



Sophia Bano, Jianguo Zhang, Stephen J. McKenna, {s.bano, j.n.zhang, s.j.z.mckenna}@dundee.ac.uk Computer Vision and Image Processing Group, Computing, School of Science and Engineering, University of Dundee, United Kingdom

1. Introduction

Focused Interaction (FI)

• Co-present individuals, having mutual focus of attention, interact by establishing face-to-face engagement and direct conversation [1]

Hypothesis

• Fusion of multimodal features can improve the overall FI detection

Challenges

- Face-to-face engagement often not maintained
- Conversational partner not always present in the video frame
- Varying illumination
- Varied scenes

Examples from our Focused Interaction Dataset



FI in which conversational partners are in the field of view of the camera



while walking



An outdoor night-time FI scenario with weak visual cues due to low illumination

Existing methods

• Off-line processing of video clips or photo streams captured in quite constrained conditions and interacting people always in view [2, 3, 4]

4. Conclusion and Future Work

- Automatic online classification of Focused Interaction in continuous, egocentric videos
- Multimodal features: face track, VAD and camera motion profile
- Best performance with multimodal feature fusion and SVM with RBF kernel
- Future work involves the use of recurrent neural network for classification and to extend this work to identify conversational partners

Finding Time Together: Detection and Classification of Focused Interaction in Egocentric Video



References

- [1] E. Goffman. Encounters: Two studies in the sociology of interaction. Bobbs-Merrill, 1961.
- [2] M. Aghaei, M. Dimiccoli, P. Radeva. With whom do I interact? Detecting social interactions in egocentric photostreams. IEEE ICPR, 2016.
- [3] S. Alletto, G. Serra, S. Calderara, F. Solera, R. Cucchiara. From ego to nos-vision: Detecting social relationships in first-person views. IEEE CVPRW, 2014.
- [4] A. Fathi, J. K. Hodgins, J. M. Rehg. Social interactions: A first-person perspective. IEEE CVPR, 2012.
- [5] R. Chaudhry, A. Ravichandran, G. Hager, R. Vidal. Histograms of oriented optical flow and Binet-Cauchy kernels on nonlinear dynamical systems for the recognition of human actions. IEEE CVPR, 2009.
- of human voice. INTERSPEECH, 2013.

[6] M. Van Segbroeck, A. Tsiartas, S. Narayanan. A robust frontend for VAD: exploiting contextual, discriminative and spectral cues

Acknowledgements

This work is supported by the UK Engineering and Physical Sciences Research Council (EPSRC) under grant Engineering and Physical Sciences EP/N014278/1: ACE-LP: Augmenting Research Council Communication using Environmental Data to drive Language Prediction.





