Emotion Recognition from Audio-Visual Data

Asha Sugave[1], Sahil Mulani[2], Trupti Shinde[3], Swapnali Tolnur[4]

Department Of Computer Engineering,TSSM's Bhivarabai Sawant College of Engineering and Research, Narhe Pune[1-4]

Abstract: Emotion Recognition Systems is used to identifying the emotions of humans with their accuracy. This paper by using Audio-visual Data to recognizing emotion. We have to use machine learning. This emotion recognition system automatically identifies the human emotional states from his or her voice and face images. An audiovisual emotion recognition system is used to develop uses fusion algorithm. In this system firstly separate emotion recognition systems that use voice and facial expressions were tested separately. These two databases are used in this system Berlin EMODB database and Assamese database. This paper is dependent on recognizing human emotion and voice emotion. This system uses many emotional persons.

Keywords: Facial Expression Recognition, Emotional Model, Social Interfaces, Emotion Recognition, Machine learning.

1 INTRODUCTION

Emotions are very complex states or processes of human beings. Automatic facial and voice expression recognition has been a growing topic in computer vision. Emotion recognition is an important part of social life. The future human-computer emotional interaction research is used the complementary for the multi-modal signals to an emotional state. A multi-modal emotional interaction algorithm is used in this system. In the current research, we investigate such inferential mechanisms on the recognition of social emotions. Modeling such a process using synthesized facial expressions allowed us to focus not simply on a restricted subset of facial movements, but also to include other movements that serve communicative purposes, adjusting dynamically to changing events. Specifically, we focused on the recognition of facial expressions. emotional intelligence is not enough to interact and communicate with users smartly. Emotions include three parts: subjective experience (personal feelings), physiological arousal (physiological signal changes in the body), and outer performance (quantified response to actions for each body part).

2 EMOTION RECOGNITION ARCHITECTURE

The decision about underlying emotion in a video is taken by combining the results of both speech and image-based systems. The complete system has been built in two phases: the training phase and the testing phase. In the training phase, separate classifiers were trained with facial expression and speech features extracted from the training video data. At the time of testing, emotion in an unknown test video was decided based on the classification results of both speech and image-based systems.

In fig(a) shows the architecture of emotion recognition. Following are the steps of a system.

a.   Voice Processing Unit: processes the audio signal of the emotional videos.

1.Speech Preprocessing: It is used to The speech signal has been preprocessed to make it more suitable for feature extraction.

fig(a) Architecture of Emotion Recognition
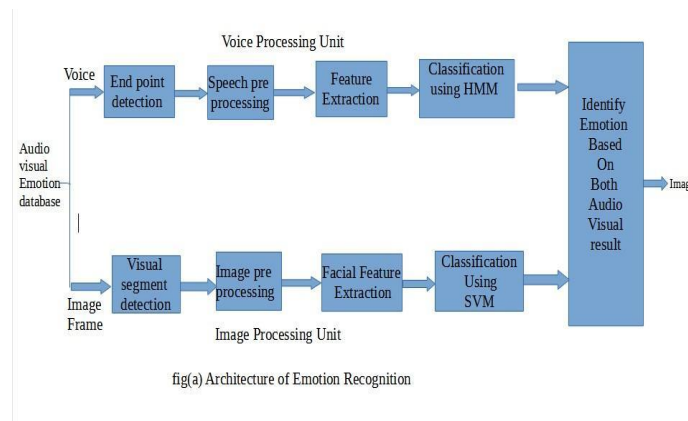
2.Feature Extraction: It is used for the preprocessed speech signal to consider only the required prosodic information.

3.Classification using Hidden Markov Model: HMM is a powerful statistical tool for modeling sequences that can be characterized   by an underlying process generating observable sequences.

b.Image Processing Unit: processes the image frames of the emotional videos.

a.   Image preprocessing: The emotion recognition system has been trained and tested using the eNTREFACE05 database that contains videos of emotional expressions. The image frames, coming from the video, contain the face as well as the background.

b.   Facial feature extraction: Image features representing facial expressions have been extracted from the incoming image frames of the video.

c.   Classification using Support Vector Machine (SVM): For facial emotion recognition system, Support Vector Machine (SVM) classifier has been used instead of HMM to avoid data sparsity problem that happens in the case of image.

What is librosa?

Librosa is for analyzing audio and music. It is a flatter package layout, standardizes interfaces and names, backward compatibility, and readable code. Further, in this Python mini-project.

What is OpenCV?

OpenCV (full form: Open Source Computer Vision Library) is a machine learning software library and open-source computer vision.   OpenCV was to accelerate the use of machine perception in commercial products and was built to provide a common infrastructure for computer vision applications.

3 LITERATUREREVIEW

1)   FERDOUS AHMED, Emotion Recognition From Body Movement. This paper proposed the system that Automatic emotion recognition from the analysis of the body.

2)   RUHUL AMIN KHALIL1, EDWARD JONES 2, Speech Emotion Recognition Using Deep Learning Techniques: A Review. This paper presents an overview of Deep Learning techniques where these methods are utilized for speech-based emotion recognition.

3)   GUOSHEN G YANG, RUI JIAO, HUI PING JIANG, AND TING ZHANG, Ground Truth Dataset for EEG- Based Emotion Recognition With Visual Indication. This paper stated that Along with extensive and successive applications, emotion recognition based on electro scope hologram.

4)   GUANGLO NG DU, SHUAIYIN G LONG, AND HUA YUAN, Non-Contact Emotion Recognition Combining Heart Rate and Facial Expression for Interactive Gaming Environment. In this study, we proposed a method to detect a player's emotions based on heartbeat (HR) signals.

5)   JIANZHU GUO1,2, ZHEN LEI 1,2, (Senior Member, IEEE), JUN WAN 1,2, (Member, IEEE), EGILS AVOTS3, (Student Member, IEEE), Dominant and Complementary Emotion Recognition From Still Images of Faces. This paper proposed the system that Emotion recognition has a key role in affective computing.

6)   IGOR BISIO, ALESSAN DRO DELFINO, FABIO LAVAGET TO, MARIO MARCHESE, AND ANDREA SCIARRO NE, GenderDriven Emotion Recognition Through Speech Signals for Ambient Intelligence Applications. This paper proposes it allows recognizing a person's emotional state starting from audio signal registrations.

7)   Subhasmita Sahoo, Changho Seo, Joel J.P.C. Rodrigues(Fellow, IEEE), (7 May 2016), Emotion Recognition From Audio-Visual Data Using Rule-Based Decision Level Fusion. In this paper, a multimodal emotion recognition system has been proposed that uses audio and visual data. Single features of speech and image have been used to make it computationally less complex.

8)   Sidorov M., Sopov E., Ivanov I., and Minker W.. (3 March 2015), Feature and Decision Level Audio-visual Data Fusion in Emotion Recognition Problem. The paper explores the problem of human emotion recognition and proposes the solution of combining audio and visual features.

4 PROBLEM STATEMENT

Recognition audio and visual dataset detect emotions with the basis of decision level. The system consumes two different emotion and according to emotion accuracy send the result.

5 OBJECTIVE

●   Emotion recognition systems identifying the emotions of human subjects from underlying data with acceptable accuracy.

● The emotion recognition system must automatically identify the human emotional states from his or her voice and facial image and identify human current behavior.

● Implement of recognizing emotion from speech using the Librosa and Sklearn libraries and the RAVDESS dataset.

6 PROPOSED SYSTEM

In the Proposed system is while taking interviews and finding human current behavior. The most fundamental requirement in generating suitable machine responses in Human-Machine-Interface (HMI) or Human-Computer-Interface (HCI) applications is prior knowledge about the affective states of the person being interacted with.

7 CONCLUSION

emotion recognition systems that use speech and image data were analyzed separately and tested with a standard emotion dataset. Then, audio and Visual information was used together with a rule-based decision level fusion to detect emotions with greater accuracy. The decision rule that has been set here is based on the performance of the unimodal systems on the dataset.

REFERENCES

1. O.-W. Kwon, K. Chan, J. Hao, and T.-W. Lee, "Emotion recognition by speech signals." in INTERSPEECH. Citeseer, 2003.
2. B. Schuller, G. Rigoll, and M. Lang, "Hidden Markov model-based speech emotion recognition,". Proceedings. (ICASSP'03). 2003 IEEE International Conference on, vol. 2. IEEE, 2003, pp.II–1.
3. Busso, Z. Deng, S. Yildirim, M. Bulut, C. M. Lee, A. Kazemzadeh, S. Lee, U. Neumann, and S. Narayanan, "emotion recognition using facial expressions, and speech," in Proceedings of the 6th international conference on Multimodal interfaces. ACM, 2004, pp. 205–211.
4. N. Sebe, I. Cohen, T. S. Huang, et al., "Multimodal emotion recognition," Handbook of Pattern Recognition, vol. 4, pp. 387– 419,2005.
5. F. Burkhardt, A. Paeschke, M. Rolfes, W. F. Sendlmeier, and B. Weiss, "A database of german emotional speech, vol. 5, 2005, pp.1517–1520.