Editorial

Introduction To The Special Issue On Affect Analysis In Continuous Input

Human affective behavior is multimodal, continuous and complex. In day-to-day interactions people naturally communicate subtle affective states by means of language, vocal intonation, facial expression, hand gesture, head movement, body movement and posture, and possess a refined mechanism for understanding and interpreting information conveyed by these behavioral cues. Despite major advances within the affective computing research field, modeling, sensing, recognizing, interpreting and responding to such human affective behavior still remain a challenge for automated systems.

Accordingly, this Special Issue focuses on "*Affect Analysis in Continuous Input*" and aims to bring together a selected body of works investigating the issues and the challenges pertinent in sensing, recognizing and responding to continuous human affective behavior from diverse communicative cues and modalities, in both research and applied contexts.

Seventeen submissions were received for this Special Issue, out of which seven were accepted, and underwent on average two rounds of rigorous revisions. Only the opening article by the guest editors was handled through an independent review process by the Editors-in-Chief. The opening article "*Categorical and Dimensional Affect Analysis in Continuous Input: Current Trends and Future Directions*" aims at providing a definition of continuous affect analysis together with the latest developments and future perspectives in this field [1].

"*Tracking Continuous Emotional Trends of Participants during Affective Dyadic Interactions Using Body Language and Speech Information*" by Metallinou, Katsamanis, and Narayanan focuses on continuous analysis, in time and in value, of human body movements, posture and behavior together with speech information to predict a participant's activation, valence and dominance level in the context of affective dyadic interactions [2]. Extracted audio-visual cues are mapped onto the above-mentioned affect dimensions using a Gaussian Mixture Model-based approach, treating both time and affect variables as continuous. Promising results are obtained for tracking trends of participants' activation and dominance values. To date, most of the automatic affect analyzers in the literature have performed one-sided analysis by looking only at one party irrespective of the other party they interacted with [1]. Focusing on affect analysis in continuous input in the context of dyadic interactions [2] is an important step towards assessing reciprocity between the interacting parties that is crucial for obtaining a better understanding of the interaction taking place.

In "*LSTM-Modeling of Continuous Emotions in an Audiovisual Affect Recognition Framework*", Wöllmer, Kaiser, Eyben, Schuller, and Rigoll demonstrate the effective use of context-sensitive Long-Short Term Memory Recurrent Neural Networks on the Audio/Visual Emotion Challenge (AVEC) task [3]. They combine word-level acoustic, linguistic, and video features using a fully automatic audio-visual recognition approach. Linguistic features are based on real-time automatic speech recognition, and also include the non-linguistic vocalizations of breathing, laughing, and sighing. Visual features are based on a novel facial movement feature extractor. With 65.2% weighted average accuracy over the four dimensions of the Audiovisual Sub-challenge, namely, arousal, expectation, power, and valence, the proposed technique leads to the best average recognition performance that has been reported so far for this task.

In "*Fusion of Facial Expressions and EEG for Implicit Affective Tagging*", Koelstra and Patras demonstrate the practical use of multimodal affect analysis in continuous input for a relatively new application domain. The focus of the work is on implicit tagging by analyzing users' facial expression and Electroencephalography (EEG) responses during interaction with multimedia content to generate descriptive tags [4]. Tags are generated in terms of values along arousal, valence, and control dimensions by employing both binary classification and regression methods, as well as feature-level fusion and decision-level fusion. Classification and regression results are improved when using information from both modalities.

As pointed out in the opening article, dynamics of human behavior have implications for meaningful interpretation of sensor data and therefore need to be carefully considered when designing automatic affect analyzers [1]. The article on "*Recognizing Expressions from Face and Body Gesture by Temporal Normalized Motion and Appearance Features*" by Chen, Tian, and Metaxas [5] focuses on categorical recognition of affect by combining local motion and appearance features to model the temporal dynamics of face and body gesture. The work employs Histogram of Oriented Gradients (HOG) on the Motion History Image (MHI) and Image-HOG features, and implicitly models local temporal dynamics of an expression through a bag-of-words-based representation. Utilizing the face and body expression temporal dynamics, rather than using a frame-based approach, appears to contribute to improving the accuracy of automatic expression recognition.

As emphasized in the opening article, for creating and fine-tuning automatic analyzers, there seems to be an ever increasing need for data with high number of participants, data with participants from diverse backgrounds in terms of culture, personality, language, age, and speaking style, as well as data with richer and novel annotations [1]. Petridis, Martinez and Pantic's article entitled "*The MAHNOB Laughter Database*" aims to make a contribution in this direction [6]. The article introduces a new publicly available audiovisual database that contains recordings from 22 subjects, recorded using two microphones, a video camera and a thermal camera while watching stimulus material. Recordings constitute elicited laughter as well as posed smiles, posed laughter, and speech. Baseline experimental results for discrimination between laughter-vs-speech and between voiced laughter, unvoiced laughter and speech, are provided together with annotated 563 laughter episodes, 849 speech utterances, 51 posed laughs, 67 speech-laughs episodes and 167 other vocalizations.

The Special Issue concludes with "*Towards the Automatic Detection of Spontaneous Agreement and Disagreement based on Nonverbal Behavior: A Survey of Related Cues, Databases, and Tools*" by Bousmalis, Mehu, and Pantic that broadens the concept of affect analysis and constitutes a link towards social signal processing by focusing on conceptualizing and modeling the behavioral dimension of agreement [7]. The survey paper starts with the identification of nonverbal behavioral cues that could be present during displays of agreement and disagreement, compiles the existing methods and databases that could be used to detect these cues automatically, and examines the existing attempts at agreement and disagreement classification.

As guest editors, our hope is that the collection of papers in this Special Issue will provide a firmer foundation and further inspiration for realization of automatic, naturalistic, multimodal, and continuous affect and affective behavior analysis and understanding.

Acknowledgments

The guest editors of this Special Issue would like to thank all the authors for their submissions, and the contributors for the careful revision of their manuscripts and for making this issue a valuable source of information for the research community. They would also like to acknowledge the following reviewers for their invaluable time and effort: Oya Aran, Nadia Berthouze, George Caridakis, Ginevra Castellano, Felix Burkhardt, Carlos Busso, Sidney D'Mello, Thierry Dutoit, Donald Glowinski, Didem Gökcay, Hayrettin Gürkok, Zakia Hammal, Dirk Heylen, Hayley Hung, Ashish Kapoor, Jonghwa Kim, Sander Koelstra, Shrikanth Narayanan, Mihalis A. Nicolaou, Christopher Peters, Stavros Petridis, Massimo Piccardi, Albert A. Salah, Stefan Scherer, Caifeng Shan, Nicu Sebe, Mohammad Soleymani, Jianhua Tao, YingLi Tian, Yan Tong, Khiet Truong, Michel Valstar, Gualtiero Volpe, Steffen Walter, and Shangfei Wang. Many thanks go to the editorial team of Image and Vision Computing for their help and patience during the preparation process of this issue.

References

[1] H. Gunes, B. Schuller, Categorical and dimensional affect analysis in continuous input: Current trends and future directions, Image and Vision Computing, Special Issue on Affect Analysis in Continuous Input, Image Vision Comput. 31 (2) (2013) 120–136.

[2] A. Metallinou, A. Katsamanis, S. Narayanan, Tracking continuous emotional trends of participants during affective dyadic interactions using body language and speech information, Image and Vision Computing, Special Issue on Affect Analysis in Continuous Input, Image Vision Comput. 31 (2) (2013) 137–152.

[3] M. Wöllmer, M. Kaiser, F. Eyben, B. Schuller, G. Rigoll, Lstm-modeling of continuous emotions in an audiovisual affect recognition framework, Image and Vision Computing, Special Issue on Affect Analysis in Continuous Input, Image Vision Comput. 31 (2) (2013) 153–163.

[4] S. Koelstra, I. Patras, Fusion of facial expressions and EEG for implicit affective tagging, Image and Vision Computing, Special Issue on Affect Analysis in Continuous Input, Image Vision Comput. 31 (2) (2013) 164–174.

[5] S. Chen, Y. Li Tian, Q. Liu, D.N. Metaxas, Recognizing expressions from face and body gesture by temporal normalized motion and appearance features, Image and Vision Computing, Special Issue on Affect Analysis in Continuous Input, Image Vision Comput. 31 (2) (2013) 175–185.

[6] S. Petridis, B. Martinez, M. Pantic, The MAHNOB laughter database, Image and Vision Computing, Special Issue on Affect Analysis in Continuous Input, Image Vision Comput. 31 (2) (2013) 186–202.

[7] K. Bousmalis, M. Mehu, M. Pantic, Towards the automatic detection of spontaneous agreement and disagreement based on nonverbal behavior: A survey of related cues, databases, and tools, Image and Vision Computing, Special Issue on Affect Analysis in Continuous Input, Image Vision Comput. 31 (2) (2013) 203–221.

Hatice Gunes
School of Electronic Engineering and Computer Science,
Queen Mary University of London, UK
Corresponding author. Tel.: +44 207 882 5349.
E-mail address: hatice@eecs.qmul.ac.uk.

Björn Schuller
Institute for Human-Machine Communication,
Technische Universität München, Germany

5 November 2012