A ES
ISSN: 2096-3246
Volume 54, Issue 02, December, 2022
Special Issue

EMOTION RECOGNITION USING HUMAN SPEECH AND FACIAL TECHNIQUE WITH MACHINE LEARNING APPROACHES

D.Shanthi
shanthid@dsce.ac.in
Dhanalakshmi Srinivasan College of Engineering, Coimbatore, Tamil Nadu, INDIA

T.Joby Titus
Dhanalakshmi Srinivasan College of Engineering, Coimbatore, Tamil Nadu, INDIA

D.Indhu
Dhanalakshmi Srinivasan College of Engineering, Coimbatore, Tamil Nadu, INDIA

V.K. Maurya
Babu Banarasi Das University, Lucknow (U.P.) India

Abstract: Identifying the Human using emotional status is the key parameter in image recognition as speech and facial expressions are the human cognitive-communication factors. The recent trends in Human computer Interaction demands the computer processing capability and recognition of facial and speech parameters. In this project, Human Speech and Facial based emotion recognition technique using a KNN has been proposed for improving the performance of detection with multi-emotions effectively. The obtained results of the proposed technique show that the average rate of recognition is higher than other recently existing techniques.

1. INTRODUCTION

Emotion recognition is being actively explored in Computer Vision research. With the recent rise and popularization of Machine Learning and Deep Learning techniques, the potential to build intelligent systems that accurately recognize emotions became a closer reality. However, this problem is shown to be more and more complex with the progress of fields that are directly linked with emotion recognition, such as psychology and neurology. Micro-expressions, electroencephalography (EEG) signals, gestures, tone of voice, facial expressions, and surrounding context are some terms that have a powerful impact when identifying emotions in a human. When all of these variables are pieced together with the limitations and problems of the current Computer Vision algorithms, emotion recognition can get highly complex. Facial expressions are the main focus of this systematic review. Generally, an FER system consists of the following steps: image acquisition, pre-processing, feature extraction, classification, or regression.

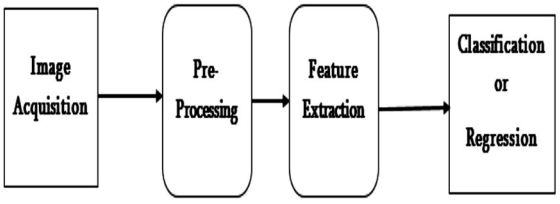

Fig 1: Block diagram of Facial Expression Recognition

To be able to get a proper facial expression classification, it is highly desirable to provide the most relevant data to the classifier, in the best possible conditions. In order to do that, a conventional FER system will firstly pre-process the input image. One pre-processing step that is common among most reviewed papers is face detection. Face detection techniques are able to create bounding boxes that delimit detected faces, which are the desired regions of interest (ROIs) for a conventional FER system. This task is still challenging, and it is not guaranteed that all faces are going to be detected in a given input image. This is especially true when acquiring images from an uncontrolled environment, where there may be movement, harsh lighting conditions, different poses, great distances, among other factors.

2. BACKGROUND WORK

2.1 Theory of Emotion

In automatic emotion recognition, it is important to examine the ideas proposed on the nature of emotions in so far as they shape the way emotional states are described. These ideas can guide us in determining what an emotional state is and what the relevant features are which distinguish this state from others. Looking back at the history of emotional theories we should mention Aristotle, who classified emotions into opposite and explained the physiological and hedonic qualities associated with emotions. Later Rene Descartes introduced the idea that a few emotions underlie the whole of human emotional behaviour. There are four different general theoretical perspectives in psychology about how to define and explain emotions. These perspectives are Darwinian, Jamesian, Cognitive and Social constructive perspectives. After studying the relationship between emotions and facial expressions and bodily movements.

2.2 Face Detection

Given an image, detecting the presence of a human face is a complex task due to the possible variations of the face. The different sizes, angles and poses a human face might have within the image can cause this variation. The emotions which are deducible from the human face and different imaging conditions such as illumination and occlusions also affect facial appearances. The approaches of the past few decades in face detection can be broadly classified in to four sections: knowledge-based approach, feature invariant approach, template based approach and appearance-based approach.

2.3 Knowledge-Based Approach

Knowledge-based approach is based on rules derived from the knowledge on the face geometry. A typical face used in this approach is shown in figure 2.1. The most common way of defining the rules is by the relative distances and positions of facial features. By applying these rules faces are detected, then a verification process is used to trim the incorrect detections. Translating knowledge about the face geometry into effective rules is one difficulty faced in this approach, since strict rules may fail to detect faces but if the rules are too general it can increase incorrect detections. This approach is too limited since extending this to detect faces in all cases is impossible.

2.4 Feature Invariant Approach

In feature invariant approach, facial features are detected and then grouped according to the geometry of the face. Selecting a set of appropriate features is very crucial. This approach is not suitable for images with noise, illuminations and occlusions since they can weaken the feature boundaries. The main drawback of template-based approaches is the sensitivity to rotation and scaling. Since feature-based approach is not affected by this sensitivity, it provides a better solution to facial detection problem. A face model which is defined in terms of features or a face texture model which is defined in terms of a set of inequalities can be used for face detection in the feature invariant approach. Recently human skin color has caught the attention of the researchers as a significant feature, since skin color reside in a small color range in different color spaces regardless of the race. Thus contemporary studies on various skin color detection techniques can be found. Approaches which combine multiple facial features have also being proposed.

2.5 Template-Based Approach

A standard pattern of a human face is used as the base in the template-based approach. The pixels within an image window are compared with the standard pattern to detect the presence of a human face within that window. Though the approach is simple, the scaling required for the image and template is a drawback. Besides, this approach is incapable of dealing with the variations of the human face. A predefined template based and deformable template based approaches are the two classes of the template-based approach

2.6 Appearance-Based Approach

Appearance-based approach considers the human face in terms of a pattern of pixel intensities. Since non face patterns are not used in the training process of this approach it is not robust enough. Even the time taken is lengthy, as the number of patterns which needs to be tested is large. A neural network is the commonly used solution for capturing complex facial patterns from facial images. Both supervised and unsupervised learning approaches have being used to train the network. Since finding a sufficient training data set is questionable, unsupervised neural networks are more preferable. Apart from neural networks, Support Vector Machines (SVM), eigenfaces, Distribution based approaches, Nave Bayes classifiers, Hidden Markov Models (HMM) and Information theoretical approaches can also be used for face detection in the appearance-based approach. Rather than minimizing the training error as in neural networks, SVM operate by minimizing the upper bound on the generalization error. Eigenfaces is a probabilistic visual learning method which uses eigenspace decomposition. Nave Bayes classifier provides a better assessment of the conditional density functions in facial sub-regions. The HMM does not require exact alignment as in template-based and appearance-based approaches. HMM usually see a face pattern as a sequence of observation vectors.

The most significant preparation for emotion recognition techniques is the selection of appropriate learning methods and appropriate feature set as a description of the emotional. In this project, an efficient emotion recognition technique based on the features of human speech and facial using a KNN has been proposed.

3. PROPOSED METHODOLOGY

The proposed technique of emotion recognition is capable of recognizing a person's emotion from his face image and speech. This technique includes several stages, like preprocessing, feature extraction, feature selection and classification as illustrated in Fig

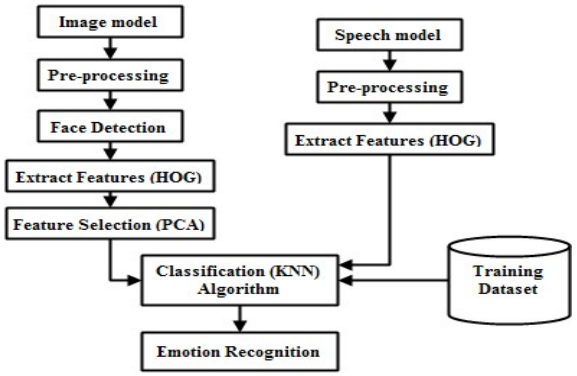

Fig 2: Proposed Emotion Recognition Method

3.1 Image Pre-processing Step

Increasing the quality of image gives a better results in face recognition rate than noisy images but it has a major difficult to extract features from such noisy images. This kind of drawback can be overcome by preprocessing especially in low quality images that can be done before the extraction process. Both the input and output has the lowest level of abstraction by a name called preprocessing that are in common images which has the intensity among all the othe processing image data. An unwanted image with some distortion enhance the improvement of image data that are featured as a future processing which has the main aim of preprocessing image data. The image can be converted into gray color that can be normalized to have uniform intensity due to the color grading process could be done by a stable image. Cropping schemes, resizing are some suitable color grading process that are used in face detection in which the image is cropped and then resized to meet the requirement. Then the image is filtered using low pass filter as shown in Figure 2 the block diagram of image pre-processing.

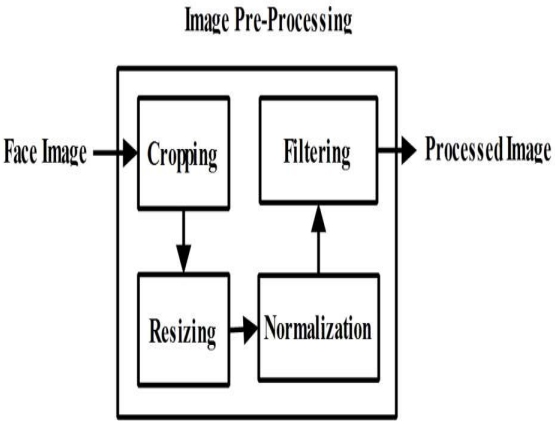

Fig 2: Block diagram of Image Preprocessing

Convert the image from RGB color to grayscale as shown in figure. This process of conversion simplifies the extraction of features from the face image.

$$\text{Grayscale}(i,j) = \big((0.3 * R) + (0.59 * G) + (0.11 * B)\big) \quad (1)$$

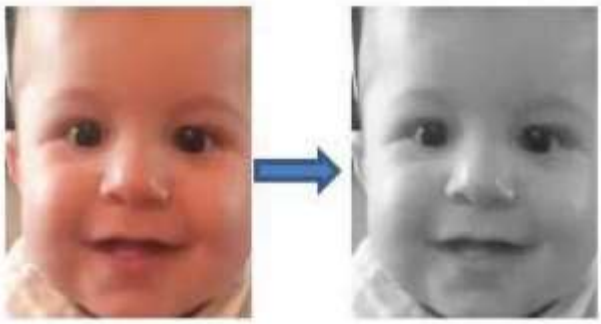

Fig 3: Gray Scale Image

Perform the histogram equalization for improving the quality of face image that has low contrast. It works on changing the dynamic range of the image then some significant features of facial will be more obvious.

The size of the face image is modified getting on by default, for instance, 128 x 128, image size normalization is done. Then high pass filtering is applied for stressing on the significant points of an image. The features that are extracted based on the outlines of facial can benefit from the results that are obtained from the scheme of edge detection. The image is subjected to median filter for removing the noise from images with no loss of valuable information. Normalize the illumination of images that were taken under various lightings for increasing the performance of recognition.

3.2 Face detection

The method of Viola and Joins is utilized for detecting the face in each input image. The main merit of this method is a fast detection process with high accuracy in cropping the regions of the face, nose, and mouth from a detected image. This method works on eliminating the insignificant background and extracting the face. It is based on several points; Firstly, the face pixels are consisted of distinguishing features like hair, eye, eyebrows, mouth, nose, etcetera, with a related structure, here, only noise and mouth are taken. Secondly, pixels of the background are with different features, so they are of diverse kinds. Thirdly, a series of strong cascaded classifiers are formed by the method of Viola and Jones for the fast discarding of unpromising image regions. These strong cascaded classifiers are ascendingly ordered based on the complexity. Depending on this, a considerable number of regions that are unlikely to include faces are removed via the initial classifiers with little impact, but more computations are consumed on candidate regions by the later, more developed classifiers.

Equation (2) explain this method

$$F = \prod_i^n f_i \qquad (2)$$

Where F indicates the stage cascade, fi is the number of the face in the dataset, ⊔ is a round classifier or factorial. The method of Viola and Joins is efficiently increasing the performance of detection and reducing the time of computation. Fig. 6 explains the cascade of stages.

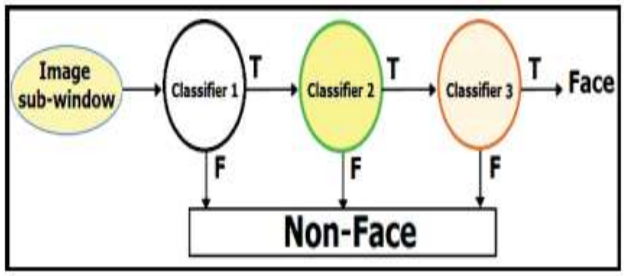

Fig.4: The Cascade of Stages.

3.3 Viola and Jones Based Face Detection

The Viola-Jones method is based on algorithm is to scan a sub window that is capable of detecting faces across a given input image. In some cases the approach is processed in the standard image that can rescale by the given input image at different sizes (Mutneja and Singh, 2019). This method can be processed by different size image as time consumption due to calculation and then run the fixed size detector through these images. The viola and jones is a standard method of face detection as rescale the detector instead of the input image and run the detector many times through the image at each time with a different size. Some of this method has been calculated in size based approach which can consume equal time which requires the same number at first one might suspect in both approaches to have devised a scale invariant detector. This features are constructed using integral image with simple rectangular reminiscent of Haar wavelets.

The main function is to turn the input image into an integral image in a first step of the Viola-Jones face detection. This algorithm is concentrated at left of each pixel to equalize the entire sum of all pixels that is done by make each pixel. In Figure 4 shows the rectangle form using four values that allows for the calculation of the sum of all pixels. These image values are in the integral image that coincides with the pixels in the input image.

Sum of grey rectangle = D - (B + C) + A

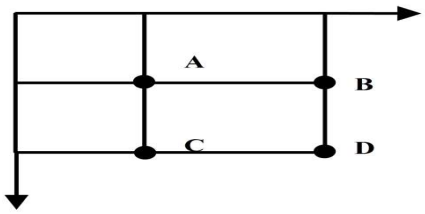

Fig.5 Sum calculation

The Figure shows the different types of features in Viola-Jones face detector analyzes to given sub window using features consisting of two or more rectangles.

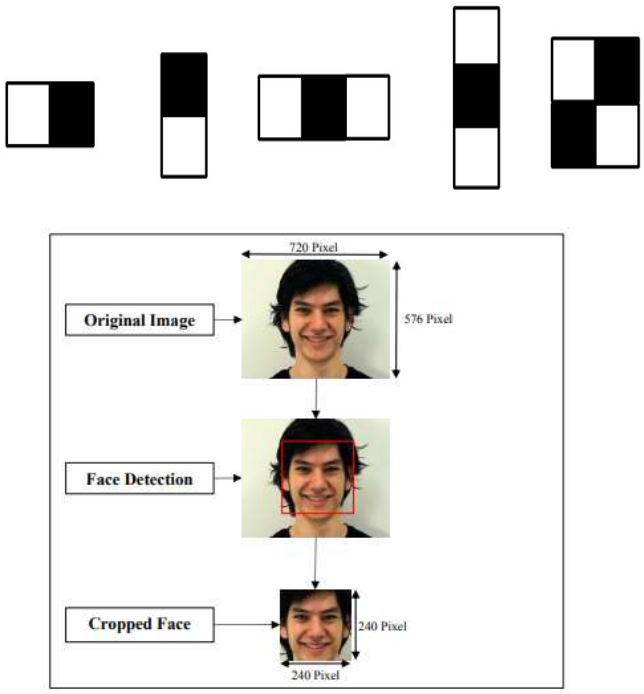

Fig.6 Viola-Jones face detector analyzes

4. CONCLUSION

In this proposed work, a technique for recognizing emotion is presented by using the algorithm of HOG to extract features from (face and speech), and using the KNN as a classification method. The four different basic emotions are smile, no-smile, crying, and laughing from two datasets (images, speech) are used for this experiment. We were able to distinguish the smile, the lack of the smile, through the images that contain the face. In this paper, half of the lower face (nose and mouth) is used only to recognize the emotion, and the obtained accuracy is 94%. This proposed technique of emotion recognition based on half of the lower face is better than other techniques that used the whole face to distinguish the emotion. Also, we were able to recognize the emotion of crying and laughing by speaking after converting the frequency from 44100 to11025, and the obtained accuracy is 86 %. In the future, the proposed system will work on other types of emotions

REFERENCES

1. F. Cavallo, F. Semeraro, L. Fiorini, G. Magyar, P. Sinčak, P. Dario, "Emotion Modelling for Social Robotics Applications: A Review“, Journal of Bionic Engineering, Vol. 15, No. 2, pp. 185–203, 2018.

2. A. A. Hayawi, J. Waleed, "Driver's Drowsiness Monitoring and Alarming Auto-System Based on EOG Signals", 2nd International Conference on Engineering Technology and their Applications 2019- IICET2019- Islamic University, Alnajaf-Iraq.

3. M. S. Hossain and G. Muhammad, "An Emotion Recognition System for Mobile Applications," in IEEE Access, vol. 5, pp. 2281-2287, 2017, doi: 10.1109/ACCESS.2017.2672829.

4. Z. Liu et al., "A facial expression emotion recognition based humanrobot interaction system," in IEEE/CAA Journal of Automatica Sinica,Vol. 4, No. 4, pp. 668-676, 2017.

5. M. B. Akcay, K. Oğuz, "Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers", Speech Communication, Vol. 116, pp. 56- 76, 2020.

6. S. Shojaeilangari, W. Yau, K. Nandakumar, J. Li and E. K. Teoh, "Robust Representation and Recognition of Facial Emotions Using Extreme Sparse Learning," in IEEE Transactions on Image Processing, vol. 24, no. 7, pp. 2140-2152, July 2015, doi: 10.1109/TIP.2015.2416634.