Original Research Paper

Fusion of Features and Extreme Learning Machine for Facial Expression Recognition

Bayezid Islam and Arfat Hossain

Department of Computer Science and Engineering,
Rajshahi University of Engineering and Technology, Rajshahi, Bangladesh

Abstract: Human emotion is highly correlated to facial expressions. Due to its growing demand in different sectors, an emotion recognition method is proposed through recognizing facial expressions. The input image is preprocessed and then the resulting image is segmented into four facial expression regions following the newly proposed segmentation method. Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP) are fused to extract the necessary features from the four segmented parts. The dimension of the feature vector is reduced using Principal Component Analysis (PCA). To classify the expressions, Extreme Learning Machine (ELM) is used. For evaluating the performance of the proposed method, three widely used and publicly available facial expression datasets (JAFFE, CK+, RaFD) are used. The proposed method achieved 95.3%, 99.84% and 98.65% accuracy while using images from JAFFE, CK+ and RaFD dataset respectively. Performance of the proposed method on these datasets is compared to other facial expression recognition methods on these datasets to indicate that the proposed method achieves state-of-the-art performance.

Keywords: Facial Region Segmentation, Fusion of HOG and LBP, Principal Component Analysis (PCA), Extreme Learning Machine (ELM)

Introduction

Ability to detect the mental condition, feelings of a person can be of great importance. Facial expressions play a vital role in both nonverbal and social communication. According to Mehrabian (1968), facial expressions contribute the most (55%) in reflecting the feelings of a speaker. So successful recognition of human emotion, feelings is highly dependent on the successful recognition of facial expressions. Recognized facial expressions can be used in different sectors of our lives for improving the everyday experience. It can be used for security purposes as well.

Whether a person is mentally fit or not for a sensitive task, can be determined by recognizing the emotion of that person in advance. Social networks can incorporate the emotion recognition feature and suggest users post on their timeline depending on the expression of their uploaded photo. Robots can be given this feature to extract the best out of them. As the world is heading towards automation, so the ultimate target is to recognize facial expressions by machines as humans can do it spontaneously in real-time.

Seven basic facial expressions are usually considered while dealing with Facial Expression Recognition (FER) problems. They are neutral, fear, disgust, sad, happy, angry and surprise (Sandbach *et al*., 2012). Some other expressions such as contemptuous, pain are considered in some works. Some works are done excluding neutral expressions. As it has many applications in many sectors, researchers have been trying to develop FER systems with the ability to recognize expressions accurately in the least possible time. So, expression recognition has been a prime research topic in the field of computer vision, image processing, human-computer interaction for the last few decades.

Usually, FER systems include image preprocessing, feature extraction and classification step. Different available methods use different techniques to perform these steps. Gabor wavelets (Islam *et al*., 2018a), Principal Component Analysis (PCA) (Mahmud *et al*., 2015), Linear Discriminant Analysis (LDA) (Mahmud *et al*., 2015), Scale Invariant Feature Transform (SIFT), Speeded-Up Robust Features (SURF), moments, Gray level co-occurrence matrix are few among the popular ones. But all of them have some problems with them. For example, Gabor

wavelets are effective against photometric disturbances but the high dimension of the extracted Gabor feature vector is a concern. Considering the issues with all the feature extraction techniques, to extract the features, a fusion of Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP) is used. This paper is not the first paper that uses the combination of both of them. Many FER systems have already used this mechanism successfully (An and Ruan, 2016; Islam *et al*., 2018b; 2018c).

For classification purpose, Artificial Neural Networks (ANN) (Islam *et al*., 2018a), multiclass Support Vector Machines (SVM) (Islam *et al*., 2018b), SVM (Jayalekshmi and Mathew, 2017), random forests (Jayalekshmi and Mathew, 2017), nearest neighbor classification (Agrawal and Yadav, 2018), AdaBoost classification (Verma and Dabbagh, 2013), Extreme Learning Machine (ELM) (Liu *et al*., 2015; Ghimire and Lee, 2014; Alphonse and Dejey, 2016) are among some which have been widely used. Now a days deep learning is also used extensively to classify expressions (Fathallah *et al*., 2017). ANN requires long training time, many training samples and many adjustable parameters to be tuned. Proper kernel function and regularization term, overfitting, training time are challenges with SVM. The number of neighbors to be considered and proper distance metric are crucial while using nearest neighbor classification. Noisy data can be problematic when using

AdaBoost classification. To handle these problems, the proposed method uses ELM to classify the expressions to their corresponding classes. Next section of the paper briefly describes the proposed method and the following few sections elaborately describe each of the steps of the proposed method. Finally, last few sections analyze the performance of the proposed method, compare performance with other methods and discuss the flaw, future possibilities of the work.

Proposed Method

At first, taking an image as an input, it is converted to grayscale. The face region from the grayscale image is detected using Viola-Jones face detection method (Viola and Jones, 2001). Then the facial image is converted to a fixed size. These three steps are cumulatively termed as image preprocessing. Following the proposed image segmentation method, the preprocessed image is segmented into four facial expression regions (right eye, left eye, nose, mouth). From these segmented parts, features are extracted using HOG and LBP. PCA is then used to reduce the dimensionality of the feature vector. Finally, ELM is used to train the system with some sample images. Rest of the images are used to test the performance of the system. The whole process is illustrated in Fig. 1.
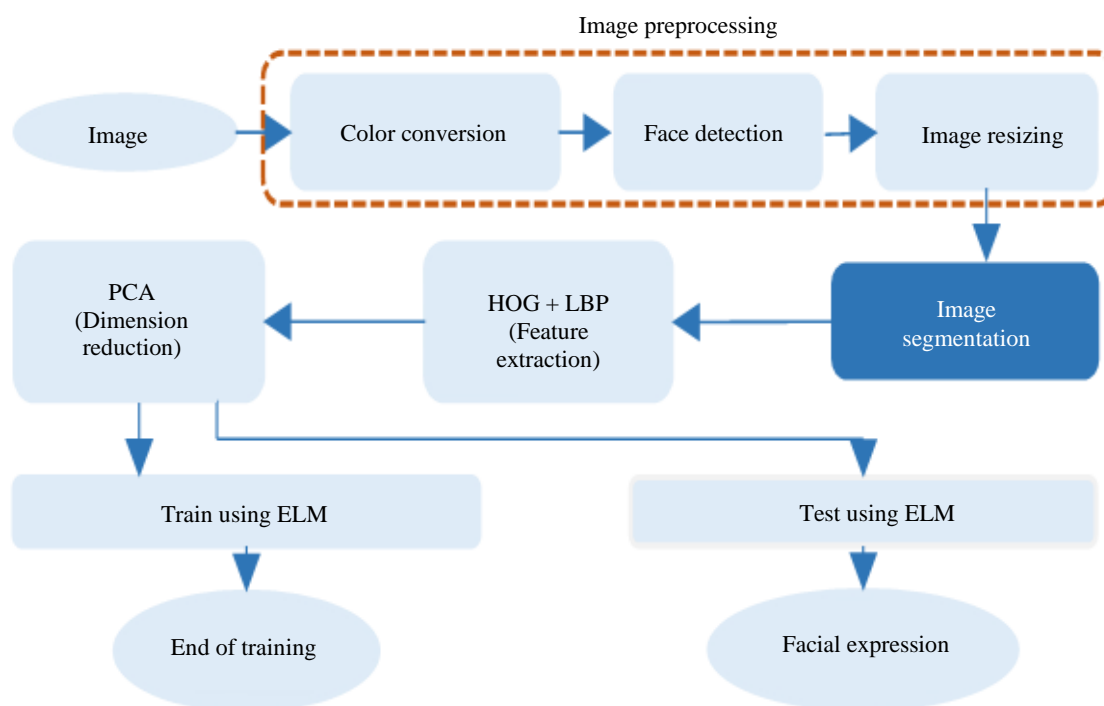

Fig. 1: The proposed method of facial expression recognition

Image Preprocessing

The input image of the system might be a color image or a grayscale image. If it is a color one then it is converted to grayscale otherwise, this step is omitted. From the grayscale image, only the facial region is detected using Viola-Jones face detection method (Viola and Jones, 2001), if it has any face in it. The facial region is then resized to a fixed size of 150×150 pixels for using it in the image segmentation step. The dimension of an image from JAFFE, CK+ and RaFD dataset are 256×256, 640×490 and 681×1024 respectively. After applying Viola-Jones face detection technique the dimension of the resultant facial images are 155×155, 319×319 and 388×388 respectively. But for developing a uniform system it is necessary to convert them to a fixed size. So they are converted to 150×150 because if they were converted to a much lower or much higher dimension then the image could lose some vital required information. The process is shown in Fig. 2 on a sample image from RaFD dataset (Langner *et al.*, 2010).
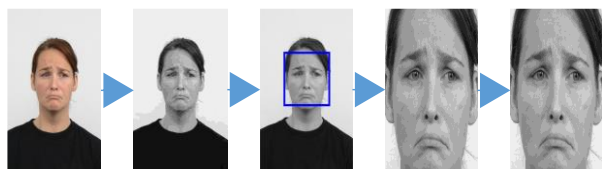

Fig. 2: Steps of image preprocessing

Image Segmentation

The purpose of this image segmentation step is to segment the facial image into four parts that contribute most in facial expressions. Different methods are available to segment the contributing parts from a facial image. Viola-Jones object detection method (Viola and Jones, 2001), Active Appearance Models (AAM) (Cootes *et al.*, 2001) are two established methods. But each has some drawbacks while using in real-world problems. When the eyes are closed or almost closed then Viola-Jones method fails to detect the eyes. Dependency on proper initialization, wide object variations in the training set are challenges with AAM. To overcome these problems, a manual segmentation method is proposed in this paper to segment right eye, left eye, nose and mouth from any facial image of size 150×150. At first, four different coordinate points are defined for the four parts and then four different regions are cropped using four defined width and height values. Let us consider the example of segmenting nose from an image to understand the procedure. At first, the coordinate point 54.33, 81.84 is found from the preprocessed facial image of size 150×150. Then a region with width 45.43 and height 38 is selected and used as the nose region. Same procedure with different coordinate values and width, height are applied to different all four parts successfully. The required values for segmenting the four parts from a preprocessed image of size 150×150 are given in Table 1.

Table 1: 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 these four parts in those images. Analyzed images were from people of different races, ethnic groups and different parts of the world. The challenge with this segmentation method was to segment the four parts as accurately as possible and at the same time with the least possible dimension. When the values of Table 1 are applied to an image of size 150×150 then the four parts get segmented as shown in Fig. 3. The whole image segmentation method is step by step illustrated on a sample image from RaFD dataset (Langner *et al.*, 2010) in Fig. 4 for clearly understanding the process.
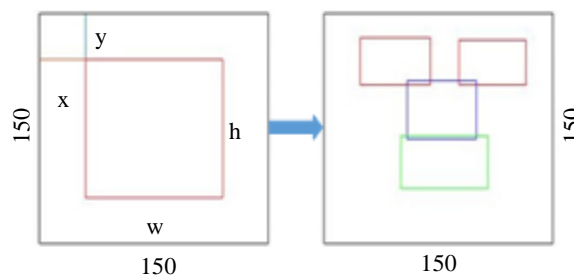

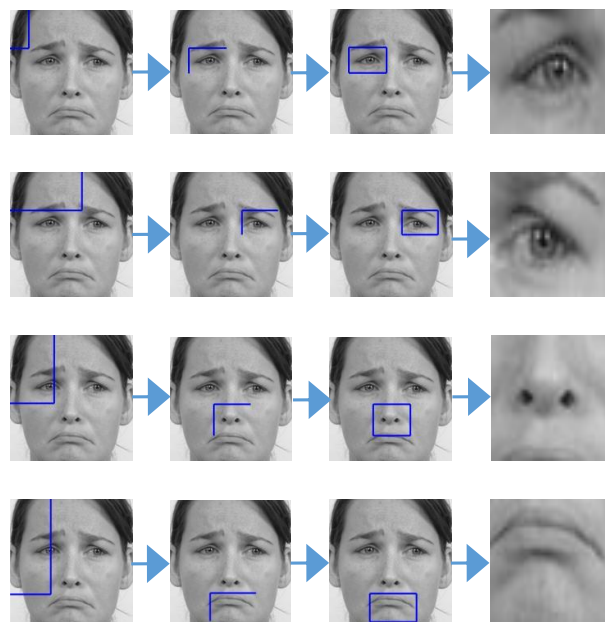**Fig. 3:** Proposed image segmentation method on a block of size

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

Feature Extraction

A fusion of HOG features and LBP features are used to extract the useful features from the segmented facial parts.

A. Histogram of Oriented Gradients (HOG)

Robustness against geometric and photometric transformations are the key benefits that led to use HOG as the feature extraction method. Due to its effectiveness, it has been used in many image processing tasks. Dalal and Briggs (2005) showed a widely accepted use of HOG to detect objects. Steps to calculate HOG features from any image are given below:

- Create small adjacent cells by dividing the whole image
- Calculate both the gradient magnitude and direction for all pixels in each cell
- Compute the appropriate bin for every gradient magnitude, direction and represent it using a histogram of gradients
- Blocks are built from adjacent cells and feature vector is calculated by performing block normalization

Choosing an appropriate number of bins, size of cells, blocks are challenging and problem specific. During the implementation, the unsigned orientation of gradients, 9 bins in the histogram, cells of size 8×8 pixels and blocks of size 2×2 cells were used. Fig. 5 shows feature extraction on the left eye of a sample image from RaFD dataset (Langner *et al*., 2010) using HOG.
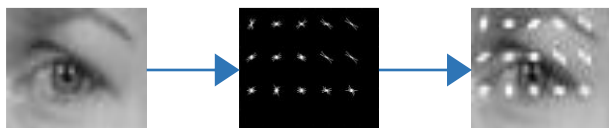

Fig. 5: HOG feature extraction on the left eye of a sample image

B. Local Binary Patterns (LBP)

LBP can handle illumination changes, they are computationally efficient as well. LBP has been widely used in many problems such as image analysis, facial analysis and motion analysis. A popular work using LBP for analyzing texture is shown by Ojala *et al*. (2002). Steps to calculate LBP features of an image are as follows:

- Create small cells with a specified radius and number of neighbors from the whole image
- Perform thresholding considering the center pixel and its neighbor pixels

- A binary number (LBP) is found as the outcome 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 and compute the feature vector

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

C. Fusion

At first, HOG features are extracted from the four segmented parts and then LBP features are extracted from the same four segmented parts. The final feature vector is formed by concatenating the features found using HOG and LBP. Length of the final feature vector was 1892, among them 1656 features were the contribution of HOG descriptor and rest were contributed by LBP.

Dimension Reduction

PCA can represent the data in terms of its principal components (Smith, 2003). For this ability, PCA is used to reduce the dimension of the feature vector. Steps of PCA are as follows:

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

During the implementation, PCA was used aiming to retain 99% variance ratio and it resulted in up to 89.85% dimension reduction. For example, with JAFFE dataset (Lyons *et al*., 1998) 1892 features were reduced to 192 features which means 89.85% (1-(192÷1892)) dimension reduction. Although the dimension of the four segmented parts was different they were same for all sample images and their summation was also same for all sample images. So when given to classifier all images had a same number of features for a segmented part and as a result, there was no conflict while comparing them for classification purpose.

Classification

Extreme Learning Machine (ELM) is a single hidden layer feed-forward neural network. It is named such because of its ability to learn thousands of time faster than traditional neural network approaches. ELM

works by finding the minimum norm least square solution of a system. As it is a feed-forward network, so there is no necessity of propagating the objective function's current state as done in backpropagation. So the solution is found just in a single forward iteration and so the algorithm is extremely fast. With ELM, the only parameter required to be tuned is the number of nodes in the hidden layer. But with conventional neural network approaches, many parameters are required to be tuned to extract the best out of the algorithm. ELM also aims to use the minimum norm of weights which leads to a better generalization performance. These above-mentioned properties of ELM make it a superior choice for classification task in image processing problems like FER (Huang *et al.*, 2004). Given a training set $\left\{\left(x_i,t_j\right)|x_i \in R^n, t_j \in R^m, i=1,2,\ldots,N\right\}$, L number of nodes in hidden layer, hidden node output function $g(w,b,x)$, then the algorithm is as mentioned below:

- Randomly generate input weights w_i and bias b_i, $i = 1,2,\ldots,L$
- Use the following formula to calculate the hidden layer output matrix H:

$$H = \begin{bmatrix} g\left(w_1.x_1+b_1\right)\ldots\ldots g\left(w_L.x_1+b_L\right) \\ .\qquad\ldots\ldots\qquad . \\ .\qquad\ldots\ldots\qquad . \\ g\left(w_1.x_N+b_1\right)\ldots\ldots g\left(w_L.x_N+b_L\right) \end{bmatrix} \qquad (1)$$

- Calculate the output weights $\beta, \beta = H+T$
- T being the target output and $H+$ the Moore-Penrose generalized inverse of H, calculated as:

$$H^+ = \left(H^T H\right)^{-1} H^T \qquad (2)$$

Sigmoid activation function and 2500 nodes in the hidden layer were used during the implementation.

Result Analysis

For implementing the proposed FER method, a computer with 64-bit operating system, 4GB of memory and core-i5 processor was used. Three publicly available facial expression dataset JAFFE (Lyons *et al.*, 1998), CK+ (Lucey *et al.*, 2010) and RaFD (Langner *et al.*, 2010) were used. All 213 images from JAFFE dataset, 1219 facial expression images of 22 people from CK+ dataset and 1407 front facing facial expression images of 67 people from RaFD dataset were used. Sample images from these datasets are shown in Fig. 6 to 8. Seven facial expressions were considered. Contemptuous expressions were not considered from RaFD dataset. K-fold cross-validation was used to avoid biased results. Splitting of training and testing sets were done randomly.

Average accuracy achieved on these three datasets by the proposed method is mentioned in Table 2. Here Correct Recognition Rate (CRR) is meant by accuracy. Average accuracy by traditional method is also mentioned. Traditional method refers to using full face to extract features. No segmentation or region of interest is considered in the traditional method. The feature-length without any dimension reduction is also mentioned in Table 2. It is apparent that there is not a huge difference in the average accuracy achieved by the proposed method and the traditional method. But there is a huge difference in the length of the feature vector. Obviously dimension reduction techniques can be applied to reduce the dimension of the feature vector as done in this paper. But reducing 1892 features to a few hundred and 10463 features to a few hundred are not computationally equivalent. The latter would require longer time and be a hindrance to creating a real-time system. Comparison of performance, of the three used datasets, by the proposed method is shown in Fig. 9.

Fig. 6: Sample images from CK+ dataset

Fig. 7: Sample images from RaFD dataset

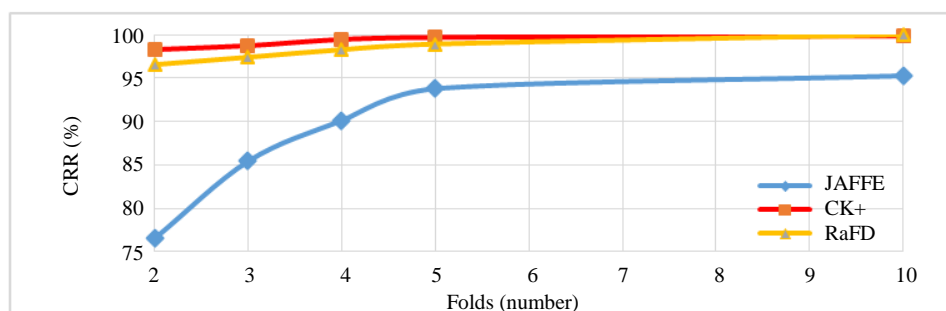**Fig. 8:** Sample images from JAFFE dataset

Fig. 9 Performance comparison on JAFFE, CK+ and RaFD dataset

Table 2: Accuracy of different datasets

Dataset Used	No. of Folds	Proposed Method		Traditional method	
		Average	Feature length (no dimension reduction)	Average	Feature length (no dimension reduction)
JAFFE	5	93.81	1892	93.92	10463
JAFFE	10	95.30		95.35	
CK+	5	99.67		99.75	
CK+	10	99.84		99.75	
RaFD	5	98.01		97.58	
RaFD	10	98.65		98.51	

It is a plot of CRR versus a different number of folds for the three datasets. It shows that with limited number of training data the results are inferior with the JAFFE dataset. But with less number of training images the results are quite similar for CK+ and RaFD dataset. As the number of folds and thus the number of training image increases, correct recognition rate also gets increased.

Accuracy is not always the ultimate metric to analyze any FER system. Confusion matrices are also analyzed to measure the performance of FER systems. Table 3, to 5 represent confusion matrices for JAFFE, CK+, RaFD datasets respectively on random test cases. From the confusion matrix, it is evident that sad expression images had the lowest CRR for JAFFE dataset for that particular test case. The system classified some of the sad images as anger images and some as disgust images for a certain test case. Fear expression images had the lowest correct recognition rate for CK+ dataset on a random test case. Only neutral and disgust images were classified properly while considering RaFD dataset. As these are random test cases so the same confusion matrices are not guaranteed when new test cases are generated.

Table 3: Confusion matrix of CRR on JAFFE dataset

	Neutral	Happy	Anger	Surprise	Fear	Disgust	Sad	Accuracy
Neutral	100	0	0.00	0	0	0.00	0.00	100.00
Happy	0	100	0.00	0	0	0.00	0.00	100.00
Anger	0	0	100.00	0	0	0.00	0.00	100.00
Surprise	0	0	0.00	100	0	0.00	0.00	100.00
Fear	0	0	0.00	0	100	0.00	0.00	100.00
Disgust	0	0	0.00	0	0	100.00	0.00	100.00
Sad	0	0	14.29	0	0	14.29	71.42	71.42

Table 4: Confusion matrix of CRR on CK+ dataset

	Neutral	Happy	Anger	Surprise	Fear	Disgust	Sad	Accuracy
Neutral	98.55	0	0	1.45	0.00	0.00	0	98.55
Happy	0.00	100	0	0.00	0.00	0.00	0	100.00
Anger	0.00	0	100	0.00	0.00	0.00	0	100.00
Surprise	0.00	0	0	100.00	0.00	0.00	0	100.00
Fear	0.00	0	0	0.00	96.61	3.39	0	93.61
Disgust	0.00	0	0	0.00	0.00	100.00	0	100.00
Sad	0.00	0	0	0.00	0.00	0.00	100	100.00

Table 5: Confusion matrix of CRR on RaFD dataset

	Neutral	Happy	Anger	Surprise	Fear	Disgust	Sad	Accuracy
Neutral	100.00	0.00	0.00	0.00	0.00	0	0.00	100.00
Happy	1.64	98.36	0.00	0.00	0.00	0	0.00	98.36
Anger	1.61	0.00	95.16	0.00	0.00	0	3.23	95.16
Surprise	1.37	0.00	0.00	98.63	0.00	0	0.00	98.63
Fear	0.00	0.00	0.00	2.70	97.30	0	0.00	97.30
Disgust	0.00	0.00	0.00	0.00	0.00	100	0.00	100.00
Sad	1.61	0.00	0.00	0.00	0.00	0	98.39	98.39

State-of-The-Art

Performance comparison of the proposed method with other available methods of FER would justify the relative position of the proposed method. Table 6 reflects the performance of the system by comparing with other FER systems on the used three datasets. Liu *et al.* (2015) used Viola-Jones face detection method to detect faces from images and then used Active Appearance Model (AAM) to segment the facial images into its parts. Gabor filters were used to extract features from those parts and PCA to reduce the dimension of the feature vector. Finally ELM was used to classify the obtained features. Training sets and testing sets were generated from JAFFE and CK+ dataset and their system resulted in 94% and 95% accuracy on these two datasets respectively. In their work, Ghimire and Lee (2014) used HOG to extract features from facial images and then they used ensemble ELM and finally they selected the winner by voting scheme. Throughout the process they used the images from JAFFE and CK+ dataset and achieved 94.37% and 97.3% CRR respectively. Alphonse and

Dejey (2016) used images from JAFFE dataset to develop an FER system. They considered 77 facial points from any facial image as feature and then again they extracted features from facial images using Gabor filter. Features were then fused and PCA was used to reduce the dimensionality. Finally they used Extreme Learning Machine, Support Vector Machine, K Nearest Neighbour and Partial Least Squares (PLS) to classify the features. They concluded that ELM outperforms other classifiers used by them. Rao *et al.* (2015) proposed a SURF boosting framework and then used gentle AdaBoost with logistic regression for classification. They have considered images from RaFD dataset for their experiment. Using visual saliency and deep learning Mavani *et al.* (2017) developed an FER system. They have used CFEE and RaFD dataset. By training with CFEE dataset and testing with RaFD dataset they got an accuracy of 95.7%. Zeng *et al.* (2018) obtained 95.79% accuracy while testing images from CK+ dataset on their proposed system which considered facial geometric features and appearance based feature. Deep sparse auto encoders were the basis of their framework.

Table 6: State-of-the-art comparison

Study	Technique Followed	Dataset	Accuracy
Proposed Method	Viola-Jones face detection, **facial region segmentation**, HOG+LBP, PCA, ELM	JAFFE	**95.30**
		CK+	**99.84**
		RaFD	**98.65**
Liu *et al.* (2015)	Viola-Jones face detection, AAM, Gabor filter, PCA, ELM	JAFFE	94.00
		CK+	95.00
Ghimire and Lee (2014)	HOG, ELM ensemble using bagging	JAFFE	94.37
		CK+	97.30
Alphonse and Dejey (2016)	Facial points, Gabor filter, PCA, ELM	JAFFE	91.40
Rao *et al.* (2015)	SURF, gentle AdaBoost	RaFD	90.64
Mavani *et al.* (2017)	Visual saliency, deep learning	RaFD	95.70
Zeng *et al.* (2018)	AAM, HOG, PCA, deep sparse auto encoders	CK+	95.79

Conclusion

Unique, simple yet effective image segmentation method along with a fusion of features and selection of proper classification algorithm led to achieving higher accuracy compared to other FER systems. One of the most controversial steps of the proposed method is its image segmentation method. Even initially, the image segmentation process might seem to be absurd but the result section justifies its effectiveness. The system was tested with almost 3000 front facing images and all of them were handled properly by the system. But if any person has an unusual facial structure then that facial image might not be segmented properly by the system due to the unusual position of the four parts. Some facial expressions are even tough for a human to recognize, let alone machines.

For the time being the system is capable of handling only front-facing images and it is an obstacle in the way of creating a robust system. So developing a more robust system capable of handling images rotated at any angle in any direction would be a future work. Beside facial information, vocal information also contributes in determining emotion of a person. So developing a system to recognize the emotion of a person by using both facial and vocal information would be another interesting work in near future.

Author's Contributions

Bayezid Islam: Contributed in designing and implementing the system along with the analysis and preparation of manuscript.

Arfat Hossain: Contributed at the design and implementation level of the proposed system.

Ethics

The article is original unpublished material and so the authors ensure that after the publication of this manuscript no ethical issue would arise.

References

Agrawal, S. and S. Yadav, 2018. Approach based on HOG Descriptor and K nearest neighbour for facial expressions recognition. Int. J. Innovat. Adv. Comput. Sci., 7: 239-249.

Alphonse, A.S. and Dejey, 2016. Performance analysis of classifiers for facial expression recognition under constrained settings. Proceedings of the International Conference on Computing Technologies and Intelligent Data Engineering, Jan. 7-9, IEEE Xplore Press, India, pp: 1-6. DOI: 10.1109/ICCTIDE.2016.7725337

An, S. and Q. Ruan, 2016. 3D facial expression recognition algorithm using local threshold binary pattern and histogram of oriented gradient. Proceedings of the IEEE 13th International Conference on Signal Processing, Nov. 6-10, IEEE Xplore Press, China, pp: 265-270. DOI: 10.1109/ICSP.2016.7877838.

Cootes, T.F., G.J. Edwards and C.J. Taylor, 2001. Active appearance models. IEEE Trans. Patt. Anal. Machine Intell., 23: 681-685. DOI: 10.1109/34.927467

Dalal, N. and T. Briggs, 2005. Histograms of oriented gradients for human detection. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Jun. 20-25, IEEE Xplore Press, USA, pp: 886-893.
DOI: 10.1109/CVPR.2005.177

Fathallah, A., L. Abdi and A. Douik, 2017. Facial expression recognition via deep learning. Proceedings of the IEEE/ACS 14th International Conference on Computer Systems and Applications, Nov. 3-3, IEEE Xplore Press, Tunisia, pp: 745-750. DOI: 10.1109/AICCSA.2017.124

Ghimire, D. and J. Lee, 2014. Extreme learning machine ensemble using bagging for facial expression recognition. J. Inform. Process. Syst., 10: 443-458. DOI: 10.3745/JIPS.02.0004

Huang, G.B., Q.Y. Zhu and C.K. Siew, 2004. Extreme Learning Machine: A new learning scheme of feedforward neural networks. Proceedings of the IEEE International Joint Conference on Neural Networks, Jul. 25-29, IEEE Xplore Press, Hungary, pp: 985-990. DOI: 10.1109/IJCNN.2004.1380068

Islam, B., F. Mahmud and A. Hossain, 2018a. Facial expression region segmentation based approach to emotion recognition using 2D Gabor filter and multiclass support vector machine. Proceedings of the 21st International Conference of Computer and Information Technology, Dec. 21-23, IEEE Xplore Press, Bangladesh, pp: 1-6. DOI: 10.1109/ICCITECHN.2018.8631922

Islam, B., F. Mahmud and A. Hossain, 2018b. High performance facial expression recognition system using facial region segmentation, fusion of HOG and LBP features and multiclass SVM. Proceedings of the 10th International Conference on Electrical and Computer Engineering, Dec. 20-22. IEEE Xplore Press, Bangladesh, pp: 42-45. DOI: 10.1109/ICECE.2018.8636780

Islam, B., F. Mahmud, A. Hossain, P.B. Goala and M.S. Mia, 2018c. A facial region segmentation based approach to recognize human emotion using fusion of HOG and LBP features and artificial neural network. Proceedings of the 4th International Conference on Electrical Engineering and Information and Communication Technology, Sept. 13-15, IEEE Xplore Press, Bangladesh, pp: 1-6. DOI: 10.1109/CEEICT.2018.8628140

Jayalekshmi, J. and T. Mathew, 2017. Facial expression recognition and emotion classification system for sentiment analysis. Proceedings of the International Conference on Networks and Advances in Computational Technologies, Jul. 20-22, IEEE Xplore Press, India, pp: 1-8. DOI: 10.1109/NETACT.2017.8076732

Langner, O., R. Dotsch, G. Bijlstra. D.H. Wigboldus and S.T. Hawk *et al.*, 2010. Presentation and validation of the radboud faces database. Cognit. Emot., 24: 1377-1388. DOI: 10.1080/02699930903485076

Liu, Z.T., G.T. Sui, D.Y. Li and G.Z. Tan, 2015. A novel facial expression recognition method based on extreme learning machine. Proceedings of the 34th Chinese Control Conference, Jul. 28-30, IEEE Xplore Press, China, pp: 3852-3857. DOI: 10.1109/ChiCC.2015.7260233

Lucey, P., J.F. Cohn, T. Kanade, J. Saragih and I. Matthews *et al.*, 2010. The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Jun 13-18, IEEE Xplore Press, USA, pp: 94-101. DOI: 10.1109/CVPRW.2010.5543262

Lyons, M.J., S. Akemastu, M. Kamachi and J. Gyoba, 1998. Coding facial expressions with gabor wavelets. Proceedings of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition, Apr. 14-16, IEEE Xplore Press, Japan, pp: 200-205. DOI: 10.1109/AFGR.1998.670949

Mahmud, F., M.T. Khatun, S.T. Zuhori, S. Afroge and M. Aktar *et al.*, 2015. Face recognition using principle component analysis and linear discriminant analysis. Proceedings of the International Conference on Electrical Engineering and Information Communication Technology, May 21-23, IEEE Xplore Press, Bangladesh, pp: 1-4. DOI: 10.1109/ICEEICT.2015.7307518

Mavani, V., S. Raman and K.P. Miyapuram, 2017. Facial expression recognition using visual saliency and deep learning. Proceedings of the IEEE International Conference on Computer Vision Workshops, Oct. 22-29, IEEE Xplore Press, Italy, pp: 2783-2788. DOI: 10.1109/ICCVW.2017.327.

Mehrabian, A., 1968. Communication without words. Psychol. Today, 2: 53-56.

Ojala, T., M. Pietikainen and T. Maenpaa, 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Patt. Anal. Machine Intell., 24: 971-987. DOI: 10.1109/TPAMI.2002.1017623

Rao, Q., X. Qu, Q. Mao and Y. Zhan, 2015. Multi-pose facial expression recognition based on SURF boosting. Proceedings of the International Conference on Affective Computing and Intelligent Interaction, Sept. 21-24, IEEE Xplore Press, China, pp: 630-365. DOI: 10.1109/ACII.2015.7344635

Sandbach, G., S. Azfeiriou, M. Pantic and L. Yin, 2012. Static and dynamic 3D facial expression recognition, A comprehensive survey. Image Vis. Comput., 30: 683-697. DOI: 10.1016/j.imavis.2012.06.005

Smith, L.I., 2002. A tutorial on principal component analysis. department of computer science, University of Otago, Dunedin, New Zealand.

Verma, R. and M.Y. Dabbagh, 2013. Fast facial expression recognition based on local binary patterns. Proceedings of the 26th IEEE Canadian Conference on Electrical and Computer Engineering, May 5-8, IEEE Xplore Press, Canada, pp: 1-4. DOI: 10.1109/CCECE.2013.6567728

Viola, P. and M. Jones, 2001. Rapid object detection using a boosted cascade of simple features. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Dec. 8-14, IEEE Xplore Press, USA, pp: 1-1. DOI: 10.1109/CVPR.2001.990517

Zeng, N., H. Zhang, B. Song, W. Liu, Y. Li and A.M. Dobaie, 2018. Facial expression recognition via learning deep sparse autoencoders. Neurocomputing, 273: 643-649. DOI: 10.1016/j.neucom.2017.08.043