International Journal of Research in Advent Technology, Vol.7, No.3, March 2019
E-ISSN: 2321-9637
Available online at www.ijrat.org

Human Emotion Recognition through Facial Expression: An Approach Based on LBP Features and SVM Classifier

Shagolshem Thonglen Singh[1], Dinesh R[2]
Demonstration School RIE (NCERT), Research Scholar, Bharathiar University Coimbator[1]
Department of Information Science and Engineering, Jain University Bangalore[2]
Email: stsinghr@gmail.com[1], dr.dineshr@gmail.com [2]

Abstract- In this paper we have proposed a method for recognizing human emotion based on facial expressions. The proposed method detects the face in real time by camera attached to handheld devices like mobile phone. The detected face image is further processed to determine the emotion. Face images are represented using LBP features and SVM classifier is used to recognize the emotion. The efficacy of the proposed method has been ascertained by conducting extensive experiments on large set of face databases including standard databases. It is found through the experiments that the proposed method has higher recognition rate of over 98% and the method can determine the emotion in under 100 milliseconds and hence suitable for real time applications.

Keywords- Emotion Recognition, Face Detection, HAAR Features, Local Binary Patterns, Support Vector Machine.

1. INTRODUCTION

Facial emotions are very important and play significant role in communication among humans. Emotions convey the motive of the persons. Facial emotional recognition (FER) is an interdisciplinary domain standing at the crossing of behavioral science, psychology, neurology and artificial intelligence. In this new competitive world security are very much concerned in public and private sectors. Emotional face recognition has become one of the most active research areas at present and many organizations are showing interest to invest. Due to this reason emotional face recognition is one of the hottest research areas in computer vision and machine learning. Very few different emotional face recognition algorithms have been designed. Face is the most commonly used for identification, recognition and verification. Emotional face recognition are so important in day today's life. Passwords and cards are the most extensive used in biometric system; because of non-contact process face recognition is the most advantageous and effective systems. It is the most complex object recognition biometric system. Because human faces are the most similar structure. Due to this reason many automatic face recognition algorithm have been designed but still the recognition rate is not up to the mark. Still there is no fixed benchmark database for testing existing algorithms. The most drawback of face recognition are inclusion of facial expression, illumination, aging, hair style, pose, scaling, frontal, present or absent of spectacle and occlusion. Lighting condition, facial expression and deep learning are the most talking in the recent past. Bright lighting causes image saturation. Emotions often mediate and facilitate interactions among human beings. Thus, understanding emotion often brings context to seemingly bizarre and/or complex social communication. Emotion can be recognized through a variety of means such as voice intonation, body language, and more complex methods; such electroencephalography (EEG) [1]. However, the easier, more practical method is to examine facial expressions. There are seven types of human emotions shown to be universally recognizable across different cultures [2]: anger, disgust, fear, happiness, sadness, surprise, contempt. Interestingly, even for complex expressions where a mixture of emotions could be used as descriptors, cross-cultural agreement is still observed [3]. Therefore a utility that detects emotion from facial expressions would be widely applicable. Such advancement could bring applications in medicine, marketing and entertainment [4]. The task of emotion recognition is particularly difficult for two reasons: 1) There does not exist a large database of training images and 2) classifying emotion can be difficult depending on whether the input image is static or a transition frame into a facial expression. The latter issue is particularly difficult for real-time detection where facial expressions vary dynamically. Most applications of emotion recognition examine static images of facial expressions. We investigate the application of subspace method for determining human emotion from facial expressions in this paper. Over the last two decades, researchers have significantly advanced human facial emotion recognition with computer vision techniques. Historically, there have been many approaches to this problem, including using pyramid histograms of gradients (PHOG) [5], AU aware facial features [6], boosted LBP descriptors [7], and RNNs [8]. Emotional face recognition is a part of deep learning which is part of a broader family of machine learning method. The architecture involved in deep learning is deep neural networks, deep belief networks, recurrent neural network. These architecture have been applied in biometric, speech recognition, computer vision, natural language processing, audio recognition, social networks filtering, machine learning, drug designing, medical image analysis etc., where better result is given by the human brain. Few works have been proposed in emotional facial expression recognition. For instance Diah A P, et al., [14] used convolution neural network (CNN) for feature extraction and classification of 6 basic human face emotions. Face detection, cropping, resize, adding noise and data normalization are used as pre-processing steps. He used first convolution layer with a kernel size 5 X 5 and stride 1 and extract features like edges, oriented edges, corners and shape. They applied max pooling in fourth layer and followed by fully connected layer that has 25 neurons. In their paper cross-entropy was used for error function. But hard to classify sadness, anger, and fear. Jiaxing, et al., [15] proposed Faster Regions with Convolutional Neural network (Faster R-CNN) for implicit feature extraction by using trainable convolution kernel. Then, maximum pooling is used to reduce the dimensions of the extracted implicit features. Next, RPNs (Region Proposal Networks) is used to generate high-quality region proposals, by Faster R-CNN for detection. Finally, the softmax classifier and regression layer is used to classify the facial expressions and boundary of the test sample. Jun Ou [16] introduced new Classification technique in emotional facials called k-nearest neighbor such as given n data points of d-dimensional space, divided into k group. Ramprakash, et al., [17] introduced a

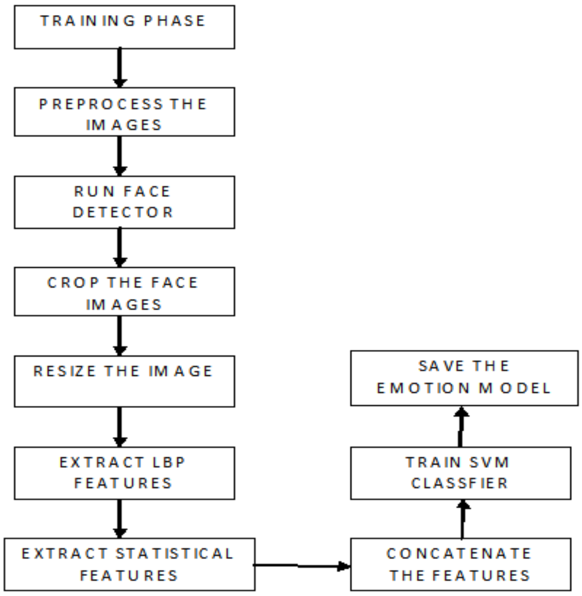

Fig 1 (a) Training phase of the proposed method

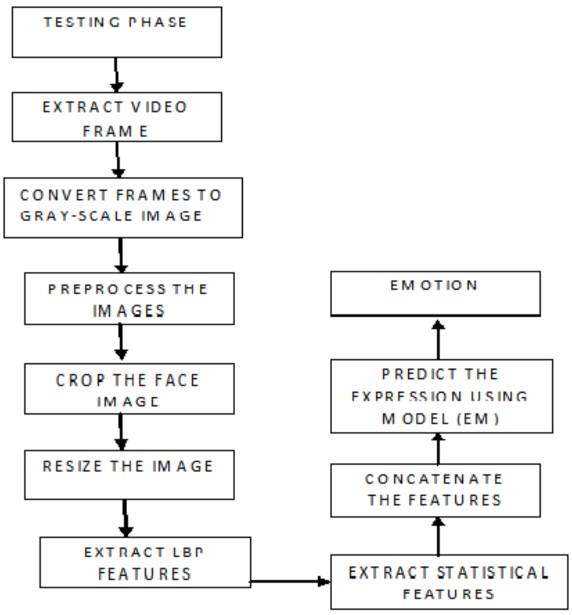

Fig 1(b). Testing phase of the proposed method

new method called action units (AU) which describe brain region thought to be responsible for the changes in the facial expression. Anurg, et al., [18] used Eigen space and PCA has used for classification. Euclidean distance has used for recognition.

2. PROPOSED METHOD

Fig 2. Image before and after application of CLAH Algorithm.

This section of the paper presents the details of the proposed method. The proposed method has been trained and tested on various standard benchmarked datasets. As it can be seen from Fig-1 the proposed method for emotion detection has two stages; (a) Training stage which mainly focuses on building an Emotion Recognition model (EM) and (b) Testing stage, which mainly focus on predicting the emotion of the person/people appear in the video frame. The following sub-section gives the details of each steps of the proposed method.

2.1 ***Image Processing***

Most of the times when the images are acquired from vision sensors; there will be lot of lighting variations. Large variations in the illumination will adversely impact the performance of the face detection and also the emotion recognition system. To overcome the problem of significant illumination variations, in the proposed approach we pre-process all the images with the Contrast Limiting Adaptive Histogram Equalization (CLAHE) [10]. Fig-2 shows the images before and after pre-preprocessing, it can be clearly seen from the figure that the excessive illumination variations are addressed effectively by the pre-processing technique.

The pre-processed images are further cropped and resized to the standard size. In the proposed method, we have resized the image to 128x128. In case of testing phase, we subject the image/video frame to the face detection. In the proposed work we have used the face detector proposed by Viola-Jones [11] and we have made use the code available in OpenCV by using already trained model. Once the faces are detected, the detected rectangle coordinates are used to crop the face region and further they are resized to a standard size of 128x128.

The characteristics of Viola–Jones algorithm which make it a good detection algorithm are:

- Robust – very high detection rate (true-positive rate) & very low false-positive rate always.
- Real time – For practical applications at least 2 frames per second must be processed.

Face detection only (not recognition) - The goal is to distinguish faces from non-faces (detection is the first step in the recognition process).

The algorithm has four stages:

(1) Haar Feature Selection

(2) Creating an Integral Image

(3) Adaboost Training

(4) Cascading Classifiers

The features sought by the detection framework universally involve the sums of image pixels within rectangular areas. As such, they bear some resemblance to Haar basis functions, which have been used previously in the realm of image-based object detection. However, since the features used by Viola and Jones all rely on more than one rectangular area, they are generally more complex. The figure on the right illustrates the four different types of features used in the framework. The value of any given feature is the sum of the pixels within clear rectangles subtracted from the sum of the pixels within shaded rectangles. Rectangular features of this sort are primitive when compared to alternatives such as steerable filters. Although they are sensitive to vertical and horizontal features, their feedback is considerably coarser.

Upon face detector successfully, to detect the face, we crop the images marked by the detector Minimum Bound Rectangle (MBR) and resize the image to standard size, in this research work we have resized all cropped face image to a standard size of 128x128.

In order to represent the training images in terms of its feature value, the proposed method uses feature descriptor derived by Local Binary Pattern (LBP). The

LBP feature is extracted from gray image. For LBP feature extraction the color image is converted into gray image, the salt and pepper noise is removed by

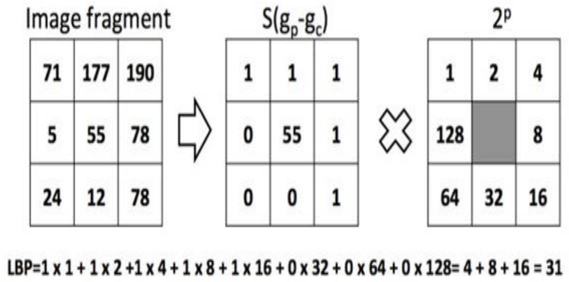

Fig. 3. LPB Operation.

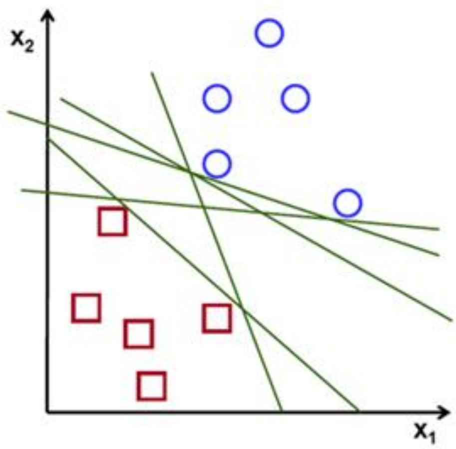

Fig 4. Samples in feature space and possible hyper plane.

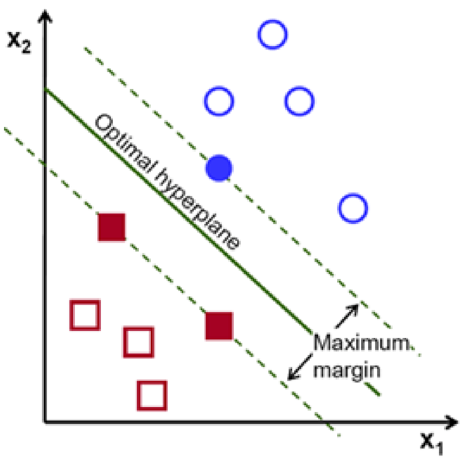

Fig 5. Samples in feature space and optimal hyper plane with maximum margin

averaging M x M neighborhood pixel and replaced centre pixel in the output image. The LBP feature extraction method first introduced by Ojala et al [12]. Local Binary Pattern is describing the texture details of a surface. The texture patterns are summarized into a histogram. The histogram values converted into single feature vector. The LBP operator works in a 3x3 pixel block of image, the central pixel of block adjusted to a threshold, the texture details can be obtain based on sign difference between centre pixel and its neighborhood pixels. For each pixel in the image, the binary value is obtained by thresholding neighbor pixel with central pixel. The neighbor pixel become one, if the compared value greater than or equal to threshold otherwise zero.

Fig 3 pictorially describes the LBP operation. The operator uses in two patterns one is uniform and one more is non uniform, the uniform pattern has number of bitwise transition from 0 to 1or vice versa is at most 2. For example 01110000 (2 transitions) and 11001111 (2 transitions) are uniform, where as pattern has more than 2 transition, such pattern called non-uniform pattern example 11001001 (4 transitions) and 01010011 (6 transitions). The occurrence of output label in non-uniform pattern is single label only where as in uniform pattern have 59 labels. In this work, we have considered an image of size 48x48. The LBP features extracted in a two level in the first level 59 features from whole image and second level the image divided into 8 cells horizontally and 8 cells vertically, each cell with 6x6 pixels, totally 64 cells are generated, again extract 59 features from each cell, totally it generates 59+ 64x59 = 3835 features. For statistical feature the input component image converted to gray images, the gray images bearing large variation in intensity. These variations are poses a challenge on recognition. For operational convinces image has to change bi-level intensity. The gray image is converted into a binary level by adjusting threshold using Otsu's approach [13].

Further to improve the Emotion recognition accuracy, the proposed method extracts additional features based on pixel density statistics. Statistical features are extracted by creating virtual Zoning on the detected face image. The statistical features viz, Euler number, Eccentricity, and Extent and orientation are extracted. In Zoning method the whole image divided into small zones, one feature value from each zone by considering number of 1's out of total element, in the Euler method by choosing number of connected component region minus number of holes present in the same connected component region, from Eccentricity method it create minimum bounding ellipse, major axis and minor axis, and calculate aspect ratio. The Extent is a ratio of number of pixel in the component region to number of pixel in a bounding box. The orientation is calculated by finding angle between X-axis to the major axis of ellipse.

By augmenting LBP features and Statistical feature, the proposed method forms a feature vector. The dimension of the feature vector is 3949. These feature vectors are used to train the classifier to recognize the emotion.

SVM classifier is used to recognize / classify the giving facial expression from the image. A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyper plane. In other words, given labeled training data (*supervised learning*), the algorithm outputs an optimal hyper plane which categorizes new examples. In which sense is the hyper plane obtained optimal? Let's consider the following simple problem:

For a linearly separable set of 2D-points which belong to one of two classes, find a separating straight line.

In Figure 4, you can see that there exist multiple lines that offer a solution to the problem. Are any of them better than the others? We can intuitively define a criterion to estimate the worth of the lines:

A line is bad if it passes too close to the points because it will be noise sensitive and it will not generalize correctly. Therefore, our goal should be to find the line passing as far as possible from all points as shown in Fig 5.

Then, the operation of the SVM algorithm is based on finding the hyper plane that gives the largest minimum distance to the training examples. Twice, this distance receives the important name of margin within SVM's theory. Therefore, the optimal separating hyper plane maximizes the margin of the training data.

For the purpose of training the component images, we have considered about 1000 images from nine different classes. Each class has about 100 images in the training set. We trained the SVM classifier by considering 80% of images for training and 20% for validation using 10 fold cross validation algorithm.

3. EXPERIMENTAL RESULTS

In this work, we have considered 180 images for each category of expressions; hence in totality we have 1080 images. Out of 180 images from each facial expression categories, we have randomly selected 100 images from each expression categories for training and 80 images for testing.

Emotion	Accuracy
Happy	91.25%
Sad	98.75%
Anger	100%
Surprise	93.75%
Excitement	96.25%
Neutral	90%

Table 1. Accuracy for different categories of Facial Expression.

	Happy	Sad	Anger	Surprise	Excitement	Neutral
Happy	73	0	2	2	0	1
Sad	0	79	0	0	0	1
Anger	2	0	80	0	0	0
Surprise	2	0	0	75	2	3
Excitement	0	0	0	2	77	3
Neutral	1	1	0	3	3	72
	91.25 %	**98.75%**	**100 %**	**93.75%**	**96.25 %**	**90.00 %**

Table 2. Confusion Matrix

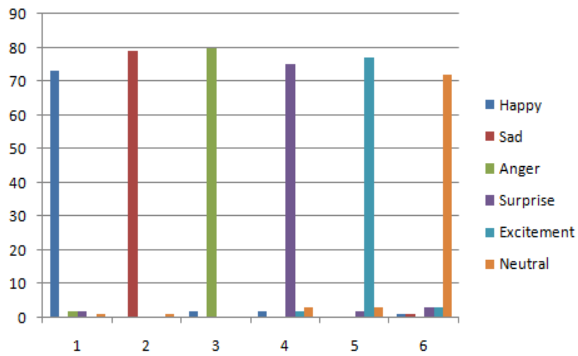

Fig 6. **Classification Accuracy of the proposed method**

Table-1 presents the recognition accuracy of the proposed method for each expression categories when tested the trained model on 80 images from each expression categories. Table-2 presents the details of the confusion matrix on test images for each category.

Following Figure shows the graphical representation of test image classification into different classes. It can be observed from the figure that the proposed method has very high classification rate for Anger emotion. Also, the proposed method has good recognition rate overall (>95%). However, the method has relatively less classification rate for Neutral facial expression; this is because, for most of the other expression, variations are not so vivid and looks similar to Neutral expression. Hence, the method has not successfully learnt the Neutral expression. This can be overcome, by adding more samples for the neutral facial expression or with appropriate sample augmentation, which we will take up in our future work.

4. CONCLUSION

In this paper we have presented a method for recognizing the facial expressions from both images and videos. The proposed method first detect the face using face detector and followed by the face images are represented with the features derived by Local Binary Patterns and Statistical features. Both LBP and Statistical features are combined to form a single feature vector. These feature vectors are trained using SVM classifier. The proposed method has been tested on standard facial expression datasets viz., AT&T Face dataset and Jaffe Datasets. From the experimental results it is evident that the proposed method has very high overall recognition rate of over 95%.

REFERENCES

[1] P. Abhang, S. Rao, B. W. Gawali, and P. Rokade, “Article: Emotion recognition using speech and EEG signal a review”, International Journal of Computer Applications , vol. 15, pp. 37–40, February 2011.

[2] P. Ekman, “Universals and cultural differences in facial expressions of emotion”, Nebraska, USA: Lincoln University of Nebraska Press, 1971.

[3] P. Ekman and W. V. Friesen, “Universals and cultural differences in the judgments of facial expressions of emotion”, Journal of Personality and Social Psychology, vol. 53, 1987.

[4] A. Kołakowska, A. Landowska, M. Szwoch, W. Szwoch, and M. R. Wrobel, “Human-Computer Systems Interaction: Backgrounds and Applications 3rd ch. Emotion Recognition and Its Applications”, pp. 51–62. Cham: Springer International Publishing, 2014.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks”, NIPS , vol. 1, p. 4, 2012.

[6] A. Yao, J. Shao, N. Ma, and Y. Chen, “Capturing au-aware facial features and their latent relations for emotion recognition in the wild”, in Proceedings of the 2015 ACM on International Conference on Multimodal Interaction , ICMI ’15, (New York, NY, USA), pp. 451–458, ACM, 2015.

[7] C. Shan, S. Gong, and P. W. McOwan, “Facial expression recognition based on local binary patterns: A comprehensive study”, Image and Vision Computing, vol. 27, no. 6, pp. 803 – 816, 2009.

[8] S. Kahou, V. Michalski, K. Konda, R. Memisevic, and C. Pal, “Recurrent neural networks for emotion recognition in video”, ICMI, pp. 467–474, 2015.

[9] Z. Yu and C. Zhang, “Image based static facial expression recognition with multiple deep network learning”, in Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, ICMI ’15, (New York, NY, USA), pp. 435–442, ACM, 2015.

[10] Zuiderveld, Karel. “Contrast Limited Adaptive Histograph Equalization”, *Graphic Gems IV*. San Diego: Academic Press Professional, 1994. 474–485.

[11] Viola Paul and Michael Jones, "Rapid object detection using a boosted cascade of simple features", International Conference on Computer Vision and Pattern Recognition, 2001.

[12] Ojala T., Pietikäinen M., Harwood D., “A comparative study of texture measures with classification based on feature distributions. Pattern Recognition. **29**(1), 51–59 (1996)

[13] Otsu, N., "A Threshold Selection Method from Gray-Level Histograms", IEEE Transactions on Systems, Man, and Cybernetics, Vol. 9, No. 1, 1979, pp. 62-66.

[14] Diah A P, Ajeng W, Basaruddin T, Dewi Y L, “Enhancing CNN with Preprocessing Stage in Automatic Emotion Recognition”, 2nd International Conference on Computer Science and Computational Intelligence 2017, ICCSCI 2017, 13-14 October 2017, Bali Indonesia.

[15] Jiaxing L, Dexiang Z, Jingjing Z, Jun Z, Teng L, Yi X, Qing Y, Lina X, “International Congress of Information and Communication Technology”, (ICIT 2017).

[16] Jun Ou, “Classification Algorithm Research on Facial Expression Recognition”, International Conference on Solid State Devices and Materials Science- 2012.

[17] Ramprakash S, Julie D G, Aleix M M, “A neural Basis of Facial Action Recognition in Humans”, the journals of neuroscience, April 20, 2016, 36(16):4434-4442.

[18] Anurag De, Ashim Saha, Dr. M C Pal, “A Human facial Expression Recognition Model based on Eigen Face Approach”, International Conference on Advanced Computing Technologies and Applications (ICACTA-2015).