system whereas gradient based back-propagation neural network (BPN) requires 1 second to train same number of images in the same environment. Table.I shows the training time required in ELM and BPN for same number of training samples.

After proper training, images are tested against the system for classifying the facial expressions. Table.II is a confusion matrix that shows the various expressions recognition accuracy of images tested against the system. For example, when happy faces are tested, 96% of the faces are correctly recognized as happy and the proposed method is confused for 4% of the faces and shows the incorrect expression. Among 4% of the incorrect expression 2% matches with surprise and rest of the 2% matches with fear expression. Again, when surprize faces are tested, 87.50% faces are correctly recognised as surprize and the model is confused for rest of the percentage and shows the incorrect results. The best performance in terms of accuracy is 96%, which is happy expression and the lowest accuracy is 82%, which is fear expression. All expression of training set successfully recognized, but not all expressions of testing set were equally well recognized by the system. The confusion matrixes of Table II and Table IV show that some of the happy faces were misclassified as others, which is 4%. For the sad expression, 5.63% were misclassified as others. The rate was 12.5% for surprise expression, 10.86% for angry expression, 7.98% for disgust expression, and 18% for fear expression. The overall misclassification rate is not so bad.

TABLE I: TRAINING TIME OF ELM AND BPN.

No. of samples	Algorithm	Training time (seconds)
42	ELM	0.0936
	BPN	1

Table III shows the expression recognition accuracy of ELM and BPN based on the proposed method. The proposed method achieved overall 90.17% expression recognition accuracy using ELM as classifier whereas it achieved 86.85% expression recognition accuracy using BPN as classifier in the same environment.

TABLE III: RCOGNITION ACCURACY OF ELM AND BPN

Algorithm	Recognition accuracy (%)
ELM	90.17
BPN	86.85

TABLE IV: FALSE POSITIVE RATE FOR EACH EXPRESSION

Facial Expression	False Positive Rate (%)
Happy	4
Sad	5.66
Surprise	12.5
Angry	10.86
Disgust	7.98
Fear	18
Total	9.83

Table IV shows the summary of the false positive rate for each expression. The whole experiment is performed by a machine 32 bit, Intel Corei3 CPU 2.3 GHz processor, 2GB RAM, and Matlab-2014a is used as simulator.

TABLE II: THE CONFUSION MATRIX FOR PROPOSED METHOD ON JAFFE DATABASE.

	Happy	Sad	Surprise	Angry	Disgust	Fear
Happy	**96%**		02%			02%
Sad		**94.34%**	04.03%		01.63%	
Surprise	09.5%	03%	**87.50%**			
Angry		06.57%		**89.14%**		04.29%
Disgust		04.08%		03.90%	**92.02%**	
Fear	01.5%		11%		05.5%	**82%**

7. Conclusion

An approach of facial expression recognition using ELM is proposed. Salient facial feature segments are detected by using Viola-Jones objection detection algorithm. These feature segments are then converted into binary images from gray-level images by using Otsu's Optimum Global Thresholding algorithm. Then morphological operation is applied on the binary features segments to omit noises and make the edges of feature segments smooth. Laplacian of Gaussian filter (LoG) is used to detect edges from the morphologically operated image and then feature vectors are created. Extreme Learning Machine algorithm is used as classifier to recognize the six basic types of expressions. The performance of the proposed model is revealed by the experimental results. Training time of ELM is much less than traditional gradient based neural network and overall recognition accuracy of ELM is satisfactory.

Although training time of the system is fast enough, but the overall accuracy of the system is less attractive. Because effectiveness of ELM depends on feature extraction and number of its basic classifier. The difficulties of facial features extraction and confusion of subtle expression change results the recognition accuracy lower. It will be difficult for the proposed system to recognize the facial expression if the face angle of the image or head position is changed. Another limitation of the proposed model is, it will be unable to correctly recognize the facial expressions if there is a of presence hairs in the face area and the illumination of the face images are changed.

The future attempt is to extract more facial features appropriately to form the feature vectors and try to improve the accuracy. Further task is to implement the proposed method on some other standard face databases for analysing the performance efficiently.

References

[1] P. Ekman, "Emotion in the Human Face", Cambridge University Press, 1982.
[2] I. Cohen, N. Sebe, A. Garg, S. C. Lawrence, S. H. Thomas, "Facial expression recognition from video sequences: temporal and static modeling", Computer Vision and Image Understanding, Special Issue on Face Recognition, vol. 91, pp. 160-187, Issues 1-2, July-August 2003.
[3] L. Ma, K. Khorasani, "Facial expression recognition using constructive feedforward neural networks," Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on , vol. 34, no. 3, pp.1588-1595, June 2004.
[4] I. Buciu, C. Kotropoulos, and I. Pitas, "Comparison of ICA approaches for facial expression recognition," Signal, Image and Video Processing, vol. 3, pp. 345–361, 2009.
[5] X.H. Wang, A. Liu & S.Q. Zhang, "New facial expression recognition based on FSVM and KNN," Optik-Int. J. for Light Electron Optics, vol. 21, no. 126, pp. 3132–3134, 2015.
[6] W. Wei, C. Tian, S. J. Maybank,, & Y. Zhang, " Facial expression transfer method based on frequency analysis," Pattern Recognition, vol. 49, pp.115-128, 2016.
[7] S. Nikan & M. Ahmadi, "Effectiveness of various classification techniques on human face recognition," High Performance Comput. & Simulation (hpcs), 2014 Int. Conf. , pp. 651–655, 2014.
[8] S. P. Khandait, R. C. Thool, & P. D. Khandait. "Automatic facial feature extraction and expression recognition based on neural network." arXiv preprint arXiv:1204.2073, 2012.
[9] G.-B. Huang, Q.-Y. Zhu, & C.-K. Siew. "Extreme learning machine: a new learning scheme of feedforward neural networks." Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on, Vol. 2, pp. 985-990, 2004.
[10 Jia, Bo, et al. "Two-Dimensional Extreme Learning Machine." Mathematical Problems in Engineering 2015 (2015).
[11] Uçar, Ayşegül, Yakup Demir, and Cüneyt Güzeliş. "A new facial expression recognition based on curvelet transform and online sequential extreme learning machine initialized with spherical clustering." Neural Computing and Applications 27.1 (2016): 131-142.
[12] Chibelushi, Claude C., and Fabrice Bourel. "Facial expression recognition: A brief tutorial overview." CVonline: On-Line Compendium of Computer Vision 9 (2003).
[13] Le An , Songfan Yang , Bir Bhanu, Efficient smile detection by Extreme Learning Machine, Neurocomputing, v.149 n.PA, p.354-363, February 2015 [doi>10.1016/j.neucom.2014.04.072].
[14] Dewan Ibtesham, "A Survey on Facial Expression Recognition" 2004.
[15] M. Lyons, S. Akamatsu, M. Kamachi, & J. Gyoba, "Coding facial expressions with gabor wavelets," Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on. IEEE, pp. 200-205, 1998.
[16] Z.-T. Liu, G.-T. Sui, D.-Y. Li & G.-H. Tan, "A novel facial expression recognition method based on extreme learning machine," Control Conference (CCC), 2015 34th Chinese. IEEE, pp. 3852-3857, 2015.
[17] Yi-Qing Wang, "An Analysis of the Viola-Jones Face Detection Algorithm", Image Processing On Line, 4 (2014), pp. 128–148. http://dx.doi.org/10.5201/ipol.2014.104.
[18] Mahmud, F.; Afroge, S.; Mamun, M.A.; Matin, A., "PCA and Back-Propagation Neural Based Face Recognition System," inComputer and Information Technology (ICCIT), 2015 18th International Conference on , vol., no., pp.582-587, 21-23 Dec. 2015, doi:10.1109/ICCITechn.2015.748813.
[19] Matin, A; Mahmud, F.; Mosst Tasnim Binte Shawkat, M.T.B, " Recognition of an Individual using the Unique Features of Human face 2016 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE) 19-21 December 2016, AISSMS, Pune, India.
[20] Mahmud, F.; Haque, M.E.; Zuhori, S.T.; Pal, B., "Human face recognition using PCA based Genetic Algorithm," Electrical Engineering and Information & Communication Technology (ICEEICT), 2014 International Conference on , vol., no., pp.1,5, 10-12 April 2014, doi: 10.1109/ICEEICT.2014.6919046.
[21] Mahmud, F.; Khatun, M.T.; Zuhori, S.T.; Afroge, S.; Aktar, M.; Pal, B., "Face recognition using Principle Component Analysis and Linear Discriminant Analysis," in Electrical Engineering and Information Communication Technology (ICEEICT), 2015 International Conference on , vol., no., pp.1-4, 21-23 May 2015, doi: 10.1109/ICEEICT.2015.7307518.