Workshop summary for the 3rd international audio/visual emotion challenge and workshop (AVEC'13)

Michel Valstar, Björn Schuller, Jarek Krajewski, Roddy Cowie, Maja Pantic

Workshop Summary for the 3rd International Audio/Visual Emotion Challenge and Workshop (AVEC'13)

Michel Valstar
University of Nottingham
Mixed Reality Lab

Björn Schuller[*]
TU München
MISP Group, MMK

Jarek Krajewski
University of Wuppertal
Schumpeter School of
Business and Economics

Roddy Cowie
Queen's University
School of Psychology

Maja Pantic[†]
Imperial College London
Intelligent Behaviour
Understanding Group

ABSTRACT

The third Audio-Visual Emotion Challenge and workshop AVEC 2013 will be held in conjunction ACM Multimedia'13. Like the 2012 edition of AVEC, the workshop/challenge addresses the interpretation of social signals represented in both audio and video in terms of the high-level continuous dimensions arousal and valence, but importantly this year the data is that of a large number of clinically depressed patients and controls, with a sub-challenge in self-reported severity of depression estimation. Like both previous AVECs, the aim is to bring together the audio and video analysis communities.

Categories and Subject Descriptors

I [**Pattern Recognition**]: Applications

Keywords

Affective Computing, Social Signal Processing, Depression

1. INTRODUCTION

The third Audio-Visual Emotion Challenge and workshop (AVEC 2013) has been organised in conjunction with the 21st ACM International Conference on Multimedia, held in Barcelona, Spain, 21-25 October 2013 (ACM-MM'13). The goal of the challenge is to provide a common benchmark test set for individual multimodal information processing and to bring together the audio and video emotion recognition communities, to compare the relative merits of the

---

[*]The author is further affiliated with Imperial College London, Department of Computing, London, U.K.

[†]The author is further affiliated with Twente University, EEMCS, Twente, The Netherlands.

two approaches to emotion and depression recognition under well-defined and strictly comparable conditions and establish to what extent fusion of the approaches is possible and beneficial. A second motivation is to advance the emotion recognition community beyond small sets of discrete emotions, into the recognition of multiple fully continuous affective dimensions. A third motivation is to build up a community of researchers working on automatic monitoring of objective cues of depression. The final motivation is the need to advance emotion recognition systems to be able to deal with naturalistic behaviour in large volumes of un-segmented, non-prototypical and non-preselected data as this is exactly the type of data that both multimedia retrieval, human-machine/human-robot communication interfaces, and automatic (home-based) depression monitoring systems will have to face in the real world.

2. PARTICIPATION

The call for participation and papers attracted registrations of 27 teams from all over the world, resulting in 7 submissions. The programme committee accepted 4 papers in addition to the baseline paper, which was reviewed independently. Three papers used fully audio-visual techniques and one paper used only the audio modality. Three papers addressed the depression recognition challenge, and one the affect recognition challenge. We hope that these proceedings will serve as a valuable reference for researchers and developers in the area of audio-visual emotion recognition and depression analysis.

3. THE ROAD AHEAD

Audio-visual recognition of dimensional affect is far from a solved case, as the papers included in these proceedings clearly show. While the prediction accuracy of the arousal and valence dimensions was higher than those earlier reported in the AVEC 2012 challenge, this can be attributed probably as much to the more constrained context of the data as to technological development of the automatic emotion recognition systems. And while the dataset used for the AVEC series all contained displays of natural behaviour, they were all recorded under fairly strictly controlled conditions, which means that there's still a bigger challenge waiting for us, where researchers will be challenged to deal

with both natural behaviour as well as natural environmental conditions.

Concerning the prediction of depression severity, it is clear that we have barely scratched the surface of this topic. The papers presented here provide a number of interesting different approaches to this problem, but even the best performing team only achieved moderately successful results, which are far from a level required to implement in a real-world application. Still, given the nature of the data and the seriousness of the problem it addresses, we believe the works presented here are an important step in the right direction, and it is clear that there are many avenues still to be considered.

Taking both the emotion and depression recognition into account, it is fair to say that it is highly likely that there will be more editions of the AVEC series.

Acknowledgments