Human Visual Behaviour for Collaborative Human-Machine Interaction

Andreas Bulling
Perceptual User Interfaces Group
Max Planck Institute for Informatics
Saarbrücken, Germany
bulling@mpi-inf.mpg.de

ACM 978-1-4503-3575-1/15/09...$15.00.
http://dx.doi.org/10.1145/2800835.2815378

Abstract
Non-verbal behavioural cues are fundamental to human communication and interaction. Despite significant advances in recent years, state-of-the-art human-machine systems still fall short in sensing, analysing, and fully "understanding" cues naturally expressed in everyday settings. Two of the most important non-verbal cues, as evidenced by a large body of work in experimental psychology and behavioural sciences, are visual (gaze) behaviour and body language. We envision a new class of collaborative human-machine systems that fully exploit the information content available in non-verbal human behaviour in everyday settings through joint analysis of human gaze and physical behaviour.

ACM Classification Keywords
H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous

Introduction
Human interactions are complex, adaptive to the situation at hand, and rely to a large extend on non-verbal behavioural cues. However, state-of-the-art human-machine systems still fall short in fully exploiting such cues. Despite significant advances in recent years, current human-machine systems typically use cues and sensing modalities in isolation, only cover a limited and

Figure 1: Sample images from our MPIIGaze dataset showing the considerable variability in terms of place and time of recording, directional light and shadows.

coarse set of human behaviours, and methods to perceive and learn from behavioural cues are mainly developed and evaluated in controlled settings.

We envision a new class of human-machine systems that fully exploit the information content available in natural non-verbal human behaviour in everyday settings through joint analysis of multiple behavioural cues. Multimodal analysis of non-verbal cues has significant potential and will pave the way for a new class of symbiotic human-machine systems that offer human-like perceptual and interactive capabilities. Human gaze and body language are particularly compelling cues given that a large body of work in the behavioural sciences has shown that these cues are rich sources of information and therefore most promising for realising our vision.

This vision poses three key research challenges. The first challenge is *human behaviour sensing*, i.e. the development of computational methods to unobtrusively, robustly, and accurately estimate gaze and body movement in daily-life settings. The second challenges is *computational human behaviour analysis*, i.e. the development of machine learning methods to analyse, understand, and learn from visual and physical behavioural cues and that cope with the significant variability and subtleness of human non-verbal behaviour. The third challenge is how to exploit and apply information on non-verbal cues in *symbiotic human-machine vision systems*. In the following we will provide an overview of our previous and ongoing efforts to address these challenges.

Human Behaviour Sensing

Our efforts to advance the state of the art in human behaviour sensing have so far mainly focused on visual behaviour, i.e. mobile and remote gaze estimation. These efforts have been driven by the vision of pervasive gaze estimation, i.e. unobtrusive and continuous gaze estimation in unconstrained daily-life settings [1]. More specifically, we have developed new mobile eye trackers based on Electrooculography and computer vision [3, 10]. These systems have significantly pushed the boundaries of mobile gaze estimation with respect to wearability and recording duration as well as accessibility, affordability, and extensibility. We have also presented methods for eye tracker self-calibration and for adapting calibrations to new users [8, 16]. We have also developed computer vision methods for remote gaze estimation using monocular RGB cameras to open up new usage scenarios and problem settings for gaze interaction. More specifically, we have developed methods for calibration-free gaze estimation on single and across multiple ambient displays [25, 26, 11] as well as for gaze estimation using the cameras readily integrated into handheld portable devices, such as tablets [23].

More recently, we have started to explore appearance-based methods that directly learn a mapping from eye appearance to on-screen or 3D gaze position. These methods have a number of appealing properties that make them particularly promising for use in daily-life settings, such as increased robustness to varying illumination conditions and camera resolutions. Specifically, we have presented a new large-scale dataset that we collected during everyday laptop use over more than three months and that is significantly more variable than existing ones with respect to appearance and illumination (see Figure 1). We have also presented a method based on a multimodal deep convolutional neural network for in-the-wild appearance-based gaze estimation that significantly outperforms previous methods [24].

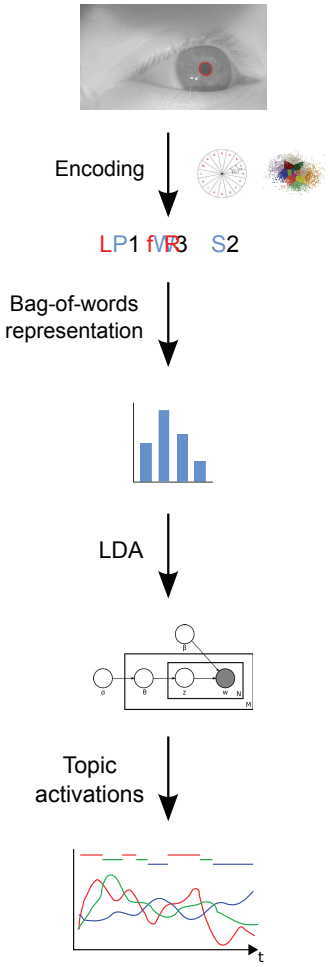

Figure 2: Unsupervised discovery of everyday activities from visual behaviour using a latent Dirichlet allocation (LDA) topic model.

Currently, we are exploring learning-by-synthesis approaches in which we use a large number of synthetic, photo-realistic eye images for pre-training the network. We have demonstrated that this approach significantly out-performs state-of-the-art methods for eye-shape registration as well as our own previous results for appearance-based gaze estimation in the wild [22].

Computational Human Behaviour Analysis

Activity recognition, and in particular recognition of users' activities from their visual behaviour, has been a focus of our research since several years. In early work we have shown, for the first time, that everyday activities, such as reading or common office activities, can be predicted in both stationary and mobile settings with surprising accuracy from eye movement data alone [5, 4]. Eye movements are closely linked to visual information processing, such as perceptual learning and experience, visual search, or fatigue. More recently we have therefore explored eye movement analysis as a similarly promising approach towards cognition-aware computing: Computing systems that sense and adapt to covert aspects of user state that are difficult if not impossible to detect using other modalities [7]. We have been among the first to demonstrate that selected cognitive states and processes can automatically be predicted from eye movement, such as visual memory recall [2], concentration [17], or personality traits, such as perceptual curiosity [9].

Despite significant advances in analysing and understanding human visual behaviour, the majority of previous works have focused on short-term behaviour lasting only seconds or minutes. We have contributed the first work on supervised recognition of high-level contextual cues, such as social interactions or being in or outside, from long-term visual behaviour [6]. Recently, we have extended that work with a new method for unsupervised discovery of everyday activities [15]. We have presented a method that combines a bag-of-words representation of visual behaviour with a latent Dirichlet allocation (LDA) topic model (see Figure 2) as well as a novel long-term gaze dataset that contains full-day recordings of natural visual behaviour of 10 participants (more than 80 hours in total).

Symbiotic Human-Machine Vision Systems

We have been studying human-computer interaction using gaze for several years. Starting from more classical problem settings in gaze-based interaction, such as interaction techniques for object selection, manipulation, and transfer across display boundaries [19, 18]. We have been particularly interested in extending the scope of gaze interaction into everyday settings and in making these interactions more natural. For example, we have introduced smooth pursuit eye movements – the movements we perform when latching onto a moving object – as a natural and calibration-free interaction technique for dynamic interfaces [21, 12]. We have also proposed social gaze as a new paradigm for designing user interfaces that react to gaze and eye contact as a form of non-verbal communication in a similar way as humans [20]. While gaze has a long history as a modality in human-computer interaction, in these works we have taken a fresh look at it and have demonstrated that there is much more to gaze than traditional but limited on-screen gaze location and dwelling.

State-of-the-art computer vision systems still under-perform on many visual tasks when compared to humans. We believe that collaborative vision systems that combine the advantages of machine and human perception and reasoning can bridge this performance gap.

We have recently started to explore both directions of collaborative vision, i.e. means of improving performance of computer vision algorithms by incorporating information from human fixations and vice versa. More specifically, we have propose an early integration approach of human fixation information into a deformable part model (DPM) for object detection [14]. We have demonstrated that our GazeDPM method outperforms state-of-the-art DPM baselines and that it provides introspection of the learnt models, can reveal salient image structures, and allows us to investigate the interplay between gaze attracting and repelling areas. In another work we have focused on predicting the target of visual search from human fixations [13]. In contrast to previous work we have studied a challenging open-world setting in which we no longer assumed that we have fixation data to train for the search targets. Both of these works as well as an increasing number of works by others in computer vision and machine learning point at the significant potential of integrating human and machine vision symbiotically.

Conclusion

In this work we have motivated the importance and potential of non-verbal behavioural cues, in particular gaze and body language, as a trailblazer for a new class of collaborative human-machine systems that are highly interactive, multimodal, and modelled after natural human-human interactions. We have outlined three key research challenges for realising this vision in unconstrained everyday settings: pervasive visual and physical human behaviour sensing, computational human behaviour analysis, and symbiotic human-machine vision systems. We have provided an overview of our previous and ongoing efforts to address these challenges, and we have highlighted individual works that represent and illustrate the state of the art in the respective area.

REFERENCES

1. Bulling, A., and Gellersen, H. Toward Mobile Eye-Based Human-Computer Interaction. *IEEE Pervasive Computing 9*, 4 (2010), 8–12.
2. Bulling, A., and Roggen, D. Recognition of Visual Memory Recall Processes Using Eye Movement Analysis. In *Proc. UbiComp* (2011), 455–464.
3. Bulling, A., Roggen, D., and Tröster, G. Wearable EOG goggles: Seamless sensing and context-awareness in everyday environments. *Journal of Ambient Intelligence and Smart Environments 1*, 2 (2009), 157–171.
4. Bulling, A., Ward, J. A., and Gellersen, H. Multimodal Recognition of Reading Activity in Transit Using Body-Worn Sensors. *ACM Transactions on Applied Perception 9*, 1 (2012), 2:1–2:21.
5. Bulling, A., Ward, J. A., Gellersen, H., and Tröster, G. Eye Movement Analysis for Activity Recognition Using Electrooculography. *IEEE Transactions on Pattern Analysis and Machine Intelligence 33*, 4 (Apr. 2011), 741–753.
6. Bulling, A., Weichel, C., and Gellersen, H. Eyecontext: Recognition of high-level contextual cues from human visual behaviour. In *Proc. CHI* (2013), 305–308.
7. Bulling, A., and Zander, T. O. Cognition-aware computing. *IEEE Pervasive Computing 13*, 3 (July 2014), 80–83.
8. Fehringer, B., Bulling, A., and Krüger, A. Analysing the potential of adapting head-mounted eye tracker calibration to a new user. In *Proc. ETRA* (2012), 245–248.

9. Hoppe, S., Loetscher, T., Morey, S., and Bulling, A. Recognition of Curiosity Using Eye Movement Analysis. In *Adj. Proc. UbiComp* (2015).

10. Kassner, M., Patera, W., and Bulling, A. Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction. In *Adj. Proc. UbiComp* (2014), 1151–1160.

11. Lander, C., Gehring, S., Krüger, A., Boring, S., and Bulling, A. GazeProjector: Accurate Gaze Estimation and Seamless Gaze Interaction Across Multiple Displays. In *Proc. UIST* (2015).

12. Pfeuffer, K., Vidal, M., Turner, J., Bulling, A., and Gellersen, H. Pursuit calibration: Making gaze calibration less tedious and more flexible. In *Proc. UIST* (2013), 261–270.

13. Sattar, H., Müller, S., Fritz, M., and Bulling, A. Prediction of search targets from fixations in open-world settings. In *Proc. CVPR* (2015), 981–990.

14. Shcherbatyi, I., Bulling, A., and Fritz, M. GazeDPM: Early Integration of Gaze Information in Deformable Part Models. arxiv:1505.05753, 2015.

15. Steil, J., and Bulling, A. Discovery of everyday human activities from long-term visual behaviour using topic models. In *Proc. UbiComp* (2015).

16. Sugano, Y., and Bulling, A. Self-calibrating head-mounted eye trackers using egocentric visual saliency. In *Proc. UIST* (2015).

17. Tessendorf, B., Bulling, A., Roggen, D., Stiefmeier, T., Feilner, M., Derleth, P., and Tröster, G. Recognition of hearing needs from body and eye movements to improve hearing instruments. In *Proc. Pervasive* (2011), 314–331.

18. Turner, J., Alexander, J., Bulling, A., and Gellersen, H. Gaze+rst: Integrating gaze and multitouch for remote rotate-scale-translate tasks. In *Proc. CHI* (2015), 4179–4188.

19. Turner, J., Bulling, A., Alexander, J., and Gellersen, H. Cross-device gaze-supported point-to-point content transfer. In *Proc. ETRA* (2014), 19–26.

20. Vidal, M., Bismuth, R., Bulling, A., and Gellersen, H. The Royal Corgi: Exploring Social Gaze Interaction for Immersive Gameplay. In *Proc. CHI* (2015), 115–124.

21. Vidal, M., Bulling, A., and Gellersen, H. Pursuits: Spontaneous interaction with displays based on smooth pursuit eye movement and moving targets. In *Proc. UbiComp* (2013), 439–448.

22. Wood, E., Baltrusaitis, T., Zhang, X., Sugano, Y., Robinson, P., and Bulling, A. Rendering of eyes for eye-shape registration and gaze estimation. arxiv:1505.05916, 2015.

23. Wood, E., and Bulling, A. Eyetab: Model-based gaze estimation on unmodified tablet computers. In *Proc. ETRA* (2014), 207–210.

24. Zhang, X., Sugano, Y., Fritz, M., and Bulling, A. Appearance-based gaze estimation in the wild. In *Proc. CVPR* (2015), 4511–4520.

25. Zhang, Y., Bulling, A., and Gellersen, H. Sideways: A gaze interface for spontaneous interaction with situated displays. In *Proc. CHI* (2013), 851–860.

26. Zhang, Y., Chong, M. K., Müller, J., Bulling, A., and Gellersen, H. Eye tracking for public displays in the wild. *Personal and Ubiquitous Computing* (2015), 1–15.