

Pioneering Method to Analyse Depression in Human Being Using Audio-Video Parameters

Shamla Mantri#1 , Deepali Sanap#2 ,Vimal Dongre#3
Information Technology Dept.,Pune University
MIT College Of Engineering, Paud Road,Kothrud,Pune-411038
deepalissanap@gmail.com
vimaldongare980@gmail.com

Abstract---In medical science there are different health disorders have high economical cost. There are many diseases present in medical science,mental disorder is the one of the most powerful disease which can have large impact on society. We are going to implement the innovative approach for sensing the human emotion by using automatic facial tracking in video, measuring facial activity, recognition of facial expression and by using the frequency of audio signal.

The main goal is to develop and diagnostic mental disorder using affective sensing technique. The sound and visual communication complement each other, by considering this hypothesis for mental disorder analysis. Using multimodal approach for mental disorder diagnosis we are using combination of audio and video fusion method. A different audio and video features are generated separately. Frame level used to classify and combined video and audio features by using support vector machine (SVM).The innovative techniques show the proposed framework's effectiveness in mental disorder analysis.

Keywords- human audio and video detection, human emotion detection.

I. INTRODUCTION

The modern applications are build very fastly in day to day life, the people using computer application for social, professional and personal used.To detect emotion more naturally by the computer system and convert it into electronic signal to find the mental condition of human.the proposed work should able to analysed the mental disorder based on obtain featuers of face and speech,both this features are depend on each other to analyse the result[1].

Multimodal framework having effective sencing technique ,which work on real world example for depression monitoring .To detect the mental disorder multiple models gives the better result than that of single model.The portion of human body is used to analyse the mental disorder by using the area up form the shoulder that is the head movement is capture from the video frame and the audio frequency from per frame is calculate and detecting the disorder [2].

The result of emotion can be depend on many facial expression given by human because the human can be happy and scared at the same time.In real time application is used by machine –human interaction. Human computer interaction generating the mood or emotion of human[3].

In this work , we are using k-Means and Support Vector machines to classify the opposite emotions.Reasearchers are debating which features are require or recognition emotion using speech to detect correct emotion[4].

II. LITERATURE SURVEY

In this paper, authors outline the approach he developed to construct an emotion-recognizing system. The nature of emotion in its interaction with attention is depends on psychological the fusion of different features of audio and video handle by neural network architecture. The neural network architecture are gives the result and their implications discusses for future direction of the research[5].

The second paper describes the audio-video fusion for depression diagnosis progress to date and proposes a novel multimodal framework. The author exploit the proposition that the auditory and visual human communication complement each other[6]. We investigate hypothesis for depression analysis. For video feature data analysis, intra-facial muscle movements, head movements and shoulder movements are analyzed by computing temporal interest points of human. The various audio features like fundamental frequency (f0), intensity, loudness and MFCC are computed. In this work, we compare fusion methods at feature level, decision level and score level. Researchers do study on an age and gender matched clinical dataset of 10-30 patients and 10-30 healthy controls. This work shows results from the multimodal experiments show the framework's effectiveness in depression analysis [7].

III. FUSION USING AGGREGATION MAETHOD

The study of combining two or more result of audio and video extraction the fusion method is used.the multimodel facial expression recognition method is used to fuse number of information.the number of results can be fuse at differente levels such as score,decision and feature level.multiple features which is extarcted by biometric can be fusion using feature level.matching score obtain by biometrics can be combined by score level.different classifiers gives a the different decision and the final result decision is based on all of the decision classifiers. We are using OR Gate to combine the results of the audio and video feature extraction.the two decision make a score bulid by classifier for two sets of features.the fusion methos represented logically[8].

The OR fusion Method represented as.

$$F=V|A$$

F-Fusion of final result

V-video result

A-Audio result

V	A	F
0	0	0
0	1	1
1	0	1
1	1	1

Figure 1 Aggregation table

IV. SUPPORT VECTOR MACHINE

SVM is basic ideas which comes from learning theory of statistics. SVM is use as linear learning module .Linear learning module require. SVM is the kernel level method which is used to create valid kernel functions.SVM is used in feature extraction process. To solving quadratic optimization problem and regression and classification is used for training SVM.SVM supports efficient and simple machine learning algorithm and it is mostly used for classification and pattern recognition problem[9].

A. Classification

Feature vector is constructed using the above mentioned features which were fed to ML algorithms for recognition. MLP is used for gender recognition and Ad boost with C4.5 is used for emotion detection. In this paper we are using analysis techniques to detect mental and psychological problems [10].

V. FEATURE EXTRACTION

In this work, there are many common features is extracted such as speech and energy rate, spectrum features and pitch rate. For example Mel-Frequency cepstrum coefficients (MFCC).

A. Pitch Related Frequency

One of the features of pitch is speech emotion recognition. Fundamental frequency F0 is vibration of vocal. Pitch is also called as glottal-wave; it is information about emotions. The signal from the pitch is depend on vocal folds for tension and air pressure depend on sub glottal and the variation in pitch value describe the various emotional statuses[11].

B. Mel-Frequency cepstrum coefficients

MFCC is widely used feature extraction method for audio frequency, it has anti-noise and other good ability to extract voice signals.MFCC has a good frequency resolution in the low frequencies and it is efficient to Noisy signals. It does not give the satisfactory results over high frequency coefficients [12].

C. Video Features

Visual feature is calculated by using frames, each frame the face feature are detected like eyes, nose, forehead is detected and analysis the movements of these feature from per frame. In this proposed work the drawbacks of previously researched work we diploid preprocessing and featured extraction of video features and audio frequency which work on multimodal technique [13].

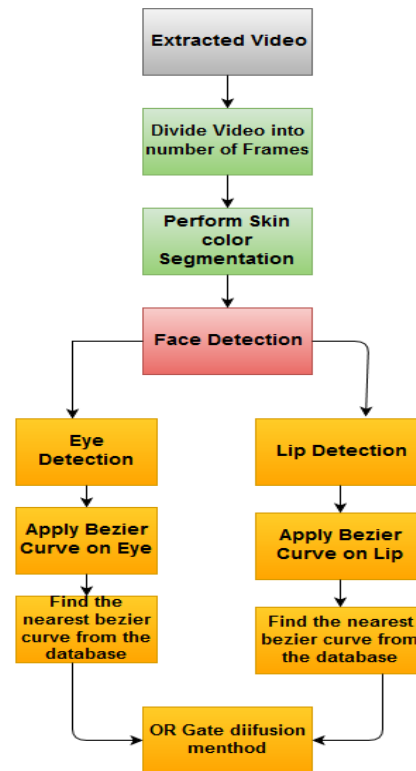


Figure 2 Video Extraction

D. Audio Features

In psychological study the investigators found the four different feature of audio – MFCC, Fundamental Frequency (f0), Loudness and Intensity. The speech frequency is divided into number of frames. The size of frame is set to particular mile-second and the overlapping of frames is avoided for better result [14].

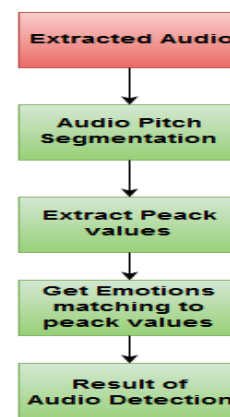


Figure 3 Audio Extraction

- 1) *Feature extraction*: F0 is extracted for gender recognition and pitch, energy and speaking rate are extracted for emotion detection. Features were extracted from the voice signals. By correct features are extracted by using proper algorithm technique it will give proper result toward analysis of face detection. All features which are extracted are transfer into input data is called feature extraction.
- 2) *Pre-processing*: Pre-processing has been done for removing the noise present in the collected samples. This includes pre-emphasis and windowing. Pre - emphasis process removes the noise from the captured signal. Spectral subtraction method has been used to remove the noise. Further hamming window is used for windowing. [15].

E. Linear And Non-Linear Features

Studies are mainly concerned with health parameters which mainly responsible for analysis depression and avoid many succeed cases which are increases these day. Audio features are performed analysis on related pattern of loudness and variation in audio frequency [15].

VI. CONCLUSION

In this work we are trying to analyze the emotional condition of human being. Because in today's fast world nobody is in fine or good mood every person is depress and had some mental illness because of many reasons. To help people and society this is the small contribution by this paper. This paper many describes about the video and audio features to get the emotion of human being. The main aim behind this to help psychiatrist and doctor's who are working in this area of medical science.

ACKNOWLEDGMENT

We wish to express our sincere thanks to the guide Prof. Shamla Mantri and Head of Department, Prof. (Dr.) A.S. Hiwale as well as our principal Prof.(Dr.) M.S.Nagmode, also Grateful thanks to our Coordinator Prof. Neha Sathe and last but not least, the departmental staff members for their support.

REFERENCES

- [1] Prof. Shamla Mantri, Dr. Pankaj Agrawal, Prof. Dipti Patil, Dr. V. M. Wadhai," Depression Analysis using ECG Signal", ISSN 2277-3061, NOVEMBER 2013.
- [2] Shamla Mantri, Vidya Dukare, Smita Yeole, Dipti Patil,V. M. Wadhai,"A Survey: Fundamental of EEG", International Journal of Advance Research in Computer Science and Management Studies, Volume 1, Issue 4, September 2013.
- [3] P. M. Chavan, Manan C .Jadhav, Jinal B. Mashruwala, Aditi K. Nehete, Pooja A. Panjari, "Real Time Emotion Recognition through Facial Expressions for Desktop Devices", Volume-1, Issue-7, May 2013.
- [4] F.Yu, E.Chang, Y.Xu, H.Shum, "Emotion detection from speech to enrich multimedia content", *Lecture Notes In Computer Science*, Vol.2195, 550-557, 2001.
- [5] Jyothi. K 1, Prabhakar C.J 2," MULTI MODAL FACE RECOGNITION USING BLOCK BASED CURVELET FEATURES", Vol.4, No.2, April 2014.

- [6] Alexandros Karatzoglou, David Meyer,Kurt Hornik," Journal of Statistical Software", April 2006, Volume 15, Issue 9.
- [7] Yixiong Pan, Peipei Shen and Liping Shen," Speech Emotion Recognition Using Support Vector Machine", International Journal of Smart Home Vol. 6, No. 2, April, 2012.
- [8] DURGESH K. SRIVASTAVA, LEKHA BHAMBHU, " DATA CLASSIFICATION USING SUPPORT VECTOR MACHINE", © 2005 - 2009 JATIT.
- [9] Shamla Mantri, Dr. Pankaj Agrawal, Dr. Dipti Patil, Dr. Vijay Wadhai," Non Invasive EEG Signal Processing Framework for Real Time Depression Analysis", *SAI Intelligent Systems Conference 2015 November 10-11, 2015 | London, UK*.
- [10] Santaji Ghorpade, Jayshree Ghorpade, Shamla Mantri,"PATTERN RECOGNITION USING NEURAL NETWORKS", International Journal of Computer Science & Information Technology (IJCSIT), Vol 2, No 6, December 2010.
- [11] Dipti Patil, Shamla Mantri, Ria Agrawal, Shraddha Bhattad, Ankit Padiya, Rakshit Rathi," A Survey: Pre-processing and Feature Extraction Techniques for Depression Analysis Using Speech Signal", International Journal of Computer Science Trends and Technology (IJCSST) – Volume 2 Issue 2, Mar-Apr 2014.
- [12] Shamla Mantri, Dr. Pankaj Agrawal, Dr. S.S.Dorle , Dipti Patil , Dr. V.M.Wadhai" Clinical Depression analysis Using Speech Features", 2013 Sixth International Conference on Emerging Trends in Engineering and Technolog..
- [13] Shamla Mantri, Dr. Pankaj Agrawal, Dr. Dipti Patil, Dr. Vijay Wadhai," Cumulative Video Analysis Based Smart Framework for Detection of Depression Disorders", 2015 International Conference on Pervasive Computing (ICPC).
- [14] N. Fragoanagos, J.G. Taylor Department of Mathematics, King's College, Strand, "Emotion recognition in human-computer interaction." London WC2 R2LS, UK , Received 19 March 2005; accepted 23 March 2005.
- [15] Jyoti Joshi, Roland Gecko, Albina Dhall, Sharifa Alghowinem, Michael Wagner, JulienEpps, Gordon Parker, Michael Breakspea," Multimodal Assistive Technologies for Depression Diagnosis and Monitoring.", University of Canberra, Australian National University, University of New South Wales.