

Accepted Manuscript

Editorial

Visual and Multimodal Analysis of Human Spontaneous Behavior: Introduction to the Special Issue of Image & Vision Computing Journal

Maja Pantic, Jeffrey F. Cohn

PII: S0262-8856(09)00150-4

DOI: [10.1016/j.imavis.2009.07.001](https://doi.org/10.1016/j.imavis.2009.07.001)

Reference: IMAVIS 2902

To appear in: *Image and Vision Computing*



Please cite this article as: M. Pantic, J.F. Cohn, Visual and Multimodal Analysis of Human Spontaneous Behavior: Introduction to the Special Issue of Image & Vision Computing Journal, *Image and Vision Computing* (2009), doi: [10.1016/j.imavis.2009.07.001](https://doi.org/10.1016/j.imavis.2009.07.001)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Visual and Multimodal Analysis of Human Spontaneous Behavior: Introduction to the Special Issue of Image & Vision Computing Journal

Maja Pantic, Imperial College London UK / University of Twente, NL
Jeffrey F. Cohn, University of Pittsburgh / Carnegie Mellon University, USA

Widely anticipated in HCI is that computing will move to the background, weaving itself into the fabric of our everyday living and projecting the human user into the foreground. To realize this goal, next-generation computing (a.k.a. pervasive computing, ambient intelligence, and human computing) will need to develop human-centered user interfaces that respond readily to naturally occurring, multimodal, human communication. These interfaces will need the capacity to perceive, understand, and respond appropriately to human intentions and cognitive-emotional states as communicated by social and affective signals.

Motivated by this vision of the future, automated analysis of nonverbal behavior has attracted increasing attention in diverse disciplines, including psychology, computer science, linguistics, and neuroscience. Promising approaches have been reported, especially in the areas of facial expression and multimodal communication. Until recently, much of this work has focused on posed, often exaggerated expressions (for reviews, see (Pantic & Rothkrantz, 2003; Tian et al., 2005; Zeng et al., 2009)). Yet, increasing evidence suggests that deliberate or posed behaviour differs in appearance and timing from that which occurs in daily life. For example, brow raises have larger amplitude, faster onset, and shorter duration when posed than when not (Schmidt et al., 2009). The morphology of facial actions differs as well. As (Littlewort, et al., 2009) in this issue report, facial actions systematically differ between spontaneous and feigned pain. Approaches to automatic behaviour analysis that have been trained on deliberate and typically exaggerated behaviours may fail to generalize to the complexity of expressive behaviour found in real-world settings.

This Special Issue of *Image and Vision Computing* brings together cutting edge work on the automatic analysis of non-posed, real-world human behavior. It includes state-of-the-art reviews of computational approaches to conversation analysis (Gatica-Perez, 2009) and social signal processing (Vinciarelli et al., 2009), recent advances in generic face modeling (Lucey et al., 2009), automated detection of pain from facial behavior (Littlewort et al., 2009) (Ashraf et al., 2009), cognitive states of interest from facial, vocal, and gestural behavior (Schuller et al., 2009), automatic detection of diverse human activities from spatio-temporal features (Oikonomopoulos et al., 2009), and automatic recognition of American Sign Language (Ding & Martinez, 2009). These papers represent an exciting advance toward human-centered interfaces that can perceive and understand real-world human behavior.

Of course, as discussed by Zeng et al. (2009) and by Vinciarelli et al. (2009) and Gatica-Perez (2009) in this issue, significant scientific and technical challenges remain to be addressed. However, we are optimistic about the continued progress. A principal reason is that automatic

multimodal analysis of human naturalistic behavior is prerequisite in achieving next-generation, human-centered computing (Jaimes et al., 2006; Pantic et al., 2006, 2009); and this topic is poised to become one of the most active research topics in the computer vision and signal processing communities. To support these efforts, infrastructure is emerging from the extensive efforts of investigators, international sponsors, and professional societies. A sampling of research activities includes basic research on machine analysis of human behavior (e.g., European Research Council (ERC) MAHNOB project), automatic analysis of face-to-face and small group interactions (e.g., see the projects of MIT Human Dynamics Laboratory and European Commission (EC) FP6 AMIDA project), social signaling (e.g., EC FP7 Social Signal Processing NoE project), human-computer interactions (e.g., EC FP7 Semaine project), applications in mental health (Cohn et al., 2009), and other areas. The contributions in this Special Issue highlight recent advances and point to continued progress toward the goal of human-centered interfaces that can understand human intentions and behavior and respond intelligently.

Acknowledgments

The Guest Editors of this Special Issue would like to thank the reviewers who have volunteered their time to provide valuable feedback to the authors. They would also like to thank the contributors for making this issue an important asset to the existing body of literature in the field. Many thanks go to Image & Vision Computing Journal editorial support for their help during the preparation of this issue.

References

- Ashraf, A. B., Lucey, S., Cohn, J. F., Chen, T., Prkachin, K. M., & Solomon, P. (2009). The painful face: Pain expression recognition using active appearance models. *Image and Vision Computing*, this issue.
- Cohn, J. F., Simon Kreuz, T., Matthews, I., Yang, Y., Nguyen, M. H., Tejera Padilla, M., et al. (2009). Detecting depression from facial actions and vocal prosody. *Proceedings Int'l Conf. Affective Computing and Intelligent Interaction*.
- Ding, L., & Martinez, A. M. (2009). Modelling and recognition of the linguistic components in American Sign Language. *Image and Vision Computing*, this issue.
- Gatica-Perez, D. (2009). Automatic nonverbal analysis of social interaction in small groups: A review. *Image and Vision Computing*, this issue.
- Jaimes, A., Sebe, N., & Gatica-Perez, D. (2006). Human-Centered Computing: A Multimedia Perspective. *Proceedings ACM Int'l Conf. Multimedia*, 855-864.
- Littlewort, G. C., Bartlett, M. S., & Lee, K. (2009). Automatic coding of facial expressions displayed during posed and genuine pain. *Image and Vision Computing*, this issue, 2009.
- Lucey, S., Wang, Y., Cox, M., Sridharan, S., & Cohn, J. F. (2009). Efficient constrained local model fitting for non-rigid face alignment. *Image and Vision Computing Journal*, this issue.
- Oikonomopoulos, A., Pantic, M., & Patras, I. (2009). Sparse B-spline Polynomial Descriptors for Human Activity Recognition. *Image and Vision Computing*, this issue.
- Pantic, M., Pentland, A., Nijholt, A., & Huang, T.S. (2006). Human Computing and machine understanding of human behavior: A Survey, *Proceedings ACM Int'l Conf. Multimodal Interfaces*, pp. 239-248.

- Pantic, M., Pentland, A., & Nijholt, A. (2009). Guest Editorial Special Issue on Human Computing. *IEEE Transactions on Systems, Man and Cybernetics – Part B*, 39(1), 3-93.
- Pantic, M., & Rothkrantz, M. (2003). Toward an affect-sensitive multimodal human-computer interaction. *Proceedings of the IEEE*, 91, 1371-1390.
- Schmidt, K. L., Bhattacharya, S., & Denlinger, R. (2009). Comparison of deliberate and spontaneous facial movement in smiles and eyebrow raises. *Journal of Nonverbal Behavior*, 33, 35–45.
- Schuller, B., Muller, R., Eyben, F., Gast, J., Hornler, B., Wollmer, M., et al. (2009). Being bored? Recognising natural interest by extensive audiovisual integration for real-life application. *Image and Vision Computing*, this issue.
- Tian, Y., Cohn, J. F., & Kanade, T. (2005). Facial expression analysis. In S. Z. Li & A. K. Jain (Eds.), *Handbook of face recognition* (pp. 247-276). New York, New York: Springer.
- Vinciarelli, A., Pantic, M., & Boulard, H. (2009). Social signal processing: Survey of an emerging domain. *Image and Vision Computing*, this issue.
- Zeng, Z., Pantic, M., Roisman, G.I., & Huang, T.S. (2009). A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions. *IEEE Transactions of Pattern Analysis and Machine Intelligence*, 31(1), 39-58.