# Editorial of Special Issue on Human Behaviour Analysis "In-the-Wild"

Mihalis A. Nicolaou, Stefanos Zafeiriou, *Member, IEEE*, Irene Kotsia, *Member, IEEE*, Guoying Zhao, *Senior Member, IEEE*, and Jeffrey Cohn

## 1 Introduction

The human face and body are quite likely the most researched objects in image analysis, computer vision and signal processing. One of the main reasons behind this popularity lies in the numerous applications of automatic face and body gesture analysis algorithms, that span several fields such as Human-Computer and Human-Robot Interaction (facial expression/body gesture recognition for automatic analysis of affect), medicine and healthcare (detection of emotional and cognitive disorders), as well as biometrics (face recognition, gait recognition). Less than a decade ago, the majority of face and body analysis algorithms and systems were built and evaluated on databases captured in controlled experimental conditions, such as the FERET database for face recognition, the Cohn-Kanade and MMI databases for facial expression recognition and facial action unit detection, and the XM2VTS and BIO-ID datasets for facial landmark detection. Research has gradually shifted to the analysis of facial images captured in-the-wild with the introduction of the SEMAINE database, the Labelled Faces in-the Wild (LFW) dataset, the FDDB for face detection and 300-W series of databases for facial landmark localisation/tracking (similarly MPII database for body pose estimation). Currently, LFW is used by the majority of researchers as a benchmark for face recognition, FDDB for face detection, 300-W for facial landmark localisation and MPII for body joint estimation.

Shifting to the analysis of spontaneous facial expressions and body gestures recorded under uncontrolled settings is a crucial step towards realising the next generation of machines that can sense and interpret human emotional and social behaviour under arbitrary recording conditions. It nevertheless is more challenging, as uncontrolled settings entail the presence of artefacts or occlusions in the visual data, as well as data capturing spontaneous emotions contain much higher variability than posed emotions, with significant differences in terms of both temporal and spatial characteristics. In general, we can group research in automatic analysis of behaviour in terms of the variables that we are interested in predicting, listed in what follows: (i) recognizing a set of discrete expressions, usually confined to the recognition of the so-called six universal expressions (i.e., Anger, Disgust, Fear, Happiness, Sadness and Surprise) plus neutral, (ii) detecting particular non-universal expressions (e.g., recognition of pain and compound expressions), (iii) detectingf Facial Action Units (FAU) in expressive sequences, which relate to a standardised taxonomy of facial muscles' movement, as well as (iv) estimation of latent emotion dimensions, such as valence (how positive or negative an emotional state is), arousal (measuring the power of the activation of the emotion) and dominance (capturing sense of control over emotion).

In this special issue, we focus on recent efforts towards catalysing progress in automatic analysis of human behaviour in uncontrolled, "in-the-wild" conditions. We summarize research efforts towards the development of research methodologies, database collections and benchmarks, as well as algorithms and systems for machine analysis of human behaviour, focusing on facial expressions, body gestures, speech, as well as various other sensors. We are delighted that the special issue includes authors both from academia as well as the industry (Affectiva, Disney Research, STATS).

The special issue is organized as follows. In the paper "AM-FED+: An Extended Dataset of Naturalistic Facial Expressions Collected in Everyday Settings", D. McDuff, M. Amr and R. el Kaliouby present a public dataset containing naturalistic facial expressions collected in everyday settings. In more detail, the authors present a dataset containing more than 1000 videos, with half of them accompanied by comprehensive annotations for facial action units, as well as facial landmarks. The authors propose the utilization of this dataset as a benchmark for the evaluation of algorithms and models, while also providing a set of contextual labels as well as baseline results for action unit classification for the coded videos.

The next paper "AffectNet: A Database for Facial Expression, Valence and Arousal Computing in the


- *M.A. Nicolaou is with the Computation-based Science and Technology Research Center at the Cyprus Institute, 20 Konstantinou Kavafi Street 2121, Aglantzia, Nicosia, Cyprus. E-mail: m.nicolaou@cyi.ac.cy.*
- *S. Zafeiriou is with the Department of Computing, Imperial College Londonm 180 Queens Gate, London SW7 2AZ, United Kingdom. E-mail: s.zafeiriou@imperial.ac.uk.*
- *I. Kotsia is with the Computer Science Department, School of Science & Technology, Middlesex University, The Burroughs, Hendon, London NW4 4BT, United Kingdom. E-mail: I.Kotsia@mdx.ac.uk.*
- *G. Zhao is with the Center for Machine Vision and Signal Analysis, University of Oulu, Pentti Kaiteran katu 1, 90014 Oulu, Finland. E-mail: guoying.zhao@oulu.fi.*
- *J. Cohn is with the Department of Psychology, University of Pittsburgh 210 S. Bouquet Street, #4327, Pittsburgh, PA 15260. E-mail: jeffcohn@pitt.edu.*


*For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.*
*Digital Object Identifier no. 10.1109/TAFFC.2019.2895141*

Wild", A. Mollahosseini, B. Hasani and M. H. Mahoor present a large scale database, coined AffectNet, which contains more than 1 million facial images from the Internet using 1,250 emotion related keywords in six different languages. AffectNet is among the largest databases the provide images annotated with regards to facial expression, valence, and arousal in the wild. Two deep neural networks are used to estimate facial expressions, as well as valence and arousal.

The paper "Discriminative Spatiotemporal Local Binary Pattern with Revisited Integral Projection for Spontaneous Facial Micro-Expression Recognition", the authors X. Huang, S. Wang, X. Liu, G. Zhao, X. Feng and M. Pietikäinen present their work on detecting spontaneous facial micro-expressions from facial images. Building on previous work that incorporates spatiotemporal local binary patterns that consider dynamic texture information, the authors extend the descriptors based on an integral projection for preserving the shape-attribute of micro-expressions, using Robust Principal Component Analysis. Furthermore, a feature selection component based on the Laplacian method is introduced in order to capture discriminative information towards the recognition of facial micro-expressions, with experiments presented on several micro-expression databases, such as CASME, CASME2 and SMIC, showing very promising results.

In the paper "Estimating Audience Engagement to Predict Movie Ratings' by R. Navarathna, P. Carr, P. Lucey and I. Matthews, the authors address an interesting problem: estimating audience engagement through both subtle and coarse facial expressions and body gesture during feature length movies. This constitutes a very challenging setting, as the environment is dark during the movie, people appear at different scales and viewpoints, while movies typically last for a quite long duration. Furthermore, as the authors claim, facial expressions by the audience during the movie are short, subtle and sparse. The authors propose a method using infrared illuminated test-beds in order to detect the change in behaviour by identifying key-frames that capture audience sentiment. The authors use crow-sourced ratings to train a classifier, and use audience sentiment as a proxy for rating movies. The entire dataset consists of over 50 hours of audience behaviour, containing a large number of more than 200 subjects.

The paper "Audio-visual Emotion Recognition in Video Clips" by F. Noroozi, M. Marjanovic, A. Njegus, S. Escalera, and G. Anbarjafari presents an audio-visual approach for emotion recognition in videos. The authors extract acoustic features such as spectral and prosodic, and utilize a set of visual features on which geometric relationships are computed. Following the proposed approach, key-frames are detected in order to summarize each emotional video. Using convolutional networks, the authors end-up in a final representation that includes classifier confidence. Multiclass SVMs and Random Forest classifiers are subsequently trained on the dimensionality reduced representation, obtained by applying PCA. The authors perform extensive experiments on databases such as RML, SAVEE, and eNTERFACE'05.

Finally, the paper "Hidden Smile Correlation Discovery across Subjects using Random Walk with Restart" by H. Jiang, M. Coskun, A. Badokhon, M. Liu, and Ming-Chun Huang deals with the problem of fine-grained smile analysis. In particular, the authors propose a fine-grained smile analysis system that utilizes the head pose of the subject, and employs conditional random forests in order to handle natural head motion and orientations changes, and eventually detect fiducial points in facial images. By using random walks, the authors then present results in terms of identifying smile intensity, and uncover hidden correlations in smile patterns across subjects. The method is evaluated on datasets such as the UvA-NEMO and a selection of data from the Labelled Faces in-the-wild dataset.

We hope that this special issue has provided an opportunity for bringing together experts from both academia and the industry, towards working on several open research problems in a challenging area. The selection of papers represents a good overview of such problems, and reviews the state-of-the-art as well as various datasets that can be used for evaluating new methods.

## ACKNOWLEDGMENTS

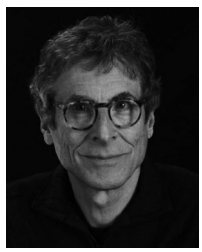
This special issue would not have been possible without the efforts and interests from all the authors who contributed their submissions. We would like to take this opportunity to thank them. We are grateful to the reviewers for their careful and valuable comments on the submitted manuscripts and for their detailed and helpful suggestions for improvement. S. Zafeiriou and G. Zhao also acknowledge support from Tekes Fidipro Program (Grant No.1849/31/2015)

**Mihalis A. Nicolaou** received the PhD degree from the Department of Computing, Imperial College London, under an EPSRC DTA, where he remained as research associate and subsequently an honorary research fellow. He is an assistant professor with the Computation-based Science and Technology Research Center, Cyprus Institute, and a Visiting Academic at Goldsmiths, University of London. His research interests span the areas of machine learning, signal processing, and computer vision, focusing on analysis and interpretation of multi-sensory high dimensional data, often conveyed via visual, auditory, social, and biomedical signals. He has received several awards for his research, including best paper awards at FG 11 and ICASSP 16. He has co-organized several workshops in his area in top venues such as CVPR, while he has been a guest associate editor at the *IEEE Transactions on Affective Computing*.

**Stefanos Zafeiriou** (M09) is currently a reader in machine learning and computer vision with the Department of Computing, Imperial College London, London, United Kingdom, a distinguishing research fellow with the University of Oulu under Finish Distinguishing Professor Programme. He was a recipient of the Prestigious Junior Research Fellowships from Imperial College London in 2011 to start his own independent research group. He was the recipient of the Presidents Medal for Excellence in Research Supervision for 2016. He is recipient of many best paper awards including the best student paper award in FG2018. In 2018, he received an EPSRC Fellowship. He currently serves as an associate editor of the *IEEE Transactions on Affective Computing* and the *Computer Vision and Image Understanding Journal*. In the past he held editorship positions in the *IEEE Transactions on Cybernetics* and the *Image and Vision Computing Journal*. He has been a guest editor of more than six journal special issues and co-organised more than 13 workshops/special sessions on specialised computer vision topics in top venues, such as CVPR/FG/ICCV/ECCV (including three very successfully challenges run in ICCV13, ICCV15, CVPR17 and ICCV'17 on facial landmark localisation/tracking). He has co-authored more than 65 journal papers mainly on novel statistical machine learning methodologies applied to computer vision problems, such as 2-D/3-D face analysis, deformable object fitting and tracking, shape from shading, and human behaviour analysis, published in the most prestigious journals in his field of research, such as the *IEEE Transactions on Pattern Analysis and Machine Intelligence*, the *International Journal of Computer Vision*, the *IEEE Transactions on Image Processing*, the *IEEE Transactions on Neural Networks and Learning Systems*, the *IEEE Transactions on Visualization and Computer Graphics*, and the *IEEE Transactions on Information Forensics and Security*, and many papers in top conferences, such as CVPR, ICCV, ECCV, ICML. His students are frequent recipients of very prestigious and highly competitive fellowships, such as the Google Fellowship x2, the Intel Fellowship, and the Qualcomm Fellowship x3. He has more than 7,000 citations to his work, h-index 44. He was the general chair of BMVC 2017. He is a member of the IEEE.

**Irene Kotsia** (M'09) received the PhD degree from the Department of Informatics, Aristotle University of Thessaloniki, Thessaloniki, Greece, in 2008. From 2008 to 2009, she was a research associate and teaching assistant with the Department of Informatics, Aristotle University of Thessaloniki. From 2009 to 2011, she was a research associate with the Department of Electronic Engineering and Computer Science, Queen Mary University of London, while from 2012 to 2014, she was a senior research associate with the Department of Computing, Imperial College London. From 2013 to 2015, she was a lecturer in creative technology and digital creativity with the Department of Computing Science, Middlesex University of London, where she is currently a senior lecturer. She has been a guest editor of two journal special issues dealing with face analysis topics. She has co-authored more than 40 journal and conference publications in the most prestigious journals and conferences of her field (e.g., the *IEEE Transactions on Image Processing*, the *IEEE Transactions on Neural Networks and Learning Systems*, CVPR, ICCV). She has published one of the most influential works in facial expression recognition in the *IEEE Transactions on Image Processing* which has received around 600 citations. She is a member of the IEEE.

**Guoying Zhao** (SM'17) received the PhD degree in computer science from the Chinese Academy of Sciences, Beijing, China, in 2005. She is currently a professor with the Center for Machine Vision and Signal Analysis, University of Oulu, Finland, where she has been a senior researcher since 2005 and an associate professor since 2014. In 2011, she was selected to the highly competitive Academy Research Fellow position. She was Nokia visiting professor in 2016. She has authored or co-authored more than 180 papers in journals and conferences. Her papers have currently more than 7,900 citations in Google Scholar (h-index 39). She is co-publicity chair for FG2018, has served as area chairs for several conferences and is associate editor of the *Pattern Recognition*, the *IEEE Transactions on Circuits and Systems for Video Technology*, and the *Image and Vision Computing Journals*. She has lectured tutorials at ICPR 2006, ICCV 2009, SCIA 2013 and FG 2018, authored/edited three books and eight special issues in journals. She was a co-chair of many International Workshops at ECCV, ICCV, CVPR, ACCV and BMVC. Her current research interests include image and video descriptors, facial-expression and micro-expression recognition, gait analysis, dynamic-texture recognition, human motion analysis, and person identification. Her research has been reported by Finnish TV programs, newspapers, and MIT Technology Review. She is a senior member of the IEEE.



**Jeffrey Cohn** is professor of Psychology, Psychiatry, and Intelligent Systems, University of Pittsburgh and adjunct faculty with the Robotics Institute, Carnegie Mellon University. He has led interdisciplinary and inter-institutional efforts to develop advanced methods of automatic analysis and synthesis of facial expression and other nonverbal behavior. He has applied those tools to research in human emotion, interpersonal processes, social development, and psychopathology. He has co-developed and distributed influential databases, including Cohn-Kanade, UNBC Pain Archive, and Binghamton-Pittsburgh 4D (BP4D and BP4D+), and has chaired international conferences in automatic face and gesture recognition, multimodal interaction, and affective computing.