

AVEC 2014: the 4th International Audio/Visual Emotion Challenge and Workshop

Michel Valstar
University of Nottingham
Mixed Reality Lab
Nottingham, United Kingdom

Björn Schuller^{*}
Imperial College London
Machine Learning Group
London, United Kingdom

Jarek Krajewski
University of Wuppertal
Schumpeter School of
Business and Economics
Wuppertal, Germany

Roddy Cowie
Queen's University
School of Psychology
Belfast, United Kingdom

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

ABSTRACT

The fourth Audio-Visual Emotion Challenge and workshop AVEC 2014 was held in conjunction ACM Multimedia'14. Like the 2013 edition of AVEC, the workshop/challenge addresses the interpretation of social signals represented in both audio and video in terms of high-level continuous dimensions from a large number of clinically depressed patients and controls, with a sub-challenge in self-reported severity of depression estimation. In this summary, we mainly describe participation and its conditions.

Categories and Subject Descriptors

I [Pattern Recognition]: Applications

Keywords

Affective Computing; Social Signal Processing; Depression

1. INTRODUCTION

This year's Audio-Visual Emotion Challenge and workshop (AVEC 2014) has been organised in conjunction with the 22nd ACM International Conference on Multimedia, held in Orlando, Florida, 3 – 7 November 2014 (ACM-MM'14).

The AVEC 2014's theme is '3D Dimensional Affect and Depression' and it is the fourth competition event aimed at comparison of multimedia processing and machine learning methods for automatic audio, visual and audio-visual emotion analysis, with all participants competing under strictly

^{*}The author is further affiliated with Technische Universität München, MISP, Munich, Germany.

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

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.
MM '14, Nov 03-07 2014, Orlando, FL, USA
Copyright is held by the owner/author(s).
ACM ACM 978-1-4503-3063-3/14/11
<http://dx.doi.org/10.1145/2647868.2647869>.

the same conditions in this first of its kind series [3, 2, 4]. However, further similar endeavours are to be noted in the meanwhile, e.g., [1].

The AVEC 2014 edition features besides emotion recognition the serious and important automatic estimation of levels of depression. Adding objective measures to what is otherwise an entirely subjective process of diagnosing and monitoring depression promises to be an invaluable support to the mental health profession besides the usage in media retrieval systems.

As before, 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 two approaches to emotion 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 the need to advance emotion recognition for multimedia retrieval to a level where behaviomedical systems are able to deal with large volumes of non-prototypical naturalistic behaviour in reaction to known stimuli, as this is exactly the type of data that diagnostic tools and other applications would have to face in the real world.

We were calling for participation in two Sub-Challenges: fully-continuous emotion detection from audio, from video, or from audio-visual information, and estimation of one depression level indicator. As benchmarking database an enhanced subset of the AVEC 2013 audio-visual depression corpus of task-specific human-computer interactions was used. Emotion needed to be recognized in terms of continuous time, *and* continuous valued dimensional affect in three dimensions: valence, arousal and dominance. Depression needed to be estimated in terms of a widely accepted clinical self-report questionnaire, the Beck Depression Inventory II.

Besides participation in the Challenge we were calling for papers addressing the overall topics of this workshop, in particular works that address the differences between audio and video processing of emotive data, and the issues concerning combined audio-visual emotion recognition.

In the following sections, we will describe the participation in this year and outline the conditions for participation in

particular in the competitive challenge event. We further acknowledge those that helped turn AVEC 2014 into reality.

2. PARTICIPATION

The call for participation and papers attracted registrations of 42 teams (15 more than last year) and 142 team members from all over the world. 6 teams submitted results for the Affect recognition Sub-Challenge (ASC), 12 teams submitted results for the Depression recognition Sub-Challenge (DSC) — overall, at the moment of writing over 60 submissions of results to both Sub-Challenges combined were received (one third of these for the ASC). Finally, 21 paper submissions were received. AVEC 2014 reviewing was double blind, and acceptance was based on relevance to the workshop, novelty, and technical quality. The programme committee accepted 8 papers in addition to the baseline paper as oral presentation (the oral acceptance rate thus equalling 43%), and further 4 as posters, which were assigned three reviewers, each, and reviewed independently. Again, 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. CHALLENGE CONDITIONS

A baseline paper explaining the dataset, the challenge evaluation procedure, baseline features and baseline results was made available during the ongoing challenge [5]. Due to the sensitive nature of the data, a strict registration for access to the challenge data policy was followed. As in 2013, a web-based database was used, however, we required to sign a new end user license agreement and AVEC 2014 data was encrypted with a new password. After downloading the data participants could directly start their own experiments with the train and development sets. Once they found their best method they should write a paper for the Workshop. At the same time they could compute their results per instance of the test set. Participants' results needed to be sent as a single packed file per Sub-Challenge to the organisers by email and scores were returned within 24 hours during typical working days. As indicated, there were two goals of AVEC 2014: the recognition of the three continuous affect labels (ASC) and the estimation of the self-reported depression indicator BDI-II (BSC).

Each participant had up to five submission attempts per Sub-Challenge. Badly formatted results were not counted towards one of these five submissions. Further, the top-two performers of both Sub-Challenges were asked to submit their program to the organisers at University of Nottingham to verify the results, both on the original test set and extra hold-out data. Delivery was as an executable or, e.g., encrypted Matlab or similar code, and they were asked to work with the organisers in validating their results. The organisers provided for each dimension Pearson's correlation coefficient or the RMS error (for the ASC/DSC, respectively), which is used to rank participants. For the ASC the RMS error was provided as well, which could be used by the authors to further discuss their results in the paper accompanying their submission. Participants could download the Matlab evaluation code for the two Sub-Challenges. This code was not supposed to 'run' on one's machine (more files and the test labels would be needed for that); it is rather just a detailed description of how we calculate the scores.

4. PROGRAM AND COMMITTEE

The workshop is a full-day event held on 7 November 2014 starting with a keynote speech, followed by an introduction to the challenge, a series of paper presentations (oral and posters), a panel discussion, and finally an overview of the challenge results and an announcement of the winners of the two Sub-Challenges.

The organisers — the authors of this summary — would like to thank all participants and in particular also the highly dedicated program committee of this year:

Anton Batliner, FAU, Germany, *Felix Burkhardt*, Deutsche Telekom, Germany, *Rama Chellappa*, University of Maryland, USA, *Mohamed Chetouani*, ISIR, France, *Julien Epps*, UNSW, Australia, *Roland Göcke*, University of Canberra, Australia, *Hatice Gunes*, QMUL, UK, *Marc Méhu*, University of Geneva, Switzerland, *Louis-Philippe Morency*, USC, USA, *Richard Morriss*, University of Nottingham, UK, *Stefan Scherer*, USC, USA, *Stefan Steidl*, FAU, Germany and *Fernando de la Torre*, CMU, USA.

Acknowledgments

The work of Michel Valstar is partly funded by the NIHR-HTC 'MindTech' and Horizon Digital Economy Research, RCUK grant EP/G065802/1. The work of Jarek Krajewski is partly funded by the German Research Foundation (KR3698/4-1).

The challenge in general has been generously supported by the Association for the Advancement of Affective Computing (AAAC, former HUMAINE association) and the EU network of excellence on Social Signal Processing SSPNet (EC's 7th Framework Programme [FP7/20072013] under grant agreement no. 231287).

The authors further acknowledge funding from the EC and ERC (grants nos. 289021, ASC-Inclusion and 338164, iHEARu).

5. REFERENCES

- [1] A. Dhall, R. Goecke, J. Joshi, M. Wagner, and T. Gedeon. Emotion Recognition In The Wild Challenge 2013. In *Proceedings ICMI*, pages 509–516, Sydney, Australia, December 2013. ACM.
- [2] B. Schuller, M. Valstar, F. Eyben, R. Cowie, and M. Pantic. AVEC 2012 – The Continuous Audio/Visual Emotion Challenge. In *Proceedings ICMI*, pages 449–456, Santa Monica, CA, October 2012. ACM.
- [3] B. Schuller, M. Valstar, F. Eyben, G. McKeown, R. Cowie, and M. Pantic. AVEC 2011 - The First International Audio/Visual Emotion Challenge. In *Proceedings ACII*, volume II, pages 415–424. Springer, Memphis, TN, October 2011.
- [4] M. Valstar, B. Schuller, J. Krajewski, R. Cowie, and M. Pantic. Workshop summary for the 3rd international audio/visual emotion challenge and workshop (AVEC'13). In *Proceedings MM*, pages 1085–1086, Barcelona, Spain, October 2013. ACM.
- [5] M. Valstar, B. Schuller, K. Smith, T. Almaev, F. Eyben, J. Krajewski, R. Cowie, and M. Pantic. AVEC 2014 – The Three Dimensional Affect and Depression Challenge. In *Proceedings of the 4th ACM international workshop on Audio/Visual Emotion Challenge*, Orlando, FL, November 2014. ACM.