INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume 4, Issue 2)
Available online at: www.ijariit.com

Emotion recognition using face and voice analysis

Rohan Raphy Thattil
rrttrills@zoho.com
Sahrdaya College of Engineering and Technology, Kodakara, Kerala

Joseph Gigo Ignatiaus
razorinstinct@gmail.com
Sahrdaya College of Engineering and Technology, Kodakara, Kerala

Shafeer P. N
shafeerpn77@gmail.com
Sahrdaya College of Engineering and Technology, Kodakara, Kerala

Shyam Krishna M
shyamkrishna@sahrdaya.ac.in
Sahrdaya College of Engineering and Technology, Kodakara, Kerala

ABSTRACT

The goal of this project is to design a system that reads the face and voice of a person in conjunction to detect the sentimental and emotional state of a person based on that data. Humans are said to express almost 50% of what they want to convey in nonverbal cues. This concept can be used to analyze on the tone and facial expressions both of them nonverbal cues that can be used to detect the sentimental and emotional state of a person based on his facial expressions and speech features. Here make use preexisting data bases, classify them according to their emotions and create classifiers for the purpose of emotion recognition of new input data. There are many data sets available from previous surveys. The database we use in here is Extended Cohn Kanade database for the facial expressions and SAVEE database for speech data. We then make use of the SVM classifier to classify the respective emotional labels appropriate for each of the image and then create an SVM classifier which is able to classify the input given to it. The emotional state of the face and the voice of the user is found out and then we read them both in conjunction to get a more accurate representation of the user's emotional state. We then play an appropriate music depending on the emotion of the user.

Keywords: *CK+ Dataset, Emotion Recognition, Feature Extraction, MFCC, SAVEE Dataset, SVM.*

1. INTRODUCTION

The purpose of this project is to create an effective solution for the purpose of detection emotions and recognition of emotional states of an individual from the analysis of face and voice of the person. Emotion recognition is a field of computer science that comes under artificial intelligence and machine learning. These days' intelligent computers are able to recognize the emotional state from the face of a person. Emotion detection helps computers and humans to come closer to each other as recognizing the emotions of another person and responding to it as such is one of the characteristics that makes us human beings. Similarly, if computers are made to learn emotions and even understand the emotional thought process that a person is undergoing we might be able to develop intelligent machines that will be able to interact with us in a 'human' way. However such machines are often the stuff of speculative fiction and as such we are still far away from creating such machines.

In this project, we aim to create an effective solution for a computer to recognize the emotional state a person is undergoing from the analysis of his face and voice in which he speaks. It is said that humans express about 50% of what they want to communicate in their body language. Their tone, facial expressions, pitch, etc... Can all be used to recognize the emotional state a person is undergoing.In fact, such methods of body language reading methodology are used by interviewers and councollers for various purposes. These include interrogating a person for information, for interviewing prospective employees or clients, for understanding a patient and so on.

Here we capture the voice and face of a person using a camera and a microphone and sent it to the computer for analyzing. The computer analyses these two components separately to recognise what they are expressing individually. After that, the computer combines these two to understand what the emotional state the person is going through after analyzing them both in conjunction.

This way we can employ a far more effective method to find the emotional state a person is undergoing under the circumstance he is in.

We first, however, train an appropriate SVM classifier that is capable of recognizing the emotional state of the user by analyzing both his face and voice. This SVM classifier is trained with the help of both the Extended Cohn Kanade Database and the Savee Database. The Extended Cohn Kanade Database contains the transition of a test subject to 6 different facial expressions from an initial neutral facial expression. There are over 500 test subjects in this database. SAVEE database has the voices of a 6 different actors speaking about in 7 different emotional states, each emotional state containing around 60 different sentences. Both the databases are appropriately labeled and send to an SVM trainer and then the SVM classifier is saved to which the test cases, as well as the user inputs, are sent. The classifier differentiates between various types of facial expressions and voice tones and then it reads both in conjunction to produce a more accurate representation of the user's emotional state. Then we play an appropriate song depending on the emotional state of the user.

2. EXISTING SYSTEM

Most of the available systems merely use one of the following-face or voice-to understand the emotion of a person. Such systems usually don't take into consideration that it's still possible for humans to fake their facial expressions to a degree and ignore other body language cues that a giveaway his emotions. There's also the limitation that those machines are designed to just recognize basic emotions such as happy, sad, etc... As they only take into factor one of the above cues. They use raw image processing to understand the expressions of the person on his face, not his emotional state. As such there are many limitations to this approach as I already said it's rather easy to fake or cheat such an approach. Another disadvantage is that such devices are rather inflexible because such devices are designed based on a preexisting fundamental. It doesn't take into consideration a person's cultural aspects that might change the way he behaves and as a result is impractical to sample on a scale as a result.

As a result, there is a need to improve these existing systems to take into consideration other body language cues and to improve upon previous knowledge.

3. TECHNOLOGY AND CONCEPT USED

A. Modules Used

i. Support Vector Machines

Support Vector Machines often shortened as SVMs have supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Consider a set of training examples, each is given a particular label, an SVM training algorithm builds a classification model that assigns each new example of a particular category a label that it can be associated with based on the what it learned from the training examples. SVM classifiers are commonly considered as a non-probabilistic binary classifier. An SVM model is a representation of the examples as points in space, marked so that the examples of each category are divided by a clear gap that is as wide as possible. New examples are then labeled and mapped into that same space and predicted to belong to a particular category based on what the classifier learned during the training phase.

Usually SVMs are used only for the purpose of linear classification, however, it's also possible to train a classifier for nonlinear classification using various 'kernel tricks'.Here we make use of the principle of One vs All classifier for classifying 6 different labels of facial expression and 7 different labels for tonal classification. The basic principle is to train the classifier to label a particular kind of label and mark all others as not belonging to it for each case. After that use the classification models in series with the help of a loop statement.

Various tools are available which implements SVM algorithm for the purpose of classification and regression analysis. Some of them are LibSVM, Gaia, OpenCV, etc… Here we make use of the SVM trainer and classifier algorithms provided by the OpenCV library.

ii. Landmarks Detector

A landmark point or a landmark is a point in a shape or an object in which correspondences between and within the populations of the object are preserved. It's also called vertices, anchor points, control points, sites, profile points, sampling points, nodes, markers, fiducial markers, etc. Landmarks can be defined either manually by experts or automatically by a computer program. Here we make use of a tool called landmarks extractor to obtain facial landmark points that are then sent to the SVM training program/classifier program.

Many tools and libraries contain algorithms for landmark detector training and classification. Here we make use of a library called Dlibs library and make use of its inbuilt pre trained facial landmark extractor to obtain 68 different facial landmark points (Figure 1).

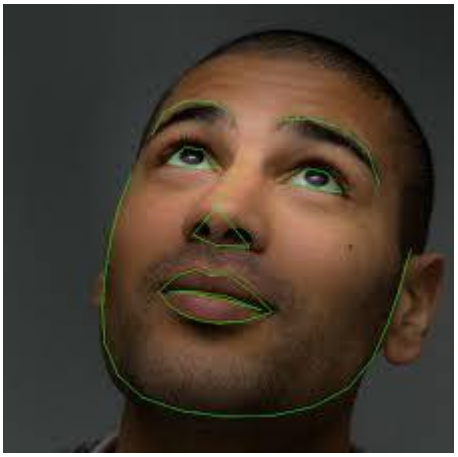

Figure 1 Facial Landmark Points

iii. Extended Cohn Kanade Database

Extended Cohn Kanade Database also is known as CK+ is a dataset of around 486 different sequences of 97 different posers transitioning from an initial neutral expression to 6 or 7 different facial expressions. This dataset is prepared for the purpose of facial emotion classification training and testing.

iv. Surrey Audio Visual Expressed Emotion Database

Surrey Audio Visual Expressed Emotion Database also known as SAVEE database contains 480 different utterances of English sentences in British accent ini 7 different emotions using the recorded voice of 4 different actors. This dataset is prepared for the purpose of automatic speech emotion recognition.

B. Concept

i. Facial Expression Recognition System

Here we classify the CK+ dataset into 6 different classes-neutral, happy, anger, sad, surprise, disgust and contempt. We then label each of these images and associate with them their associated label. We then obtain the facial landmark points for each of these images and represent them in a matrix form for each image. We then normalize these matrixes. We then send these labeled matrixes for the training examples to the one vs all SVM classifier. We use 80% of the dataset for training and 20% for testing. For each phase in which the SVM classifier is trained for a particular emotion, the cases in which the expression comes is labeled as positive and the rest are labeled as negative. We then save this 6 class SVM classifier model. In case we need to classify any new examples we make use of the landmark extractor algorithm to obtain its facial landmark points and then represent them in a matrix format and normalizes this matrix. We then sent this matrix to the 7 class SVM classifier which then associates an appropriate label for that class depending on the expression that face is trying to convey. The diagramatic representation of the facial expression recognition system is given as the below figure 2.

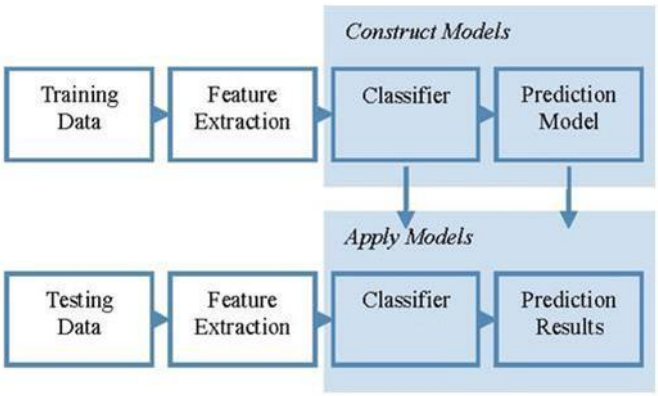

Figure 2 Facial Expression Recognition System

ii. Speech Emotion Recognition

Now we classify the SAVEE dataset into 7 different classes-neutral, happy, anger, sad, surprise, fear and disgust. We then label each of these audio signals and associate with them their respective labels. We then obtain the MFCCs for each of these audio signals and represent them in a matrix form. Each audio signal is represented with the help of 13X3 matrix. We then send these labeled matrixes for the training examples to the one vs all SVM classifier. We use 80% of the dataset for training and 20% for testing. For each phase in which the SVM classifier is trained for a particular emotion, the cases in which the expression comes is labeled as positive and the rest are labeled as negative. We then save this 6 cl7ss SVM classifier model. In case we need to classify any new examples we make use of the MFCC extraction algorithm again to obtain 13 different low, mid and high MFCCs and then represent them in a matrix format. We then sent this matrix to the saved 7 class SVM classifier which then associates an appropriate

label for that class depending on the emotion that the recorded voice is trying to convey is trying to convey. The diagramatic representation of the speech emotion recognition system is given as the below figure 3.

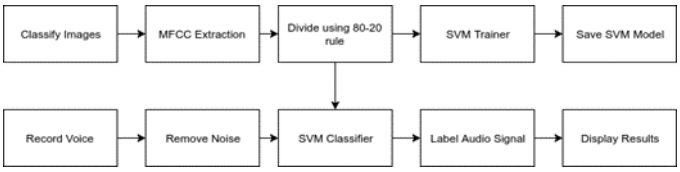

Figure 3 Speech Emotion Recognition System

iii. Combining Both

First, we try to classify them both face and tone individually. Then we take both in conjunction and then create 49 different combinations of different facial and tonal states. For each case, we make use of preexisting body language studies to find out what the user is really expressing. For eg- a nuetral face along with a stressed voice can indicate that the person is holding some grudge. This way we find the appropriate body language cue for 49 different combinations containing voice tones and facial expressions and then play out an appropriate music for each of this case. We make use of the Python Tkinter library to implement a GUI and make use of the PyAudio library to play the music.

4. EXPERIMENTS AND RESULTS

The given datasets are made use of to create a facial expression and speech emotion SVM classifiers. The face of a new user is captured with the help of a camera and his facial expression is recognized. The voice of the user is captured using the recorder and noise signals are removed and then it is sent to the SVM classifier for speech emotion recognition. We then combine them both and form 49 different combinations. The sentiment expressed by the user is recognized by using body language studies to understand each of these 49 different combinations and the sentiment is informed to the user. We then play an appropriate music that is fitting depending on the sentiment the user is expressing. We have tried testing the system with 10 different cases using 4 subjects and found out the accuracy to be around 60%. We have found the main challenge is with the non-British origin of the test subjects since the voice dataset we used is based on British actors' voices. Hence we had to ask the test subjects to emulate the British accent when testing out voice to improve the reading of speech emotion recognition system.

5. CONCLUSION

We have hence created a system that is capable of reading both the facial expression and speech emotion state and found out a method in which both can be read in conjunction to create a more effective method to find out the emotional state of the user and then play out a music for each state. This system can be employed in many applications such as interviews, surveys, etc… This can even further improve by integrating more dialects and voices and even more faces to get a more accurate reading.

6. REFERENCES

[1] Rajesh KM & Naveen Kumar M, A Robust Method for Face Recognition and Face Emotion Detection System using Support Vector Machines, 2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques, Pg no.2 .2017

[2] Mehdi Abdollahpour,lafar Zamani & Hamidreza Saligheh Rad, a Feature representation for speech emotion Recognition, 25th Iranian Conference on Electrical Engineering, Pgno 4, 2017

[3] Nancy Speech Emotion Semwal, Abhijeet Detection Kumar System using & Sakthivel Narayanan, Automatic Multi-domain Acoustic Feature Selection and Classification Models,Bhabha Atomic Research Centre Conference, Pg No 4, 2017

[4] Surjya Ghosh, Niloy Ganguly, Bivas Mitra, Pradipta De, Towards Designing an Intelligent Experience Sampling Method for Emotion Detection, 14th IEEE Annual Consumer Communications & Networking Conference, Pg No.1,2017

[5] Arnab Bag and Md. Aftabuddin, A Review on Emotion Recognition using Speech, Saikat Basu, Jaybrata Chakraborty, International Conference on Inventive Communication and Computational Technologies, Pg No 5,2017

