Review Paper on Recognizing Emotions Using Multimodal Information From Spontaneous Non-basic & micro expressions

Ms.Deepti Rathod
M.Tech CSE
Department of Computer Science and Engineering
G. H. Raisoni Institute of Engineering And Technology for Women,
Nagpur, Maharashtra, India

Ms. Ranjana Shende
Assistant Professor
Department of Computer Science and Engineering
G.H.Raisoni Institute of Engineering And Technology for Women,
Nagpur,Maharashtra,India

ABSTRACT

Emotion plays an important role in human communications. The proposed approach is based on the analysis of video sequences which combines facial expressions observed visually with acoustic features to recognize emotion classes. In order to locate and track facial feature points, an Active Appearance Model for facial images with all kinds of expressions is used. Short-term mean energy, fundamental frequency and formant frequencies from each frame as speech features is extracted. The two modalities are combine at both feature and score level to compare the respective joint emotion recognition rates. The emotions are instantaneously classified using a Support Vector Machine and sequentially aggregated based on their classification probabilities. The proposed approach also helps in recognizing all the emotions with equal accuracy rate.

Keywords:Differential evolution Markov chain,PRAAT &Mel-frequency Cepstral Coefficient (MFCC),Support Vector Machine (SVM) classifier,facial expressions.

I. INTRODUCTION

Emotion recognition is an important research field of pattern recognition. Emotion takes a significant role in human communications, and has an effect on perception and decision making. Emotion recognition is widely used in human-computer interaction, medical care, security, communication, and many other fields. Psychological researchers have found six kinds of affective states, including happiness, anger, sadness, fear, surprise, and disgust. Facial expressions are derived from motions of facial muscles.One of the most interesting aspects of facial expression analysis is recognizing spontaneous non-basic & micro-expression**.** Many researchers have used speech signals to recognize emotions of people. Three fusion strategies have been applied in multi-modal emotion recognition: feature-level fusion, model-level fusion, and decision-level fusion. Feature- level fusion concatenates speech features and facial expression features to construct combined feature vector. Model-level fusion relaxes the requirement of synchronization and makes use of the correlation of multi-streams as well. Decision-level fusion which independently models features from video and audio, and combines these unimodal recognition results in the end.This multi-modal approach extractspontaneous non-basic & micro-expression features from video stream and speech features from audio stream respectively. The input video is split into audio stream and video stream. Emotional features are extracted from facial image sequences and speech signals, and fused together to find the exact emotions of a person.

ISSN: 2278 – 7798

II. LITERATURE REVIEW

Yun Tie, Member, IEEE, and Ling Guan describe the facial region is first detected with local normalization in the input frames. The 26 fiducial points are then located on the facial region and tracked through the video sequences by multiple particle filters. Depending on the displacement of the fiducial points, may be used as landmarked control points to synthesize the input emotional expressions on a generic mesh model. And also extracts the deformation feature from the realistic emotional expressions [1].

B. Schuller, G. Rigoll and M. Lang. describe the audio component of the system extracts features related to pitch and intensity along with Mel-frequency Cepstral Coefficient (MFCC) and formant frequencies. The feature extraction process involves the preprocessing of the audio signal by removing the leadingand trailing silent edges of the signal. The pitch and intensity contours of the audio signal are obtained using PRAAT, the speech analysis software. Once, the features are extracted they are normalized separately for each subject before performing the classification using a multi-class SVM [6].

Malika Meghjani, Frank Ferrie and Gregory Dudek explains the high dimensional feature vectors obtained from the two modalities are reduced using a feature reduction method which apply Recursive Feature Elimination (RFE) method to obtain a minimum subset of the most discriminative feature set. The RFE method, iteratively removes the input features using a ranking criterion. The ranking criterion is based on the weights obtained from a classifier like SVM. The reduced features are classified using a Support Vector Machine (SVM). The SVMs are binary classifiers which can be extended for classifying more than two classes using two techniques:(a) one-against-rest and (b) one-against-one classification. One-against-one technique to classify five emotion classes. This method compares pair-wise classes which results in combinations of the classifier. The final classification result is based on maximum-wins voting scheme. The score level fusion combines the two modalities based on the scores or the probability estimates obtained after individually classifying the two modalities using the multi-class SVM. The score for the visual system is obtained by temporally aggregating the weighted scores from an interval of frames in the visual sequence. The weights and the interval of frames for the temporal aggregation are decided based on two criteria: (a) maximum audio intensity and (b) minimum entropy of the probability distribution [7].

Chao Xu, Pufeng Du, Zhiyong Feng, Zhaopeng Meng, Tianyi Cao, and Caichao Dong describe an emotion classifier designed to fuse facial expression and speech based on Hidden Markov Models and Multi-layer Perceptron. Emotional features are extracted from facial image sequences and speech signals, and fused on decision level. In the phase of emotion recognition, Hidden Markov Models are constructed for every expression and emotional speech. Multi-layer Perceptron is applied to fuse expression features and emotional speech features, and exact emotion are acquired [2].

III. Proposed Methodology

The outline of proposed approach for emotion recognition using multimodal analysis is given in figure1. The details of each processing step are discussed in the following sections. First, the input video is split into audio stream and video stream. After that feature extraction is takes place. Feature extraction is a critical procedure of pattern recognition issue. The accuracy of the extracted affective

ISSN: 2278 – 7798

features has a significant impact on the result of emotion recognition.

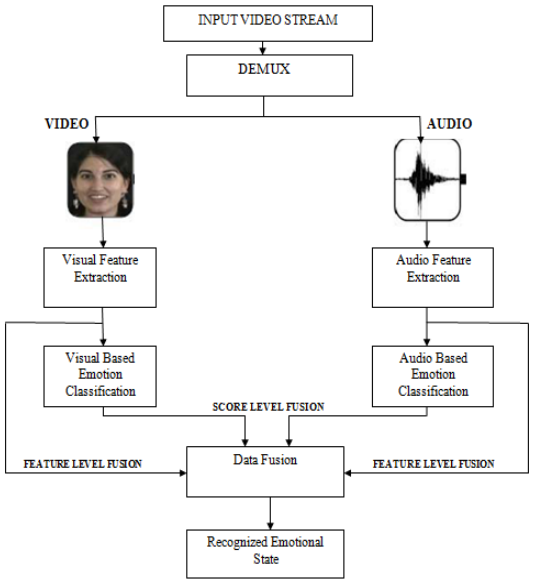

Figure 1: Flow diagram of the proposed methodology

Visual Feature Extraction:

Facial expression features and emotional speech features are extract simultaneously from video sequences and speech signals. The facial region is first detected automatically in the input frames using the local normalization-based method .Then locate 26 fiducial points over the facial region using scale-space extrema and scale-invariant feature examination. The fiducial points are tracked continuously by multiple particle filters throughout the video sequences. Elastic Body Spline (EBS) is used to extract the deformation features.

Acoustic Feature Extraction:

The feature extraction process involves the preprocessing of the audio signal by removing the leading and trailing silent edges of the signal. From the audiosignal : the fundamental frequency or pitch, the energy of the signal,the first three formant, the harmonicity of the signal, the first nine linear predictive coding coefficients,the first ten mel–frequency cepstral coefficients are extracted.These 26 features are collected with the use of PRAATand down sampled to 25 samples per second to help synchronization with video features. The processing time of the audio analysis is compatible with real-time constrains. Speech rate and pausing structure are two other features which are thought to carry emotional information but they are related to long term analysis of the speech (several words seconds) and are therefore not compatible with real–time constraints.

Classification:

Once the features are extracted they are normalized separately for each subject before performing the classification. The features selected from the above process are classified using a Support Vector Machine (SVM). The SVMs are binary classifiers which can be extended for classifying more than two classes using two techniques: (a) one-against-rest and (b) one-against-one classification. The one-against-one technique is use to classify five emotion classes. This method compares pair-wise classes which results in combinations of the classifier. The final classification result is based on maximum-wins voting scheme.

Data Fusion:

The method for combining the two modalities is based on the scores or the probability estimates obtained after individually classifying the two modalities using the multi-class SVM. The score for the visual system is obtained by temporally aggregating the weighted scores from an interval of frames in the visual sequence. The weights and the interval of frames for the temporal aggregation are

ISSN: 2278 – 7798

decided based on two criteria: (a) maximum audio intensity and (b) minimum entropy of the probability distribution.

Emotion Recognition:

An emotion classifier is use to fuse facial expression and speech based on Hidden Markov Models and Multi-layer Perceptron. Emotional features are extracted from facial image sequences and speech signals, and fused on decision level. In the phase of emotion recognition, Hidden Markov Models are constructed for every expression and emotional speech. Multi-layer Perceptron is applied to fuse expression features and emotional speech features, and exact emotion are acquired.

IV. Conclusion

Multi-modal fusion emotion recognition is an important research field of pattern recognition. The multi-modal approach has relatively high recognition rate.It is the real time, user independent, identification of the primary as well as secondary emotions of a person from video. Also this approach recognize & track the spontaneous non-basic& micro-expressions of a person. The proposed approach also help in recognizing all the emotions with equal accuracy rate.

V. References

[1]Yun Tie, Member, IEEE, and Ling Guan, Fellow, IEEE,"A Deformable 3-D Facial Expression Model for Dynamic Human Emotional State Recognition",IEEE Transactions on circuits and systems for video technology, Vol. 23, No. 1, January 2013.

[2]Chao Xu, Pufeng Du, Zhiyong Feng, Zhaopeng Meng, Tianyi Cao, and Caichao Dong,"Multi-Modal Emotion Recognition Fusing Video and Audio" Applied Mathematics & Information Sciences,An International Journal, Mar. 2013.

[3] J. Ye, M. Zhang, X. Gong, Speech emotion recognition based on MF-DFA. Computer engineering and applications, (2012), Vol. 48, No. 18, 119-122.

[4]N. Zhao, Y. Liu, Product Approximate Reasoning of Online Reviews Applying to Consumer Affective and Psychological Motives Research. Applied Mathematics & Information Sciences, (2011), Vol. 5, No. 2, 45-51.

[5] P. Lucey, J.F. Cohn, T. Kanade et al., The Extended Cohn- Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. In Proceedings of IEEE workshop on CVPR for Human Communicative Behavior Analysis. (2010), 94-101.

[6]B. Schuller, G. Rigoll and M. Lang. "Speech Emotion Recognition Combining Acoustic Features and Linguistic Information in a Hybrid Support Vector Machine-Belief Network Architecture". IEEE International Conference onAcoustics, Speech, and Signal Processing, volume 1, May 2009.

[7]Malika Meghjani, Frank Ferrie and Gregory Dudek,"Bimodal Information Analysis for Emotion Recognition", Department of Electrical and Computer Engineering and School of Computer Science,2009.

[8]Y. Wang and L. Guan, "Recognizing human emotional state from audiovisual signals," IEEE Trans. Multimedia, vol. 10, no. 5, pp. 659–668, Aug. 2008.

[9]M. Song, Z. Dong, C. Theobalt, H. Q. Wang, Z. C. Liu, and H. P.Seidel, "A general framework for efficient 2-D

ISSN: 2278 – 7798

and 3-D facial expressionanalogy," IEEE Trans. Multimedia, vol. 9, no. 7, pp. 1384–1395, Nov.2007.

[10]H. Soyel and H. Demirel, "Facial expression recognition using 3-D facial feature distances," in Proc. Int. Conf. Image Anal. Recognit., vol. 4633.Aug. 2007, pp. 831–838.

First Author: Ms. DEEPTI R. RATHOD

Department of Computer Science and Engineering,

G. H. Raisoni Institute Of Engineering And Technology For Women,

Nagpur,Maharashtra,India

Nagpur, Maharashtra, India

Mobile No.-7083921462

Education Details-B.Tech (IT) from VJTI,Mumbai

Pursuing M.Tech (CSE) From G. H. Raisoni Institute Of Engineering And TechnologyFor Women,Nagpur

Second Author: Ms.RANJANA SHENDE

Department of Computer Science and Engineering,

G. H. Raisoni Institute And Technology For Women,

Nagpur, Maharashtra, India

Mobile No.-9766969604

Education Details:B.E.(CSE) from K.D.K. college Nagpur,

M.Tech(CSE) from G.H.Raisoni College Of Engineering,Nagpur

Publications: IJERT,IJCST

Membership: CSI member

ISSN: 2278 – 7798