

Expression Recognition by Using Facial And Vocal Expressions

Gholamreza Anbarjafari

IMS Lab
Institute of Technology
University of Tartu
Tartu, Estonia
sjafari@ut.ee

Alvo Aabloo

IMS Lab
Institute of Technology
University of Tartu
Tartu, Estonia
alvo.aabloo@ut.ee

Abstract

Human behaviour may be monitored by analysing facial expressions and vocal expressions. Hence an automatic technique which combines both these features will give a more accurate overall estimation of expression. In this work we propose a new method which uses facial and vocal features to estimate the expression of the subject. Facial expressions are analysed by extracting important facial features and then clustering the movement of these features. In parallel the voice is processed by using considering sudden changes in amplitude and frequency in order to recognize the expression. Finally a weighted sum rule is used to combine the decisions obtained by facial and vocal expression recognition. The proposed technique is tested on an ongoing set of real data monitored by a psychologist.

1 Introduction

Human-computer interaction has been centre of interest of many researchers for a number of years [1-3]. In order to facilitate this interaction it is essential for the computer system, for example a robot, to understand the feelings of the user [4, 5]. Recognition of emotions has been mainly achieved by using facial expression recognition [6, 7]. Techniques using k-nearest neighbour [8], hidden Markov model (HMM) [9], and translation of belief model [10] have been frequently used for implementation of facial expression recognition.

Vocal expression recognition has also been used separately for expression recognition [11, 12]. The combination of both facial and vocal expression is novel and many researchers are investigating an intelligent model for this fusion [13]. One of the main issues with these techniques is the application of HMM as the hidden states are usually unknown and need to be estimated based on assumed probability distributions of the data [14]. In this work we are proposing a new expression recognition technique which is using both facial and vocal expressions in order to estimate the expression of the speaker. The proposed technique is benefiting from the facial and vocal features. The proposed technique is evaluated with many volunteers sitting in front of a camera under controlled illumination reading texts with expressions of happiness, sadness, disgust, surprise, anger, fear, and contempt. The initial experimental results are promising showing that the proposed technique outperforms other current techniques.

2 The Proposed Expression Recognition Technique

In the proposed technique two parallel observations are made. Firstly the facial features are detected using the Viola-Jones face detector. The important features are corners of lips, upper and lower part of lips, corners and centre of eyes, nose, nose tip, corners of eyebrows, and chin. Then movements of these facial features will be used in order to recognize each of the seven basic emotions. Fig. 1 is showing the important features on the facial image which are used for expression recognition.

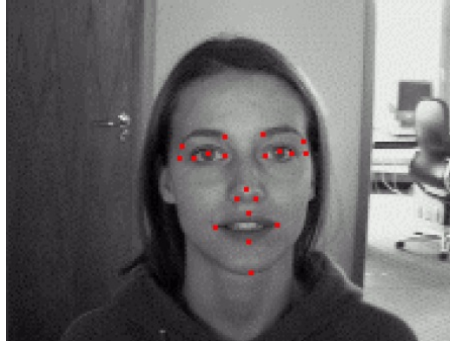


Fig. 1: 20 important facial features which are being used for facial expression recognition.

In addition important features of the voice of the subject are analysed. First the system will be trained by a vocal input with neutral emotion. In order to recognise expression changes in the tone of the voice will be monitored by the sudden changes in amplitude and mid and high frequencies.

The extracted features from face and voice will be combined by using weighted sum in which the weights are being assigned by experts and can be modified. In the early experimental results, the normalized weights were $3/5$ and $2/5$ for face and voice respectively. More improvement is expected to be achieved if supported vector machine (SVM) is employed for fusion of the facial and vocal features. The initial experiments conducted on different volunteers are showing the exact recognition of expressions for sadness, happiness, surprise, and anger. However, expressions of contempt and fear are not being recognized properly due to their high misinterpretation with other expressions. The primary tests are conducted on small datasets and on the continuation of the work a standard database containing both facial and vocal scenarios will be used.

3 Conclusion

In this paper we are proposing a new technique for recognition of expressions by using facial and vocal expression recognition using weighted sum. The proposed technique has been tested on a small set of group of people. The early experimental results show the high potential of the proposal as an accurate expression recognition technique.

Reference

[reference stub]

1. Z. Obrenovic and D. Starcevic, "Modeling multimodal human-computer interaction", *Computer*, 2004, 37(9), 65- 72.
2. S. Benbelkacem, M. Belhocine, N. Zenati-Henda, A. Bellarbi, and M. Tadjine, "Integrating human-computer interaction and business practices for mixed reality systems design: a case study", *IET Software*, 2014, 8(2), 86-101.
3. A. Kucukyilmaz, T. M. Sezgin, and C. Basdogan, "Intention Recognition for Dynamic Role Exchange in Haptic Collaboration", *IEEE Transaction on Haptics*, 2013, 6(1), 56-68.
4. Y. Tie and L. Guan, "A Deformable 3-D Facial Expression Model for Dynamic Human Emotional State Recognition", *IEEE Transactions on Circuits and Systems for Video Technology*, 2013, 23(1), 142-157.
5. M. Pantic and L. J. M. Rothkrantz, "Toward an affect-sensitive multimodal human-computer interaction", *Processing of the IEEE*, 2003, 91(9), 1370-1390.
6. M. S. Bartlett, G. Littlewort, M. G. Frank, C. Lainscsek, I. R. Fasel, and J. R. Movellan, "Automatic recognition of facial actions in spontaneous expressions", *Journal of Multimedia*, 2006, 1(6), 22-35.
7. J. F. Cohn, L. I. Reed, Z. Ambadar, J. Xiao, T. Moriyama, and T. Moriyama, "Automatic analysis and recognition of brow actions and head motion in spontaneous facial behavior", *IEEE International Conference on Systems, Man and Cybernetics*, 2004, 610-616.
8. M. Yeasin, B. Bullot, and R. Sharma, "From facial expression to level of interest: A spatio-temporal approach", *IEEE Conference on Computer Vision and Pattern Recognition*, 2004, 922-927.

9. I. Cohen, N. Sebe, A. Garg, L. S. Chen, and T. S. Huang, "Facial expression recognition from video sequences: temporal and static modeling", *Computer Vision Image Understanding*, 2003, 91(1-2), 160–187.
10. Z. Hammal and M. Kunz, "Pain monitoring: A dynamic and context-sensitive system", *Pattern Recognition*, 2012, 45(4), 1265–1280.
11. T. L. Nwe, S. W. Foo, and L. C. D. Silva, "Speech emotion recognition using hidden markov models", *Speech Communication*, 2003, 41(4), 603–623.
12. A. Nogueiras, A. Moreno, A. Bonafonte, and J. B. Marino, "Speech emotion recognition using hidden markov models", *proceeding of INTER-SPEECH*, 2001, 2679–2682.
13. H. Meng and N. Bianchi-Berthouze, "Affective State Level Recognition in Naturalistic Facial and Vocal Expressions", *IEEE Transaction on Cybernetics*, 2014, 44(3), 315-325.
14. F. Eyben, S. Petridis, B. Schuller, G. Tzimiropoulos, and S. Zafeiriou, "Audiovisual classification of vocal outbursts in human conversation using long-short-term memory networks", *International Conference on Acoustics, Speech and Signal Processing*, 2011, 5844–5847.