International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 4 Issue: 4 466 - 470

Innovative Approach to Detect Mental Disorder Using Multimodal Technique

Shamla Mantri#1, Poonam Sarnikar# , Aishwarya Bhosale#3 , Dr. Dipti Patil#4 , Dr. V. M. Wadhai#5

Information Technology Dept.,Pune University
MIT Collage Of Engineeing, Paud Road,Kothrud,Pune-411038
poonam.sarnikar@gmail.com, bhosaleaishwarya530@gmail.com

***Abstract*—**The human can display their emotions through facial expressions. To achieve more effective human- computer interaction, the emotion recognize from human face could prove to be an invaluable tool. In this work the automatic facial recognition system is described with the help of video. The main aim is to focus on detecting the human face from the video and classify the emotions on the basis of facial features .There have been extensive studies of human facial expressions. These facial expressions are representing happiness, sadness, anger, fear, surprise and disgust. It including preliterate ones, and found much commonality in the expression and recognition of emotions on the face.

Emotion detection from speech has many important applications. In human-computer based systems, emotion recognition systems provide users with improved services as per their emotions criteria. It is quite limited on body of work on detecting emotion in speech. The developers are still debating what features effect the emotion identification in speech. There is no particularity for the best algorithm for classifying emotion, and which emotions to class together.

Keywords*—human- computer interaction, human emotion, facial expression*

*****

I . INTRODUCTION

Deprssion is a disorder which affect on humans life functions.There are varieties of features that can be extracted from human speech. We use statistics graph relating to the pitch, Mel Frequency Campestral Coefficient (MFCCs) and Formats of speech as inputs to classification algorithms[1]. The emotion recognition accuracy allows us to carry the most emotional information from the features.Using these methods we achieve high emotion recognition accurately. In this paper we use k-means and Support Vector Machines (SVMs) to classify emotions. There are various phases of emotion capturing such as preprocessing, feature extraction, face detection etc.Preprocessing is nothing but removing unwanted signal from speech signal. Extraction means extract only the necessary data which are useful for computing the result [2].

Author presented a system to determine the emotions with the help of facial expressions,displayed in live video streams and video sequences [3]. The system is based on the Piece wise Baez ire Volume Deformation tracker and has been extended with face detector to initially capture the human face automatically. Author also used Naive Bays and the Tree Augmented Naive Bays (TAN) classier in person dependent and person independent tests on the Cohn- Canada database.

Author implements a framework for emotional state classification through still images of pictures and real time feature extraction and emotion analysis application[4].The application automatically detect the face and codes them in a seven different dimensions i.e. neutral, anger ,sad,fear,joy,surprise and disgust. The main aim is to analyses facial expressions.

II. PROPOSED SYSTEM

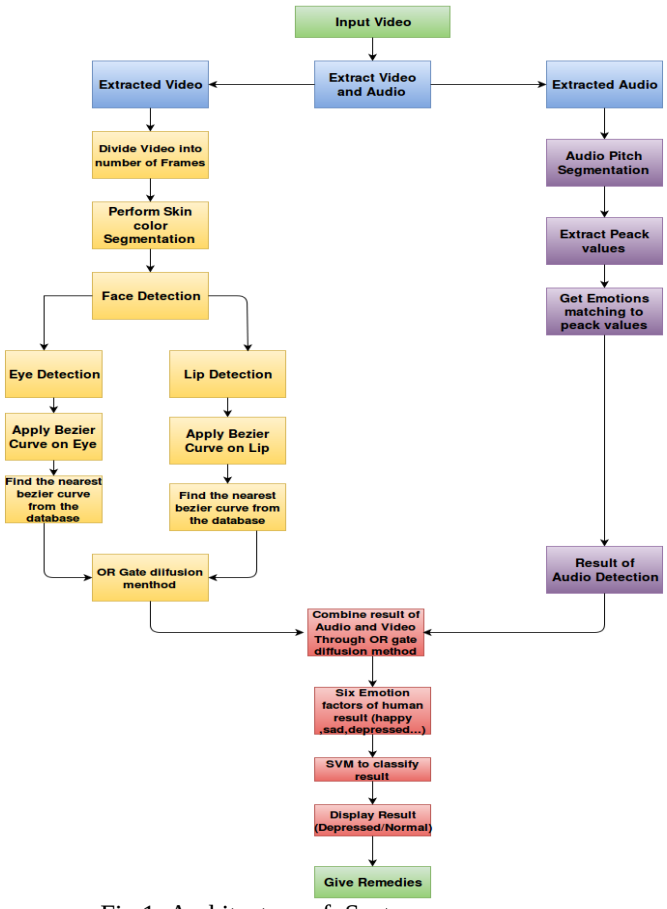

Fig 1 Architecture of System

A. Live Streaming

In fig 1, the live streaming is the important step of image acquisition in real time. In this image frames are obtained using streaming media [5]. In this stage, the application receives images from video camera device. The streaming continues until input image frame is acquired.

B. Frontal Face

This file is used for capture the image and codes with respect to two dimentational in real time like normal and abnormal[5].

C. Skin color Segmentation

Basically Skin color segmentation is used as a differential between the actual skin pixel and non-skin which are white or black pixel. It permits face detection to mainly focus on important areas of image which are used for detecting emotions like lips, eye etc. It mainly explores skin colors.

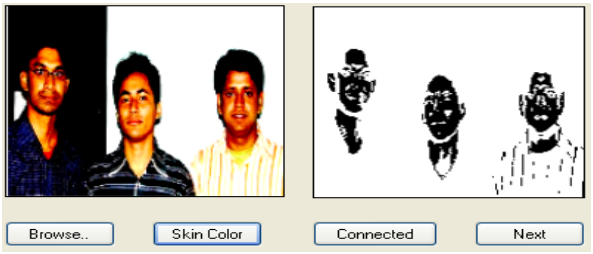

Fig 2 Skin Color Segmentation

In fig 2, for skin color segmentation we first need to contrast the image. Then, we have to find the largest connected region. Then we have to check the possibility to get a face of the largest connected region. In fig 3, if the largest connected region has the probability to become a face, it will open a new window with the largest connected region. If the height & width of connected region is large or equal than 50 and the ratio of height/width is in range of 1 to 2, then it may be face.

Fig 3 Connected Regions

D. Face Detection

It discovers size and location of objects, within an input image.it is detected by some facial features by ignoring all the elements which are not used for detecting the face elements .In fig 4,it is necessary to convert the original image into binary format and scan the whole image for forehead .Color image converting to binary image, we need measure the average value of RGB for each pixel and if the value is small than 110, we replace or covered it by black pixel or we covered it by white pixel. By this method, we get a binary image from RGB image[6].

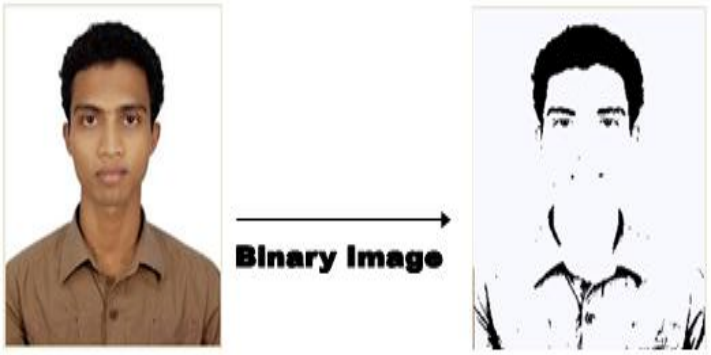

Fig 4 Binary image conversion

After that we find the forehead from binary image which a black and white image. First we start scan from the middle of the image figure, after that we want to find a series of white pixels after a series of black pixel. In fig 5, we want to search the maximum width of white pixel by finding vertically all left and right site. If new width is small than half of the previous large width, after that we cut the scan image because if we near position to the eyebrow then and then only this situation will arise. In next step we cut the face from the beginning position of the forehead of image and its height will be 1.5 multiply of its width of image.

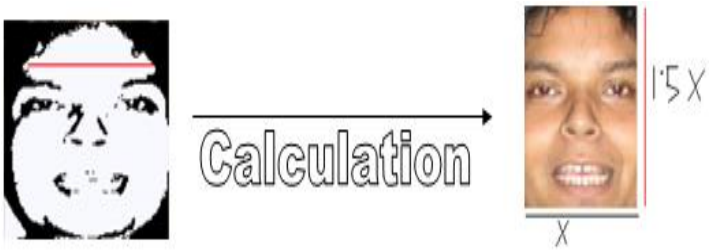

Fig 5 The actual face calculation

In the fig 5, X will be match to the maximum width of the forehead of image. At next we have an image which contain only eyes, nose and lip. Then we cut the RGB image as per to the binary image.

E.Eye Detection

Basically in eye detection first we need to convert RGB (RED-GREEN-BLUE) into the binary form. (GRAY-BLACK).Then we need to scan the image by using (W/4) formula where w is nothing but width of image.

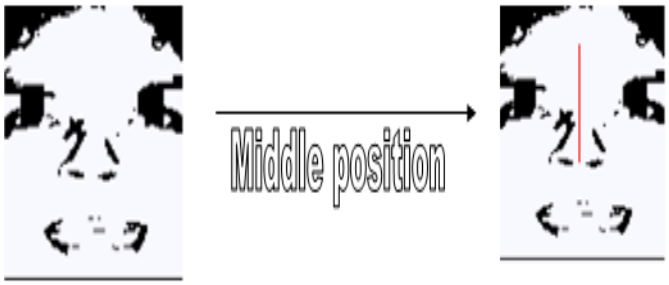

Fig 6 Middle posiotion of eye

In fig 6, we final the middle position of eye with the help of width. First we consider upper position of two eyebrows and black pixel vertically added to connect eyes and eyebrows together. Then we search line of black pixels horizontally from eye's middle position to find out left eye of right side. In fig 7, the left eye is starting from width of particular areas point and right eye indicate the end point of width of particular area [7].

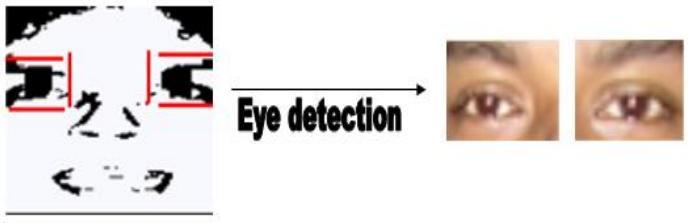

Fig 7 Eye Detection

F. Lip Detection

To detect the emotion the feature of image is to detect the lip emotion which is nothing but LIP DETECTION [8]. The different shape of lip is plain, pout, slightly curve etc. In database we already provide some definite values for lips eye curve. Based on that lip curve is calculated. As shown in fig 8, we determine the lip box for lip detection. We consider that lip need be in inside the lip box specification. So, first of all we need determine the distance between the forehead and eyes part of image. Then we need to add this distance with the measure lower height of the eye part to determine the upper height of box will contain the lip[9]. Now, the start point of box of eye part it will be the ¼ part of the left eye box part and end point will be the ¾ part of the right eye box part. The end height image of the box will be lower end of the persons face image. This box stored lip and may some of the part of the nose part. Then we cut the RGB image according the box image specification [10].

Fig 8 Lip Detection

G. Longest Binary Pattern

Basically longest binary pattern (LBP) is a type of visual descriptor.it is a classifier which is used to classify texture classifiers. It is necessary to convert the skin pixel to the white and remaining to black pixel which are not used for classifications [11]. To find a particular region in image for example lips we deal with the nearest connected region in the database.

H.Bezier Curve Algorithm

Most important applications of Bezier Curve include interpolation, approximation , curve fitting, and object representation. The aim of the algorithm is to find points in the middle of 2 nearby points and to repeat this until we have no more iterations. As shown in fig 9, the Bezier curve is applied to find out the curve of eye[12].

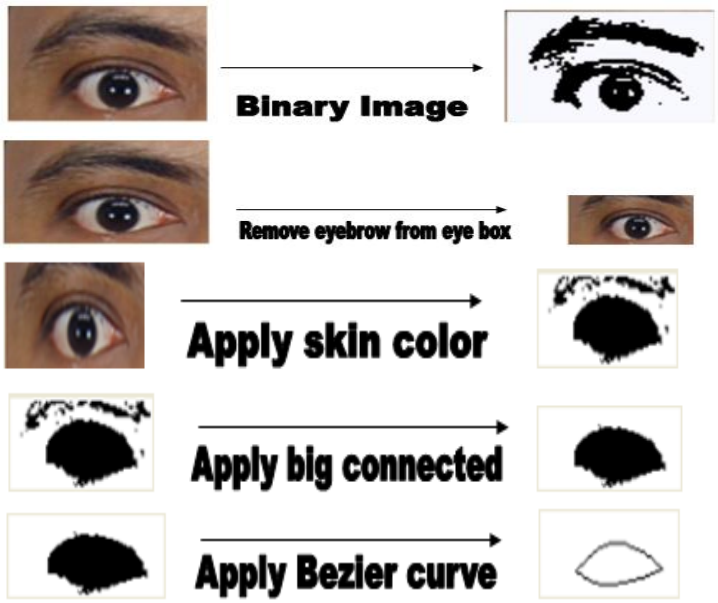

Fig 9 Bezier curve on eye detection

I.Emotion Detection

In this phase we recognize emotions of human. Bezier curve is mapped into larger regions. In fig 10, with the help of Bezier curve algorithm we compare the curve values with values which are already present in database. The most nearest value is picked and displayed as emotions. If result is not

match with database values then average result is calculated, and according to the result decision is made[13].

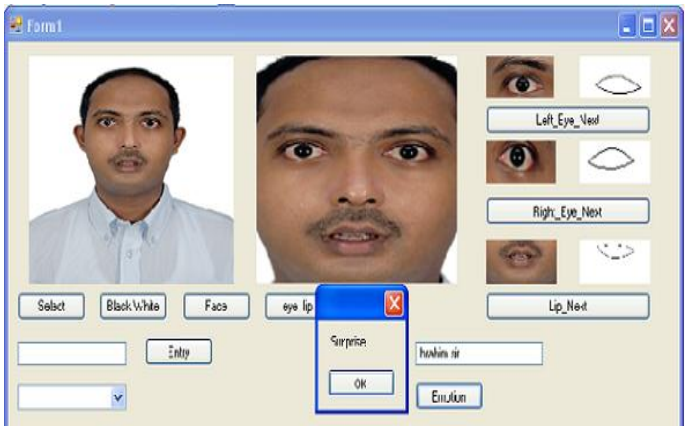

Fig 10 Emotion Detection

J.Database

In database all values are stored which are used for comparing the two nearest matching emotions to treat as a output. The result is displayed with the help of OR gate method. This results treated as a final output[14] .

K.Output Display

After all calculations completed the final output contain graphical and animated graph or figure which describes emotions of human.

L. Feature Extraction

Pitch extract from the waveform of speech using a modified version of the RAPT algorithm for pitch tracking implemented in the toolbox .It uses 50ms, the pitch for each frame was calculated and placed in a vector to correspond to that frame[15]. If the speech is unvoiced the corresponding marker in the pitch vector was set to zero.

M.SVM

Basically SVM is used to divide emotions in two or three emotions that known a priori, classifying a wider spectrum of emotions is a more pragmatic en-devour . The application of an algorithm that can choose only between two known emotions is quite limited. SVMare supervised learning models. When data are not labeled, a supervised learning is not possible, and unsupervised is needed, that would find natural clustering of the data , and map new data to these formed groups. This algorithm which provides a new way to the support vector machines is called support vector clustering[16].

III. IMPLEMENTATION METHODS

Fig 11 Facial Feature Detection and Emotion Recognition

a.Steps for video :

1) Extract video and Audio separately.
2) Divide video into number of frames and maintain an array Image[n]
3) Image[n] = {f1,f2,…..fn}, Where f = frame
4) Perform basic image processing algorithms like image filtering, image transformations and color space conversions.
5) Perform image analysis perform object tracking which gives edges of detected.
6) Detect Eyes and lips
7) Calculate curves of eyes and lips
8) Find the nearest Bezier curse from the data base & apply that data base stored Bezier curve emotion as this image emotion.

b. steps for Audio:

1) Extract video and Audio separately.
2) Perform audio pitch segmentation
3) Perform Classification on the basis of pitch peak values
4) Get emotions matching to extracted values

IV. CONCLUSION

In this Paper we proposed a system which automatically detects human emotions on the basis of facial expressions as well as speech recognition. It shows that building a fast and efficient speech emotion detector is a challenging but achievable goal. By combining a theoretical learning of machine learning along with the Computer Science with classifier optimization, a system can achieve great success in analysis of emotions detection of human being in various normal day to day life situations. Also, it further takes a step to

improve the emotion detection technique using Bezier Curve Algorithm. The system works well for faces with different shapes, skin tones as well audio speech, voice modulations. The key design principles behind this successful implementation of a large real-time system included choosing efficient data structures and algorithms and employing suitable software engineering tools. In addition to that the paper also present an understanding of a wide area of Computer Science to demonstrate that highly accurate speech and facial emotion detection analysis is possible, and that it can be done in real time.

ACKNOWLEDGMENT

I wish to express my 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.

.

REFERENCE

[1] Shamla Mantri, Dr. Dipti Patil, Dr. Pankaj Agarwal, Dr.Vijay Wadhai, "Cumulative Video Analysis Based Smart Framework for Detection of Depression Disorders", 2015 International Conference on Pervasive Computing (ICPC).

[2] Shamla Mantri, Dr.Dipt Patil, Ria Agarwal, 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 (IJCST) – Volume 2 Issue 2, Mar-Apr 2014.

[3] Aitor Azcarate, Felix Hageloh,Koenvan de Sande,Roberto Valenti, "Automatic Facial Emotional Recognition",Universiteit van Amsterdam,June 2005.

[4] LiyanageCDe Silva,ChunHui "Real Time Facial Feature Extraction and Emotion Recognotion",2003.

[5] P.M.Chavan,Manan C.Jadhav,Jinal B.Mashruwala, "Real Time Emotion Recognition through Facial Expressions for Desktop Devices",International Journal of Emerging Science and Engineering,Volume-1,Issue-7,May2013.

[6] Alex Mordkovich, Kelly Veit, Daniel Zilber,"Detecting Emotion in Human Speech", December 16th, 2011.

[7] Asthana, A., Saragih, J., Wagner, M., Goecke, "Evaluating AAM Fitting Methods for Facial Expression Recognition" Proceedings of the IEEE International Conference on Affective Computing and Intelligent Interaction, ACII'09, pp. 598–605 (2009).

[8] Casale S.,Russo A.,Scebba G.,Serrano,"Speech Emotion Classification Using Machine Learning Algorithms Semantic Computing",IEEE International Converence,2008.

[9] Zhu, X., Ramanan, "Face detection, pose estimation, and landmark localization in the wild",In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2879–2886 (2012).

[10] Ambadar. Z.,Schooler, J.,Cohn ,"Deciphering the Enigmatic face:The importance of facial dynamics to interpreting subtle facial expression",Psychological Science,2005.

[11] N.Fragopanagos,J.G.Taylor ,"Emotion Recognition in Human-computer Interaction",Neural Networks,23 March 2005.

[12] L.Cohen,N.Sebe,A.Garg,L.Chen,andT.S.Hung,"Facial Expresion Recognition From video sequences:Temporal and static modeling.Computer Vision and Image Understanding",91(1-160-187,2003.

[13] Alghowinem, S.Goecke, R.Wagner, M. Epps, J. Breakspear, M. Parker, G.From Joyous to Clinically Depressed, "Mood Detection Using Spontaneous Speech",In: Proc. FLAIRS-25 (2012).

[14] M.Pantic,L.J.M.Rothkrantz,"Automatic Analysis of Facial Expressions:the state of the art",2000.

[15] V.A.Petrushin,"Emotion Recognition in Speech Signal:ExperimentalStudy,Development,and Application",ICSLP- 2000,Vol.2,2000.

[16] Joyti Joshi, Roland Goecke,Abhinav Dhall,Sharifa Alghowinem, " Multimodal Assistive Technologies for Depression Diagnosis And Monitoring",Journal on Multimodal User Interfaces manuscript.