Automatic Emotion Analysis Through Visual And Vocal Expressions Based On AI

P.Piraveena[1], Dr. M.Chinnadurai[2]
[1, 2] Dept of Computer science and Engineering
[1, 2] E.G.S Pillay Engineering College, Nagapattinam, India

Abstract- *Analysis of observable behavior in depression primarily relies on subjective measures. New computational approaches make possible automated audiovisual measurement of behaviors that humans struggle to quantify (e.g., movement velocity and voice inflection). Emotions are normally displayed by visual, vocal, and other physiological means. One of the important way humans display emotions is through facial expressions. Facial expression is one of the most powerful ways that people bring together conversation and communicate emotions and other mental, social, and physiological cues. These tools have the potential to improve screening and diagnosis, identify new behavioral indicators of depression, measure response to clinical intervention, and test clinical theories about underlying mechanisms. Highlights include a study that measured the temporal coordination of vocal tract and facial movements, a study that predicted which adolescents would go on to develop depression based on their voice qualities, and a study that tested the behavioral predictions of clinical theories using automated measures of facial actions and head motion. The large sizes of data in recent years has led to the need for big data and streaming frameworks for mining.*

Keywords- Facial Expression, Voice Expression , FDHH, Emotion Recognize

I. INTRODUCTION

Facial Feature Point Detection

An automatic facial expression analysis system usually consists of three consecutive steps: data acquisition, feature extraction and classification. The goal of our fiducial facial point detector is to detect 20 fiducial facial points plus the irises and the medial point of the mouth in the face region. It give a definition of these points, which is needed for manual annotation of the points for training and evaluation of the facial point detection system

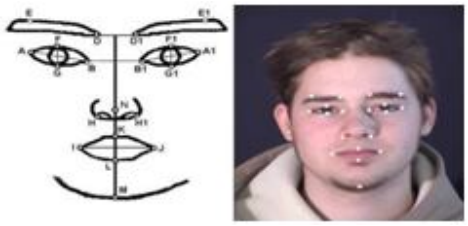

Fig: fiducial point

An affect sensor is a device that takes an input signal and processes it for some evidence of emotions. There are a number of different techniques and modalities used to detect affect. These include: physiological signals, facial expression recognition, speech prosody recognition and pressure sensors.

What is affective computing? There are two main elements involved in affective computing:

- Detect the emotion via facial expressions, gestures, tone of voice, pressure on a keyboard, or other physiological measurements.
- Put this information into context.

This may seem straight forward, as this is a process that humans do all the time. We can determine just from a glance whether someone is upset, interested or bored. However, computers are unable to do this, partially because emotional response studies were carried out on the assumption that emotions are expressed in the same way across a population. This, however, is not the case

Detecting and recognizing emotional information

Detecting emotional information begins with passive sensors which capture data about the user's physical state or behavior without interpreting the input. The data gathered is analogous to the cues humans use to perceive emotions in others. For example, a video camera might capture facial expressions, body posture and gestures, while a microphone might capture speech. Other sensors detect emotional cues by directly measuring physiological data, such as skin temperature and galvanic resistance.

Recognizing emotional information requires the extraction of meaningful patterns from the gathered data. This is done using machine learning techniques that process different modalities speech recognition, natural language processing, or facial expression detection, and produce either labels (i.e. 'confused') or coordinates in a valence-arousal space.

- The nature of motivation, emotions and feeling.
- The detection of emotional and other affective states and processes.
- The nature of intelligence and the relationships between intelligence and emotions.
- The physiology of the brain and other aspects of human physiology relevant to affective states.
- Requirements for effective human-computer interfaces in a wide range of situations.

Emotional Recognition from Facial Expressions

Knowing how facial expressions relate to the underlying emotional experiences is an important factor in using facial expression measurements as an input signal in affective computing. Therefore, the assessment of emotional experiences from objectively measured facial expressions becomes an important research topic. In the field of facial expression recognition, several efforts have been made in trying to recognize expressions of discrete emotions, especially the ones suggested . Although there is evidence for universal facial expressions of certain emotions , it is important to realize that there are also differences in the facial behavior of different people. With regard to this issue, supported that the most accurate interpretation of facial expression benefits from the knowledge of what is normative for each individual. Hence, the findings that there are considerable differences in facial behavior between individuals recommend that the best results in emotion estimation could be obtained using a person adaptive system. This system would form an individual model of facial behavior for each individual user

An important issue is that many of the existing facial recognition systems rely on analyzing single facial images instead of tracking the changes in facial expressions continuously . It would be more meaningful if the computerized learning environments could analyze the student's facial expressions continuously to be able to react to changes in the student's emotional state at the right time. The point that the lack of temporal information is a significant limitation in many facial expression recognition systems. Consequently, methods for analyzing facial expressions in human-computer interaction, especially those concerning computer-aided learning systems.

Emotional Speech Recognition

The modulation of voice intonation is one (of the) main channel(s) of human emotional expression . Certain emotional states, such as anger, fear, or joy, may produce physiologic reactions, such as an increase of cardiac vibrations and more rapid breathing. These in turn have quite mechanical and thus predictable effects on speech, particularly on pitch timing and voice quality .Some researchers have investigated the existence of reliable acoustic correlates of emotion in the acoustic characteristics of the signal. Their results agree on the speech correlates that are derived from physiological constraints and correspond with broad classes of basic emotions, but disagree and are unclear concerning the differences between the acoustic correlates of fear and surprise or boredom and sadness. This is perhaps explained by the fact that fear produces similar physiologic reactions to surprise, and boredom produces similar physiologic reactions to sadness, and consequently very similar physiological correlates result in very similar acoustic correlates .

Applications for Affective Computing

Technologies that can read human emotions can help people who may have trouble determining the emotions of others, like those with Autism, or provide companionship and support for nursing home residents.

Emotional awareness enables you to:

Accurately read other people, including the emotions they're feeling and the unspoken messages they're sending.

- Create trust in relationships by sending nonverbal signals that match up with your words.
- Respond in ways that show others that you understand.

II. LITERATURE REVIEW

2.1 Automated Audiovisual Depression Analysis

Author: M. Girarda , F. Cohna

Depression has salient, observable behavioral symptoms pertaining to general psychomotor functioning, the

expression of affective states, and the negotiation of in temporal situations. Before extracting prosodic features from an audio signal, it is useful to segment participant speech from periods of silence, noise, and the speech of other parties. Three main approaches to analyzing depression from audiovisual information have Numerous techniques for feature extraction and supervised learning have been proposed and evaluated. Automated methods for behavior analysis can provide measurements that are difficult for humans to quantify and have the benefit of high repeatability. While clinicians may vary in their degree of accuracy and consistency, anyone implementing the same automated system can be confident that it will perform consistently.

2.2 Automatic Facial Expression Interpretation: Where Human-Computer Interaction, Artificial Intelligence and Cognitive Science Intersect

Author: L. Lisetti , J. Schiano

The main motivating principle is that computers should be adapting to people rather than vice versa. Our research group is working on the construction of a computer system capable of recognizing and responding to cognitive and affective states of users while they are involved performing various tasks and activities. Facial expressions can also be considered as a modality for communication, the face being an independent channel conveying conversational signals. Given some of the newest results in neuroscience emphasizing the plasticity of the human brain, facial actions have also recently been considered as emotional activators and regulators. A possible exception can be found in the case of fear and seemed constrained by the forced choice procedure. Finally another one of the controversial issues still disputed, is whether facial expressions are perceived by humans as varying continuously along certain underlying dimensions, or as belonging to qualitatively discrete categories.

III. PROPOSED SYSTEM

Human facial expressions and voices in depression are theoretically different from those under normal mental states. An attempt to find a solution for depression scale prediction is achieved by combining dynamic descriptions within naturalistic facial and vocal expressions. A novel method is developed that comprehensively models the variations in visual and vocal cues, to automatically predict the scale of depression. The proposed framework is an extension of the previous method by replacing the hand-crafted techniques with deep face representations as a base feature to the system.

For the deep feature process, the temporal data for each sample is broken down into static image frames which are pre-processed by scaling and subtracting the given mean image. These are propagated forward into the deep network for high level feature extraction. Once the deep features are extracted for a video sample, it is rank normalized between 0 and 1 before the FDHH (Feature Dynamic History Histogram) algorithm is applied across each set of features per video. The output is transformed into a single row vector, which will represent the temporal feature of one video.

3.1 FACIAL FEATURE EXTRACTION AND EMOTION CLASSIFICATION

Facial features extracted from images or video clips can be broadly categorized as geometrical features and appearance based features. Geometrical features consist of shapes of facial components (eyes, lips ,smiling etc.) and salient points on the face (nose tip etc.).This provides us the component related to the variations resulting from facial expressions, which are then classified using Support Vector Classifiers (SVC). Appearance based features provide information about the texture of the face as well .Geometrical. It is expected that methods that use both geometrical and appearance based features give more accurate results .A major problem in classifying facial expressions is defining the emotion classes. anger, disgust, fear, happiness, sadness and surprise are universal and associated with muscular patterns in all cultures.

3.2 VOICE TONE EXPRESSION

Natural language is one of the most used communication methods, and although it has been extensively studied, relevant aspects still remain opened. As stated before, in order to obtain a more natural and trustworthy interaction, systems must be capable of responding appropriately to the users with affective feedback. Within verbal communication it implies the addition of variability in the answers and the synthesis of emotions in speech.

Facial expressions were accompanied by speech with a related tone of voice because we believed that this combination could improve recognition of the target emotions. The relevant sentences were uttered by a trained actress to convey the desired emotion. Fear and Neutral facial expressions were not easy recognized so used the voice tone expression that to analysis the some various. The Euclidean distance (ED) has been computed between the new (testing) image Eigenvector and the Eigen subspaces for each expression, and minimum Euclidean distance based

classification is done to recognize the expression of the input image. The formula for the Euclidean distance is given by

$$ED = \sqrt{\sum(x_2 - x_1)^2} \quad \text{.....(1)}$$

The Eigenvectors are obtained from the input image and voice then EDs between each Eigenvectors and the reference Eigenvectors of each trained expressions are obtainedx1 and x2 stands for expression analysis. If two expressions are same, then ED will be minimum to extract correct emotion recognize.

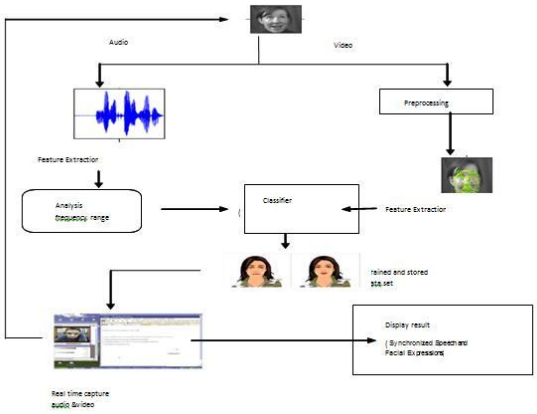

Fig: Expression Analysis

3.3 MODULES

a) Emotion recognition systems
b) Methodology
c) Feature extraction
d) Fusion for audio and visual
e) Expression analysis
f) Experimental result

IV. SYSTEM WORK

4.1 TEXTURE FEATURE ANALYSIS

We recognize texture when we see it but it is very difficult to define. This difficulty is demonstrated by the number of different texture definitions attempted by vision researchers Generally speaking, textures are complex visual patterns composed of entities, or sub patterns that have characteristic brightness, color, slope, size, etc. Thus texture can be regarded as a similarity grouping in an image.

There are four major issues in texture analysis:

1) Feature extraction: to compute a characteristic of a digital image able to numerically describe its texture properties.
2) Texture segmentation: to partition a textured image into regions, each corresponding to a perceptually homogeneous texture.
3) Texture classification: refers to the process of grouping test samples of texture into classes, where each resulting class contains similar samples according to some similarity criterion. If the classes have not been defined a priori, the task is referred to as unsupervised classification.
4) Shape from texture: to reconstruct image geometry from texture information. Approaches to texture analysis are usually categorized into

- Structural,
- Statistical,
- Model-based and transform

4.2 GEOMETRIC FACIAL FEATURE

Lips:

A different lip contour template is prepared for each lip state. The open and closed lip contours are modelled by two parabolic arcs, which are described by six parameters: the lip centre position, the lip shape, and the lip orientation. For tightly closed lips, the dark mouth line connecting the lip corners represents the position, orientation, and shape.

Eyes:

In order to detect whether the eyes are open or closed, the degree of eye opening, and the location and radius of the iris. Two eye states are proposed: open and closed.

Brow and cheek*:*

Features in the brow and cheek areas are also important for expression analysis. Each left or right brow has one model a triangular template with six parameters. Each cheek has also a similar six parameter down-ward triangular template model.

V. CONCLUSION

Automatic affect recognition is inherently a multidisciplinary enterprise involving different research fields, including psychology, linguistics, computer vision, speech analysis, and machine learning. There is no doubt that the progress in automatic affect recognition is contingent on the progress of the research in each of those fields Speech is

another important communicative modality in human-human interaction. Speech conveys affective information through explicit (linguistic) and implicit (paralinguistic) messages that reflect the way that the words are spoken.

REFERENCE

[1] M. Girarda, F. Cohna "Automated Audiovisual Depression Analysis" in 2015
[2] CL Lisetti "Automatic Facial Expression Interpretation: Where Human-Computer Interaction, Artificial Intelligence and Cognitive Science Intersect" in 2000.
[3] M. Marcus, M. T. Yasamy, M. van Ommeren, and D. Chisholm, "Depression, a global public health concern," pp. 1–8, 2012.
[4] World Health Organization, "Depression and other common mental disorders: global health estimates," Tech. Rep., 2017.
[5] R. C. Kessler and E. J. Bromet, "The epidemiology of depression across cultures." Annual review of public health, vol. 34, pp. 119–38, 2013.
[6] E. Jenkins and E. M. Goldner, "Approaches to understanding and addressing treatment-resistant depression: A scoping review," 2012.
[7] D. D. Luxton, Artificial intelligence in behavioral and mental healthcare. Academic Press, 2015.
[8] J. F. Cohn, T. S. Kruez, I. Matthews, Y. Yang, M. H. Nguyen, M. T.Padilla, F. Zhou, and F. De La Torre, "Detecting depression from facial actions and vocal prosody," in Proceedings - 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, ACII 2009, 2009.