A Survey on Various Techniques of Human Emotion Recognition and its Applications

Anjana C R

Department of Computer Science, Jawaharlal College of Engineering and Technology, Kerala, India

***Abstract*— Humans show a variety of emotions according to the situation and identifying these emotions can be helpful in efficient communication. In the case of human- machine interaction, identifying the human's emotion correctly is very important. In the current situation where machines act as an inevitable part of human life as an instructor, a helper, or even a companion, it would be helpful if the machines could understand the emotions of the human and act accordingly. Human emotions can be recognized from the tone of voice, facial expressions (which may be from image or from video), body language and using EEG (Electroencephalography). EEG is a technique that uses electrodes attached to the scalp to identify the tiny electrical charges that result from the activity of the brain cells for emotion detection [4]. But it is not always possible to wear the electrodes in order to communicate with a machine. So, more focus is given to the other techniques of emotion recognition. Emotion recognition from facial images is done by using image processing techniques. Various types of neural networks can be employed for identifying various expressions from faces in the images. Acoustic data can be collected and used for identifying emotions from sound input. The body language of a person can also convey some information about the emotional state of the individual.**

***Keywords*— Neural Network, Mel Spectrogram, Chromagram, Electroencephalography**

I. INTRODUCTION

Emotion recognition is always important for meaningful communication. Facial emotions are important cues for non-verbal communication. This is why humans can recognize emotions accurately. The basic facial expressions convey discrete and specific emotions: anger, happiness, surprise, fear, disgust, and sadness. The face communicates a lot of information such as the emotional state of the person and their identity.

For facial emotion recognition, we have to identify the face from the given image. The algorithm proposed by Viola and Jones is the most commonly used method for face identification. It used the textures that are common to all faces such as the lighter areas above and below the eye, the edge features and the darker areas to the left and right of the nose to locate the face.

Once the face is detected from the image, the next step is to identify the emotion of the face from the facial expression. This is done by examining the facial features such as position and orientation of the eyes, eyebrows, lips etc.

The fig 1 represents the steps in emotion recognition during the training and test phases.

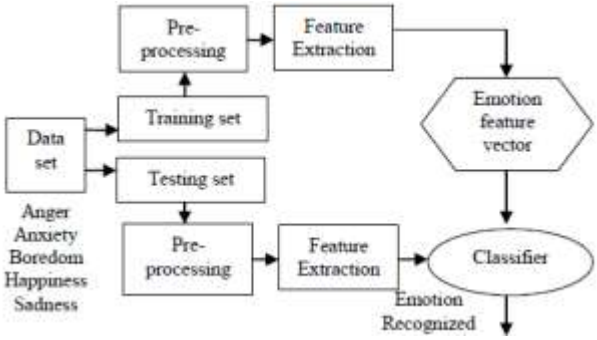

Fig 1. Steps in Training and testing the emotion recognition system.

A. Emotion Recognition from Facial Images

Facial expression can be identified by estimating the displacements in the positions of facial points such as eyebrows, lips, etc. This is the method of geometrical feature extraction. The region based algorithms are categorized in three parts gradient based, correlation based and filtering based. One of the most efficient methods for face detection is the Viola- Jones face detection algorithm. Texture analysis can also be done to identify face from a given image. The eyes, eyebrows and lips are the portions of the face that convey a lot about the expression of the human. Texture analysis is a method used to analyse the texture of these areas in order to understand the expression.

Neural networks are an interconnection of processing elements, each of them doing a specific task. Neural networks are trained with a large set of images with various expressions. The emotion recognition system works in two stages, first, train the neural network with a specific data set and second to give a test image as input to identify the facial expression from the image. Emotion recognition can also have two steps, first to identify the features from the given input image and then use these features to classify the image to one of the various expression classes if a face is detected from the image.

Various classes of neural networks, such as Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN) can be used to implement this model. The emotion recognition model using Convolutional auto-encoders is said to have an accuracy of 92.52% [1].

B. Emotion Recognition from Speech Signals

Speech signals can convey a lot about the emotional state of the human. Emotion recognition from speech has two stages, feature extraction and emotion recognizer. Various parameters of voice are considered for feature extraction such as Mel spectrogram, Harmonic percussive, Chromagram, Mel frequency cepstral, Beat tracking and Beat-synchronous features aggregation.Emotion recognition from speech signals can also have a deep neural network and a classifier such as SVM to classify the emotion from the input speech signal considering the various features of the signal. The emotions and how the parameters change according to each emotion is given in Table I.

TABLE I

DIFFERENCE IN SPEECH FOR EACH EMOTIONAL STATE

	Anger	Happiness	Sadness	Fear	Disgust
Speech Rate	slightly faster	faster or slower	slightly slower	much faster	very much slower
Pitch Average	very much higher	much higher	slightly lower	very much higher	very much lower
Pitch Range	much wider	much wider	slightly narrower	much wider	slightly wider
Intensity	higher	higher	lower	normal	lower
Voice Quality	breathy	blaring	resonant	irregular	grumbled

The following statistics are calculated from the pitch and used in the pitch feature vector: mean, median, variance, maximum, minimum (for the pitch vector and its derivative), average energies of voiced and unvoiced speech and speaking rate.

C. Emotion Recognition from Video

Emotion recognition from video input is basically done by fusing multiple features. Face images sequences and audio data are taken from the input video. The dynamic textures and geometric features are extracted from the images. These are fused with the acoustic features and fed to multiple kernel SVMs to obtain the results.

The use of video for emotion recognition is in the fact that facial expressions along with its corresponding audio modalities convey the correct emotions. Thus, if the system is fast enough, then video emotion recognition is the most useful in case of applications such as human-computer interaction, lie detection etc. The fusion method uses multiple kernels, i.e. acoustic kernel, geometric kernel and appearance kernel for each of its corresponding features. The emotion recognition from video is more dynamic and thus needs to be fast and efficient, since the emotions may be dynamic in the video and it needs to be recognized simultaneously for accuracy. There are many methods proposed for video emotion recognition that includes the use of a variety of features and SVMs. The hybrid feature method discussed in [3] is said to have an accuracy of 95.7%.

II. Conclusions

Emotion recognition finds its most advanced application in real time scenarios where human emotions must be recognized in real time. Emotion recognition is used in applications where machines acts as an instructor, helper and companion, for lie detection, for music players and similar applications based on human emotions etc.Emotion recognition from images is highly efficient but its application in real time is limited. In the case of real time emotion recognition, the video based model is required. Speech based emotion recognition is more useful in simpler applications such as emotion based music player, lie detection etc. Other methods such as EEG is highly accurate but the brain wave monitoring equipment is costly and is not widely used.

Acknowledgement

I express my profound gratitude to my guide for providing me with unfailing support and encouragement.

References

[1]. Ariel Ruiz-Garcia, Mark Elshaw, Abdulrahman Altahhan, Vasile Palade, (2017). Stacked Deep Convolutional Auto-Encoders for Emotion Recognition from Facial Expressions. IEEE.
[2]. Junkai Chen, Zenghai Chen, Zheru Chi. (2016), Facial Expression Recognition in Video with Multiple Feature Fusion. IEEE and Hong Fu.
[3]. K.Tarunika, R.B Pradeeba, P.Aruna. (2018),Applying Machine Learning Techniques for Speech Emotion Recognition. IEEE.
[4]. SnehaLukose, Savitha S. Upadhya. (2017), Music Player Based on Emotion Recognition of voice signals. International Conference on Intelligent Computing,Instrumentation and Control Technologies (ICICICT).
[5]. www.jrmc.org/services/other-services/imaging-services/