

Eyewear Computing – Augmenting the Human with Head-Mounted Wearable Assistants

Edited by

Andreas Bulling¹, Ozan Cakmakci², Kai Kunze³, and James M. Rehg⁴

1 Max-Planck-Institut für Informatik – Saarbrücken, DE, bulling@mpi-inf.mpg.de

2 Google Inc. – Mountain View, US, ozancakmakci@google.com

3 Keio University – Yokohama, JP, kai@kmd.keio.ac.jp

4 Georgia Institute of Technology – Atlanta, US, rehg@gatech.edu

Abstract

The seminar was composed of workshops and tutorials on head-mounted eye tracking, egocentric vision, optics, and head-mounted displays. The seminar welcomed 30 academic and industry researchers from Europe, the US, and Asia with a diverse background, including wearable and ubiquitous computing, computer vision, developmental psychology, optics, and human-computer interaction. In contrast to several previous Dagstuhl seminars, we used an ignite talk format to reduce the time of talks to one half-day and to leave the rest of the week for hands-on sessions, group work, general discussions, and socialising. The key results of this seminar are 1) the identification of key research challenges and summaries of breakout groups on multimodal eyewear computing, egocentric vision, security and privacy issues, skill augmentation and task guidance, eyewear computing for gaming, as well as prototyping of VR applications, 2) a list of datasets and research tools for eyewear computing, 3) three small-scale datasets recorded during the seminar, 4) an article in ACM Interactions entitled “Eyewear Computers for Human-Computer Interaction”, as well as 5) two follow-up workshops on “Egocentric Perception, Interaction, and Computing” at the European Conference on Computer Vision (ECCV) as well as “Eyewear Computing” at the ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp).

Seminar January 24–29, 2016 – <http://www.dagstuhl.de/16042>

1998 ACM Subject Classification H.5 Information Interfaces and Presentation (e.g., HCI), I.4 Image Processing and Computer Vision, K.4 Computer and Society

Keywords and phrases Augmented Human, Cognition-Aware Computing, Wearable Computing, Egocentric Vision, Head-Mounted Eye Tracking, Optics, Displays, Human-Computer Interaction, Security and Privacy

Digital Object Identifier 10.4230/DagRep.6.1.160



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 3.0 Unported license

Eyewear Computing – Augmenting the Human with Head-mounted Wearable Assistants, *Dagstuhl Reports*, Vol. 6, Issue 1, pp. 160–206

Editors: Andreas Bulling, Ozan Cakmakci, Kai Kunze, and James M. Rehg



DAGSTUHL
REPORTS Dagstuhl Reports

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Executive Summary

Andreas Bulling

Ozan Cakmakci

Kai Kunze

James M. Rehg

License © Creative Commons BY 3.0 Unported license

© Andreas Bulling, Ozan Cakmakci, Kai Kunze, and James M. Rehg

Main reference A. Bulling, K. Kunze, “Eyewear Computers for Human-Computer Interaction”, ACM Interactions, 23(3):70–73, 2016.

Computing devices worn on the human body have a long history in academic and industrial research, most importantly in wearable computing, mobile eye tracking, and mobile mixed and augmented reality. In contrast to traditional systems, body-worn devices are always with the user and therefore have the potential to perceive the world and reason about it from the user’s point of view. At the same time, given that on-body computing is subject to ever-changing usage conditions, on-body computing also poses unique research challenges.

This is particularly true for devices worn on the head. As humans receive most of their sensory input via the head, it is a particularly interesting body location for simultaneous sensing and interaction as well as cognitive assistance. Early egocentric vision devices were rather bulky, expensive, and their battery lifetime severely limited their use to short durations of time. Building on existing work in wearable computing, recent commercial egocentric vision devices and mobile eye trackers, such as Google Glass, PUPIL, and J!NS meme, pave the way for a new generation of “smart eyewear” that are light-weight, low-power, convenient to use, and increasingly look like ordinary glasses. This last characteristic is particularly important as it makes these devices attractive for the general public, thereby holding the potential to provide a research and product platform of unprecedented scale, quality, and flexibility.

While hearing aids and mobile headsets became widely accepted as head-worn devices, users in public spaces often consider novel head-attached sensors and devices as uncomfortable, irritating, or stigmatising. Yet with the advances in the following technologies, we believe eyewear computing will be a very prominent research field in the future:

- Increase in storage/battery capacity and computational power allows users to run eyewear computers continuously for more than a day (charging over night) gathering data to enable new types of life-logging applications.
- Miniaturization and integration of sensing, processing, and interaction functionality can enable a wide array of applications focusing on micro-interactions and intelligent assistance.
- Recent advances in real-life tracking of cognitive activities (e.g. reading, detection of fatigue, concentration) are additional enabling technologies for new application fields towards a quantified self for the mind. Smart eyewear and recognizing cognitive states go hand in hand, as naturally most research work in this field requires sensors.
- Cognitive scientists and psychologists have now a better understanding of user behavior and what induces behavior change. Therefore, smart eyewear could help users in achieving behaviour change towards their long term goals.

Eyewear computing has the potential to fundamentally transform the way machines perceive and understand the world around us and to assist humans in measurably and significantly improved ways. The seminar brought together researchers from a wide range of computing disciplines, such as mobile and ubiquitous computing, head-mounted eye tracking,

optics, computer vision, human vision and perception, privacy and security, usability, as well as systems research. Attendees discussed how smart eyewear can change existing research and how it may open up new research opportunities. For example, future research in this area could fundamentally change our understanding of how people interact with the world around them, how to augment these interactions, and may have a transformational impact on all spheres of life – the workplace, family life, education, and psychological well-being.

2 Table of Contents

Executive Summary

Andreas Bulling, Ozan Cakmakci, Kai Kunze, and James M. Rehg 161

Ignite Talks

Pervasive Sensing, Analysis, and Use of Visual Attention
Andreas Bulling 165

Egocentric Vision for social and cultural experiences
Rita Cucchiara 165

User Interfaces for Eyewear Computing
Steven K. Feiner 166

Action and Attention in First-person Vision
Kristen Grauman 167

Technology for Learning in Virtual and Augmented Reality
Scott W. Greenwald 168

Detecting Mental Processes and States from Visual Behaviour
Sabrina Hoppe 169

In pursuit of the intuitive interaction in our daily life
Masahiko Inami 169

Cognitive Activity Recognition in Real Life Scenario
Shoya Ishimaru 170

Reading-Life Log
Koichi Kise 171

Redesigning Vision
Kiyoshi Kiyokawa 172

Augmenting the Embodied Mind
Kai Kunze 173

Egocentric discovery of task-relevant interactions for guidance
Walterio Mayol-Cuevas 173

Haptic gaze interaction for wearable and hand-held mobile devices
Päivi Majaranta 174

From EyeWear Computing to Head Centered Sensing and Interaction
Paul Lukowicz 175

Eyewear Computing: Do we talk about security and privacy yet?
René Mayrhofer 175

A multimodal wearable system for logging personal behaviors and interests
Masashi Nakatani 176

Pupil – accessible open source tools for mobile eye tracking and egocentric vision research
Will Patera and Moritz Kassner 176

Smart Eyewear for Cognitive Interaction Technology
Thies Pfeiffer 177

Activity Recognition and Eyewear Computing
Philipp M. Scholl 178

Eyes, heads and hands: The natural statistics of infant visual experience <i>Linda B. Smith</i>	178
Head-Worn Displays (HWDs) for Everyday Use <i>Thad Starner</i>	179
Unsupervised Discovery of Everyday Activities from Human Visual Behaviour and 3D Gaze Estimation <i>Julian Steil</i>	179
Calibration-Free Gaze Estimation and Gaze-Assisted Computer Vision <i>Yusuke Sugano</i>	180
Wearable barcode scanning <i>Gábor Sörös</i>	180
Workshops and Tutorials	
Workshop: PUPIL <i>Moritz Kassner, Will Patera</i>	181
Workshop: Google Cardboard <i>Scott W. Greenwald</i>	184
Workshop: J!NS Meme <i>Shoya Ishimaru, Yuji Uema, Koichi Kise</i>	186
Tutorial: Head-Worn Displays <i>Kiyoshi Kiyokawa, Thad Starner, and Ozan Cakmakci</i>	187
Tutorial: Egocentric Vision <i>James M. Rehg, Kristen Grauman</i>	188
Challenges	189
Breakout Sessions	190
Group 1: Multimodal EyeWear Computing <i>Masashi Nakatani</i>	190
Group 2: Egocentric Vision <i>Rita Cucchiara, Kristen Grauman, James M. Rehg, Walterio W. Mayol-Cuevas</i> . .	193
Group 3: Security and Privacy <i>René Mayrhofer</i>	197
Group 4: Eyewear Computing for Skill Augmentation and Task Guidance <i>Thies Pfeiffer, Steven K. Feiner, Walterio W. Mayol-Cuevas</i>	199
Group 5: EyeWear Computing for Gaming <i>Thad Starner</i>	203
Group 6: Prototyping of AR Applications using VR Technology <i>Scott W. Greenwald</i>	204
Community Support	204
Datasets	204
Tools	205
Participants	206

3 Ignite Talks

3.1 Pervasive Sensing, Analysis, and Use of Visual Attention

Andreas Bulling (*Max-Planck-Institut für Informatik – Saarbrücken, DE*)

License © Creative Commons BY 3.0 Unported license
© Andreas Bulling

Main reference A. Bulling, “Pervasive Attentive User Interfaces”, *IEEE Computer*, 49(1):94–98, 2016.
URL <http://dx.doi.org/10.1109/MC.2016.32>

In this talk I motivated the need for new computational methods to sense, analyse and – most importantly – manage user attention continuously in everyday settings. This is important to enable future human-machine systems to cope with the ever-increasing number of displays and interruptions they cause. I provided a short overview of selected recent works in our group towards this vision, specifically head-mounted and appearance-based remote gaze estimation [2], short and long-term visual behaviour modelling and activity recognition, cognition-aware computing [3], attention modelling in graphical user interfaces [1], as well as collaborative human-machine vision systems for visual search target prediction and visual object detection.

References

- 1 Pingmei Xu, Yusuke Sugano, Andreas Bulling. *Spatio-Temporal Modeling and Prediction of Visual Attention in Graphical User Interfaces*. Proc. of the 34th ACM SIGCHI Conference on Human Factors in Computing Systems (CHI), 2016. <http://dx.doi.org/10.1145/2858036.2858479>
- 2 Xucong Zhang, Yusuke Sugano, Mario Fritz, Andreas Bulling. *Appearance-Based Gaze Estimation in the Wild*. Proc. of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4511-4520, 2015. <http://dx.doi.org/10.1109/CVPR.2015.7299081>
- 3 Andreas Bulling, Thorsten Zander. *Cognition-Aware Computing*. *IEEE Pervasive Computing* 13 (3), pp. 80-83, 2014. <http://dx.doi.org/10.1109/mprv.2014.42>

3.2 Egocentric Vision for social and cultural experiences

Rita Cucchiara (*University of Modena, IT*)

License © Creative Commons BY 3.0 Unported license
© Rita Cucchiara

Joint work of R.Cucchiara, A. Alletto, G. Serra

Main reference A. Alletto, G. Serra, S. Calderara, R. Cucchiara, “Understanding Social Relationships in Egocentric Vision”, *Pattern Recognition*, 48(12):4082–4096, 2015.
URL <http://dx.doi.org/10.1016/j.patcog.2015.06.006>

Egocentric Vision concerns computer vision models and techniques for understanding what a person sees, from the first person’s point of view, with eyewear devices and centered on the human perceptual needs. Computer vision problems become more challenging when applied in such an unconstrained scenario, with an unknown camera motion due to the body and head movements of the camera wearer. In this research, we address how egocentric vision can be exploited to augment cultural and social experiences. It can be done in many ways. A first application is in attention driven life-logging: whenever vision solutions will be able to recognize the interest of the persons, what they are looking at and how much they do, personalized video summarization will be available to keep the memory of the experience,

to share them and to use it also for educational purpose. The second is more general: it concerns recognition of targets of interests in indoor museums or in outdoor unconstrained setting, as for instance in a cultural visit around a historical/artistic area. Here egocentric vision is required to understand the social relationships among people and friends, recognize what persons would like to see and be engaged with, and localize the person's position to suggest useful information in real time. It can be done in a simplified manner by using image search for similarity or more precisely by providing 2D and 3D registration. Here, there are many open problems such as, for instance, video registration in real-time, matching images taken at different weather or time conditions. New generation of eyewear devices, possibly provided with eye-tracking and with high connectivity to allow a fast image processing, will provide a big leap- forward in this field, and will open new research areas in egocentric vision for interaction with environment.

References

- 1 P. Varini, G. Serra, R. Cucchiara. *Egocentric Video Summarization of Cultural Tour based on User Preferences*. Proc. of the 23rd ACM International Conference on Multimedia, 2015.
- 2 R. Cucchiara, A. Del Bimbo. *Visions for Augmented Cultural Heritage Experience*. IEEE Multimedia 2014.
- 3 A. Alletto, G. Serra, R. Cucchiara, V. Mighali, G. Del Fiore, L. Patrono, L. Mainetti, *An Indoor Location-aware System for an IoTbased Smart Museum*. Internet of Things Journal, 2016.

3.3 User Interfaces for Eyewear Computing

Steven K. Feiner (Columbia University, US)

License © Creative Commons BY 3.0 Unported license
© Steven K. Feiner

Joint work of Feiner, Steven K.; Elvezio, Carmine; Oda, Ohan; Sukan, Mengu; Tversky, Barbara
Main reference O. Oda, C. Elvezio, M. Sukan, S.K. Feiner, B. Tversky, "Virtual Replicas for Remote Assistance in Virtual and Augmented Reality", in Proc. of the 28th Annual ACM Symp. on User Interface Software & Technology (UIST'15), pp. 405–415, ACM, 2015.
URL <http://dx.doi.org/10.1145/2807442.2807497>

Our research investigates how we can create effective everyday user interfaces that employ eyewear alone and in synergistic combination with other modalities. We are especially interested in developing multimodal interaction techniques that span multiple users, displays, and input devices, adapting as we move in and out of their presence. Domains that we have addressed include outdoor wayfinding, entertainment, and job performance. One approach that we employ is *environment management* – supervising UI design across space, time, and devices. (A specific example is *view management*, in which we automate the spatial layout of information by imposing and maintaining visual constraints on the locations of objects in the environment and their projections in our view.) A second approach is the creation of *hybrid user interfaces*, in which heterogeneous interface technologies are combined to complement each other. For example, we have used a video-see-through head-worn display to overlay 3D building models on their footprints presented on a multitouch horizontal tabletop display.

A research theme underlying much of our research is collaboration: developing user interfaces that support multiple users wearing eyewear working together, both co-located and remote. We are especially interested in remote task assistance, in which a remote subject-matter expert helps a less knowledgeable local user perform a skilled task. Two issues here are helping the local user quickly find an appropriate location at which to perform the

task, and getting them to perform the task correctly. In one project, we have developed and evaluated ParaFrustum, a 3D interaction and visualization technique that presents a range of appropriate positions and orientations from which to perform the task [3]. In a second project, we have developed and evaluated 3D interaction techniques that allow a remote expert to create and manipulate virtual replicas of physical objects in the local environment to refer to parts of those physical objects and to indicate actions on them [2].

Finally, much of our research uses stereoscopic eyewear with a relatively wide field of view to present augmented reality in which virtual media are geometrically registered and integrated with our perception of the physical world. We are also exploring the use of eyewear with a small monoscopic field of view, such as Google Glass. In this work, we are developing and evaluating approaches that do not rely on geometric registration [1].

References

- 1 Carmine Elvezio, Mengü Sukan, Steven Feiner, and Barbara Tversky. *Interactive Visualizations for Monoscopic Eyewear to Assist in Manually Orienting Objects in 3D*. Proc. IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2015), 180–181. <http://dx.doi.org/10.1109/ISMAR.2015.54>
- 2 Ohan Oda, Carmine Elvezio, Mengü Sukan, Steven Feiner, and Barbara Tversky. *Virtual replicas for remote assistance in virtual and augmented reality*. Proc. ACM Symposium on User Interface Software and Technology (UIST 2015), pp. 405–415. <http://dx.doi.org/10.1145/2807442.2807497>
- 3 Mengü Sukan, Carmine Elvezio, Ohan Oda, Steven Feiner, and Barbara Tversky. *Para-Frustum: Visualization techniques for guiding a user to a constrained set of viewing positions and orientations*. Proc. ACM Symposium on User Interface Software and Technology (UIST 2014), pp. 331–340. <http://dx.doi.org/10.1145/2642918.2647417>

3.4 Action and Attention in First-person Vision

Kristen Grauman (University of Texas at Austin, USA)

License © Creative Commons BY 3.0 Unported license
© Kristen Grauman

Joint work of K. Grauman, D. Jayaraman, Yong Jae Lee, Bo Xiong, Lu Zheng
URL <http://www.cs.utexas.edu/~grauman/>

A traditional third-person camera passively watches the world, typically from a stationary position. In contrast, a first-person (wearable) camera is inherently linked to the ongoing experiences of its wearer. It encounters the visual world in the context of the wearer’s physical activity, behavior, and goals. This distinction has many intriguing implications for computer vision research, in topics ranging from fundamental visual recognition problems to high-level multimedia applications.

I will present our recent work in this space, driven by the notion that the camera wearer is an active participant in the visual observations received. First, I will show how to exploit egomotion when learning image representations [1]. Cognitive science tells us that proper development of visual perception requires internalizing the link between “how I move” and “what I see” – yet today’s best recognition methods are deprived of this link, learning solely from bags of images downloaded from the Web. We introduce a deep feature learning approach that embeds information not only from the video stream the observer sees, but also the motor actions he simultaneously makes. We demonstrate the impact for recognition, including a scenario where features learned from ego-video on an autonomous car

substantially improve large-scale scene recognition. Next, I will present our work exploring video summarization from the first person perspective [3, 2]. Leveraging cues about ego-attention and interactions to infer a storyline, we automatically detect the highlights in long videos. We show how hours of wearable camera data can be distilled to a succinct visual storyboard that is understandable in just moments, and examine the possibility of person- and scene-independent cues for heightened attention. Overall, whether considering action or attention, the first-person setting offers exciting new opportunities for large-scale visual learning.

References

- 1 D. Jayaraman and K. Grauman. *Learning Image Representations Tied to Ego-Motion*. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, Dec 2015.
- 2 Y. J. Lee and K. Grauman. *Predicting Important Objects for Egocentric Video Summarization*. International Journal on Computer Vision, Volume 114, Issue 1, pp. 38–55, August 2015.
- 3 B. Xiong and K. Grauman. *Detecting Snap Points in Egocentric Video with a Web Photo Prior*. In Proceedings of the European Conference on Computer Vision (ECCV), Zurich, Switzerland, Sept 2014.

3.5 Technology for Learning in Virtual and Augmented Reality

Scott W. Greenwald (MIT – Cambridge, US)

License © Creative Commons BY 3.0 Unported license
© Scott W. Greenwald

Main reference S. W. Greenwald, M. Khan, C. D. Vazquez, P. Maes, “TagAlong: Informal Learning from a Remote Companion with Mobile Perspective Sharing”, in Proc. of the 12th IADIS Int’l Conf. on Cognition and Exploratory Learning in Digital Age (CELDA’15), 2015.

URL <http://dspace.mit.edu/handle/1721.1/100242>

Face-to-face communication is highly effective for teaching and learning. When the student is remote (e.g. in a situated context) or immersed in virtual reality, effective teaching can be much harder. I propose that in these settings, it can be done as well or better than face-to-face using a system that provides the teacher with the first-person perspective, along with real-time sensor data on the cognitive and attentional state of the learner. This may include signals such as EEG, eye gaze, and facial expressions. The paradigm of eyewear computing – to provide contextual feedback using the egocentric perspective as an input and output medium – is critical in the realization of such a system. In my current work, I investigate key communication affordances for effective teaching and learning in such settings.

3.6 Detecting Mental Processes and States from Visual Behaviour

Sabrina Hoppe (*Max-Planck-Institut für Informatik – Saarbrücken, DE*)

License © Creative Commons BY 3.0 Unported license
© Sabrina Hoppe

Joint work of Stephanie Morey, Tobias Loetscher, Andreas Bulling

Main reference S. Hoppe, T. Loetscher, S. Morey, A. Bulling, “Recognition of Curiosity Using Eye Movement Analysis”, in *Adjunct Proc. of the ACM Int’l Joint Conf. on Pervasive and Ubiquitous Computing (UbiComp’15)*, pp. 185–188, ACM, 2015.

URL <http://dx.doi.org/10.1145/2800835.2800910>

It is well known that visual behaviour in everyday life changes with activities and certain environmental factors like lightening conditions, but how about mental processes? In this talk, I introduced some of our initial work in this direction: we have shown that curiosity, as an exemplar personality trait, is reflected in visual behaviour and can be inferred from gaze data [1]. However, personality is just one out of many potential mental states to be investigated in this context – further processes of interest include confusion [3] and mental health [2]. If and how this information can be leveraged from gaze data remains an interesting open question in the field of eye wear computing.

References

- 1 Sabrina Hoppe, Tobias Loetscher, Stephanie Morey, Andreas Bulling. *Recognition of Curiosity Using Eye Movement Analysis*, *Adj. Proc. of the ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*, pp. 185–188, 2015.
- 2 Tobias Loetscher, Celia Chen, Sophie Wignall, Andreas Bulling, Sabrina Hoppe, Owen Churches, Nicole Thomas. *A study on the natural history of scanning behaviour in patients with visual field defects after stroke*, *BMC Neurology*, 15(64), 2015.
- 3 Nikolina Koleva, Sabrina Hoppe, Mohammed Mehdi Moniri, Maria Staudte, Andreas Bulling. *On the interplay between spontaneous spoken instructions and human visual behaviour in an indoor guidance task*, *Proc. of the 37th Annual Meeting of the Cognitive Science Society*, 2015.

3.7 In pursuit of the intuitive interaction in our daily life

Masahiko Inami (*University of Tokyo, JP*)

License © Creative Commons BY 3.0 Unported license
© Masahiko Inami

Joint work of Fan, Kevin; Huber, Jochen; Nanayakkara, Suranga; Kise, Koichi; Kunze, Kai; Ishimaru, Shoya; Tanaka, Katsuma; Uema, Yuji

Main reference K. Fan, J. Huber, S. Nanayakkara, M. Inami, “SpiderVision: Extending the Human Field of View for Augmented Awareness”, in *Proc. of the 5th Augmented Human Int’l Conf. (AH’14)*, Article No. 49, ACM, 2014.

URL <http://dx.doi.org/10.1145/2582051.2582100>

We have established the Living Lab Tokyo at National Museum of Emerging Science and Innovation (Miraikan). The main focus is to create a living environment for users, by embedding mini sensors to sense the various interactions between users and the environment for various purposes such as entertainment, safety, enhancing communication between family members and much more. Our group worked hand-in-hand with users to learn and explore more about the users’ needs at the Living Lab Tokyo. We have been achieved some interactive systems such as Senskin [1] and JINS MEME [2]. Recently, we are trying to expand our research target from a living room to a play ground. Then we have start a new research laboratory on Superhuman sports, which is a new challenge to reinvent sports that anyone can

enjoy, anywhere and anytime. Based on this concept we have developed a new head mounted device, SpiderVision [3] that extends the human field of view to augment a user's awareness of things happening behind one's back. SpiderVision leverages a front and back camera to enable users to focus on the front view while employing intelligent interface techniques to cue the user about activity in the back view. The extended back view is only blended in when the scene captured by the back camera is analyzed to be dynamically changing. In this project, we explore factors that affect the blended extension, such as view abstraction and blending area. We succeeded using the system on a sky field. I would like to discuss possible application to enhance our physical activities with a smart eyewear.

References

- 1 Masa Ogata, Yuta Sugiura, Yasutoshi Makino, Masahiko Inami, and Michita Imai. 2013. *SenSkin: adapting skin as a soft interface*. In Proceedings of the 26th annual ACM symposium on User interface software and technology (UIST'13). ACM, New York, NY, USA, 539-544. <http://dx.doi.org/10.1145/2501988.2502039>
- 2 Shoya Ishimaru, Kai Kunze, Katsuma Tanaka, Yuji Uema, Koichi Kise and Masahiko Inami. *Smarter Eyewear – Using Commercial EOG Glasses for Activity Recognition*. Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct Publication (UbiComp2014). September 2014.
- 3 Kevin Fan, Jochen Huber, Suranga Nanayakkara and Masahiko Inami. *SpiderVision: Extending the Human Field of View for Augmented Awareness*. In Proceedings of the 5th Augmented Human International Conference (AH'14), ACM, Article 49, 2014.

3.8 Cognitive Activity Recognition in Real Life Scenario

Shoya Ishimaru (Osaka Prefecture University, JP)

License © Creative Commons BY 3.0 Unported license

© Shoya Ishimaru

Main reference S. Ishimaru, K. Kunze, K. Tanaka, Y. Uema, K. Kise, M. Inami, “Smart Eyewear for Interaction and Activity Recognition”, in Proc. of the 33rd Annual ACM Conf. Extended Abstracts on Human Factors in Computing Systems (CHI EA'15), pp. 307–310, ACM, 2015.

URL <http://dx.doi.org/10.1145/2702613.2725449>

As people can be motivated to keep physical fitness by looking back their step counts, tracking cognitive activities (e.g. the number of words they read in a day) can help them to improve their cognitive lifestyles. While most of the physical activities can be recognized with body-mounted motion sensors, recognizing cognitive activities is still a challenging task because body movements during the activities are limited and we need additional sensors. The sensors also should be for everyday use to track daily life. In this talk, I introduced three projects tracking our cognitive activities with affordable technologies. The first project is eye tracking on mobile tablets. Most of the mobile tablets like iPad have a front camera for video chat. We have analyzed the facial image from the camera and detected where the user is looking at [1]. The second project is activity recognition based on eye blinks and head motions. We have detected eye blinks by using the sensor built in Google Glass and combined head motion and eye blink for the classification [2]. The last project is signal analysis on commercial EOG glasses. We have detected eye blinks and horizontal eye movement by using three electrodes on J!NS MEME and used them for cognitive activity recognition [3]. In addition to detecting reading activity, the number of words a user read can be estimated from small eye movements appeared on the EOG signal. After presenting the three projects, I proposed an important topic we should tackle at Dagstuhl. Because of the rising of deep

learning, the critical issue for classification tasks might be switched from “What is the best feature” to “How can we create a large dataset with labeling”. Unlike image analysis, it’s difficult to correct human’s sensor data with ground truth. So I would like to discuss with participants how to organize large scale recording in real life through the seminar.

References

- 1 Kai Kunze, Shoya Ishimaru, Yuzuko Utsumi and Koichi Kise. *My reading life: towards utilizing eyetracking on unmodified tablets and phones*. Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication (UbiComp2013). September 2013.
- 2 Shoya Ishimaru, Jens Weppner, Kai Kunze, Koichi Kise, Andreas Dengel, Paul Lukowicz and Andreas Bulling. *In the Blink of an Eye – Combining Head Motion and Eye Blink Frequency for Activity Recognition with Google Glass*. Proceedings of the 5th Augmented Human International Conference (AH2014). March 2014.
- 3 Shoya Ishimaru, Kai Kunze, Katsuma Tanaka, Yuji Uema, Koichi Kise and Masahiko Inami. *Smart Eyewear for Interaction and Activity Recognition*. Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI2015). April 2015.

3.9 Reading-Life Log

Koichi Kise (Osaka Prefecture University, JP)

License © Creative Commons BY 3.0 Unported license
© Koichi Kise

Main reference K. Yoshimura, K. Kise, K. Kunze, “The Eye as the Window of the Language Ability: Estimation of English Skills by Analyzing Eye Movement While Reading Documents”, in Proc. of 13th Int’l Conf. on Document Analysis and Recognition (ICDAR’15), pp. 251–255, IEEE, 2015.

URL <http://dx.doi.org/10.1109/ICDAR.2015.7333762>

Reading to the mind is what exercise is to the body. It is a well-known sentence that indicates the importance of reading. Actually we are spending so much amount of time for everyday reading. In other words, it is rare to spend a whole day without reading anything. Unfortunately, however, this activity has not been recorded and thus we are not able to use it. Our project called “Reading-Life Log” is to record our reading activities at various levels by using a variety of sensors. As levels, we have proposed the amount of reading, the period of reading, the type of documents, the log of read words, and the level of understanding and interests. The first one, the amount of reading is measured by a method called “wordometer” [1] which estimates the number of read words based on eye movement or gaze. The period of reading is estimated by analyzing eye movements or gaze. Document types are recognized by using first person vision, or eye gaze. Read words are listed with the help of document image retrieval and eye gaze. As the estimation of the level of understanding, we have developed a method to estimate an English proficiency by analyzing eye gaze [2].

As a possible research direction, I also introduced our notion of “experiential supplement”, which is to record people’s experiences to give them to persons for their better experiences. This is based on our notion that ordinary people can be best helped by other ordinary people who have similar background knowledge and experiences.

As an important research topic for us I have pointed out that we need a way of fractional distillation of the data obtained by eye wears, because the data contains many factors such as a person, an object the person looks, his/her interests, difficulties, etc.

References

- 1 Kai Kunze, Hitoshi Kawaichi, Koichi Kise and Kazuyo Yoshimura, *The Wordometer – Estimating the Number of Words Read Using Document Image Retrieval and Mobile Eye Tracking*, Proc. 12th International Conference on Document Analysis and Recognition (ICDAR 2013), pp. 25–29 (2013-8).
- 2 Kazuyo Yoshimura, Koichi Kise and Kai Kunze, *The Eye as the Window of the Language Ability: Estimation of English Skills by Analyzing Eye Movement While Reading Documents*, Proc. 13th International Conference on Document Analysis and Recognition (ICDAR 2015), pp. 251–255 (2015-8).

3.10 Redesigning Vision

Kiyoshi Kiyokawa (Osaka University, JP)

License © Creative Commons BY 3.0 Unported license

© Kiyoshi Kiyokawa

Joint work of Kiyoshi Kiyokawa, Alexandor Plopski, Jason Orlosky, Yuta Itoh, Christian Nitschke, Takumi Toyama, Daniel Sonntag, Ernst Kruijff, Kenny Moser, J. Edward Swan II, Dieter Schmalstieg, Gudrun Klinker, and Haruo Takemura

Main reference A. Plopski, Y. Itoh, C. Nitschke, K. Kiyokawa, G. Klinker, and H. Takemura, “Corneal-Imaging Calibration for Optical See-Through Head-Mounted Displays”, in IEEE Transaction on Visualization and Computer Graphics (TVCG), 21(4):481–490, 2015.

URL <http://doi.ieeecomputersociety.org/10.1109/TVCG.2015.2391857>

Ideally, head worn displays (HWDs) are expected to produce any visual experience we imagine, however, compared to this ultimate goal, current HWDs are still far from perfect. One of fundamental problems of optical see-through HWDs is that the user’s exact view is not accessible as image overlay happens on the user’s retina. Our new concept, corneal feedback augmented reality (AR) is a promising approach to realize a closed feedback loop for better registration [1], color correction, contrast adjustment, accurate eye tracking and scene understanding, by continuously analyzing the reflections in the eye.

In the case of video see-through HWDs, our vision can be more flexibly redesigned. With a modular video see-through HWD we developed [2], our natural vision can be switched to a super eyesight on demand such as a super wide view and a super zoom view, which is controlled by natural eye gesture such as squinting. By combining advanced 3D reconstruction system, our vision can even be free from physical constraints so that we can change our viewpoint or the size of the world on demand.

References

- 1 Alexandor Plopski, Yuta Itoh, Christian Nitschke, Kiyoshi Kiyokawa, Gudrun Klinker, and Haruo Takemura, *Corneal-Imaging Calibration for Optical See-Through Head-Mounted Displays*, IEEE Transaction on Visualization and Computer Graphics (TVCG), Special Issue on IEEE Virtual Reality (VR) 2015, Vol. 21, No. 4, pp. 481–490, 2015.
- 2 Jason Orlosky, Takumi Toyama, Kiyoshi Kiyokawa, and Daniel Sonntag, *ModuLAR: Eye-controlled Vision Augmentations for Head Mounted Displays*, IEEE Transactions on Visualization and Computer Graphics (TVCG), Special Issue on International Symposium on Mixed and Augmented Reality (ISMAR) 2015, Vol. 21, No. 11, pp. 1259–1268, 2015.

3.11 Augmenting the Embodied Mind

Kai Kunze (*Keio University – Yokohama, JP*)

License © Creative Commons BY 3.0 Unported license
© Kai Kunze

People use mobile computing technology to track their health and fitness progress, from simple step counting to monitoring food intake to measuring how long and well they sleep. Can also quantify cognitive tasks in real-world requirements and in a second step can we use them to change behavior? There are patterns in the physiological signals and behavior of humans (facial expressions, nose temperature, eye movements, blinks etc.) that can reveal information about mental conditions and cognitive functions. We explore how to use these patterns to recognize mental states and in a second step searches for interactions to change human mental states by stimulating the users to change these patterns. We believe all sensing and actuation will be embedded in smart eye wear. We continue our research recognizing reading comprehension detecting cognitive load using eye movement analysis, first with slightly altered stationary setups (baseline experiments) then with extending to a more mobile, unconstrained setup, where we also use our pervasive reading detection/quantification methods (recording how much a person reads with unobtrusive smart glasses and looking for patterns in to detect what are healthy reading habits to improve comprehension) and our work to record seamless the user’s facial expression using an unobtrusive glasses design (affective wear). In initial trials with affective wear, we found indications that facial expressions change in relation to cognitive functions.

References

- 1 Amft, O., Wahl, F., Ishimaru, S., and Kunze, K. *Making regular eyeglasses smart*. IEEE Pervasive Computing 14(3):32–43, 2015.

3.12 Egocentric discovery of task-relevant interactions for guidance

Walterio Mayol-Cuevas (*University of Bristol, UK*)

License © Creative Commons BY 3.0 Unported license
© Walterio Mayol-Cuevas

Joint work of Walterio Mayol-Cuevas; Dima Damen; Teesid Leelasawassuk

Main reference D. Damen, T. Leelasawassuk, O. Haines, A. Calway, W. Mayol-Cuevas, “You-Do, I-Learn: Discovering Task Relevant Objects and their Modes of Interaction from Multi-User Egocentric Video”, in Proc. of the British Machine Vision Conf. (BMVC’14), BMVA Press, 2014.

URL <http://dx.doi.org/10.5244/C.28.30>

How to decide what is relevant from the world? What objects and situations are important to record and which ones to ignore? These are key questions for efficient egocentric perception. One first step toward egocentric relevance determination is the building of real-time models of attention which can help in building systems that are better at online data logging, systems that are assistive by knowing what to show, as well as understanding more about attention is part of the process needed to inform the type of sensors and feedback hardware needed. At Bristol, we have been developing methods that gate visual input into small snippets that contain information that is task relevant. These include objects that have been interacted with, the ways in which these objects have been used as well as video segments representative of the interactions that can later be offered to people for guidance. The ultimate goal of this work is to allow people to feel they can do much more than what they can currently do – to


have such a superhuman feeling will transform how wearables are perceived and make them more useful beyond simple recording devices. Our work is underpinned by various research strands we are exploring including attention estimation from IMU signals [1], the ability to fuse information from SLAM and gaze patterns with visual appearance for interaction relevance determination [2] and lightweight object recognition for fast and onboard operation [3]. Overall, the understanding of what is important and useful to provide guidance will ultimately lead to better informed algorithmic and hardware choices in eyewear computing.

References

- 1 T Leelasawassuk, D Damen, W Mayol-Cuevas. *Estimating Visual Attention from a Head Mounted IMU*. International Symposium on Wearable Computers (ISWC). 2015.
- 2 D Damen, T Leelasawassuk, O Haines, A Calway, W Mayol-Cuevas. *You-Do, I-Learn: Discovering Task Relevant Objects and their Modes of Interaction from Multi-User Egocentric Video*. British Machine Vision Conference (BMVC), Nottingham, UK. 2014.
- 3 Dima Damen, Pished Bunnun, Andrew Calway, Walterio Mayol-Cuevas, *Real-time Learning and Detection of 3D Texture-less Objects: A Scalable Approach*. British Machine Vision Conference. September 2012.

3.13 Haptic gaze interaction for wearable and hand-held mobile devices

Päivi Majaranta (University of Tampere, FI)

License  Creative Commons BY 3.0 Unported license

© Päivi Majaranta

Joint work of Majaranta, Päivi; Kangas, Jari; Isokoski, Poika; Špakov, Oleg; Rantala, Jussi; Akkil, Deepak; Raisamo, Roope


Gaze-based interfaces provide means for communication and control. It is known from previous research that gaze interaction is highly beneficial for people with disabilities (see e.g. work done within the COGAIN network reported in [3]). Gaze interaction could also potentially revolutionize pervasive and mobile interaction. Haptics provide a private channel for feedback, as it is only felt by the person wearing or holding the device, e.g. on eye glass frames [1] or hand-held devices. We found such haptic feedback useful for example in controlling a mobile device with gaze gestures. Haptic feedback can significantly reduce error rates and improve performance and user satisfaction [2]. In addition to gaze interaction with mobile devices, we see a lot of potential in using gaze as a channel to interact with our surroundings in the era of Internet-of-Things. One challenging question that we wish to tackle is how to enable more direct interaction with the objects instead of sending the commands via a screen.

References

- 1 Kangas, J., Akkil, D., Rantala, J., Isokoski, P., Majaranta, P., and Raisamo, R. (2014a). *Using Gaze Gestures with Haptic Feedback on Glasses*. Proc. 8th Nordic Conference on Human-Computer Interaction. ACM, 1047-1050. <http://dx.doi.org/10.1145/2639189.2670272>
- 2 Kangas, J., Akkil, D., Rantala, J., Isokoski, P., Majaranta, P. and Raisamo, R. (2014b). *Gaze Gestures and Haptic Feedback in Mobile Devices*. Proc. SIGCHI Conference on Human Factors in Computing Systems. ACM, 435-438. <http://dx.doi.org/10.1145/2556288.2557040>
- 3 Majaranta, P., Aoki, H., Donegan, M., Hansen, D.W., Hansen, J.P., Hyrskykari, A., and Riih , K-J. (Eds.) *Gaze Interaction and Applications of Eye Tracking: Advances in Assistive Technologies*, IGI Global, 2012. <http://dx.doi.org/10.4018/978-1-61350-098-9>

3.14 From EyeWear Computing to Head Centered Sensing and Interaction

Paul Lukowicz (DFKI Kaiserslautern, DE)

License  Creative Commons BY 3.0 Unported license
 © Paul Lukowicz
 URL <http://ei.dfki.de/en/home/>

Eyewear Computing in the for currently implemented by e.g. Google Glass predominately aims at making switching between the digital and the physical world faster and easier. This is the proposition on which consumers are likely to be initially adopting this technology. However, in the long run, the more profound impact is the fact that it creates the possibility of putting advanced sensing and interaction modalities at the users' head. This in turn creates access to a broad range of information sources not accessible at any other body location which is likely to revolutionize activity and context recognition.

3.15 Eyewear Computing: Do we talk about security and privacy yet?

René Mayrhofer (Universität Linz, AT)

License  Creative Commons BY 3.0 Unported license
 © René Mayrhofer

Like other wearable devices, eyewear is subject to typical threats to security and privacy. In my talk, I outlined four categories of threats that seem especially relevant to eyewear computing. Three of these are shared with other mobile device categories such as smart phones: *device-to-user authentication* is required to make sure that a device has not been physically replaced or tampered with (and which has not been sufficiently addressed for most device shapes [1]); *emphuser-to-device authentication* to prevent malicious use of devices by other users (which is mostly solved for smart phones citehintze-locked-device-usage, but still largely open for eyewear devices); and *emphdevice-to-device authentication* to ensure that wireless links are established correctly (strongly depending on the scenario, some solutions exist that may be directly applicable to eyewear computing [3]). The fourth threat of *privacy* is also common to all wearable devices, but due to typical inclusion of cameras and microphones, is potentially harder to address for eyewear devices. We expect the application n of cross-device authentication methods to have significant impact especially for security in eyewear computing.

References

- 1 Rainhard D. Findling, Rene Mayrhofer. *Towards Device-to-User Authentication: Protecting Against Phishing Hardware by Ensuring Mobile Device Authenticity using Vibration Patterns*. 14th International Conference on Mobile and Ubiquitous Multimedia (MUM'15), 2015
- 2 Daniel Hintze, Rainhard D. Findling, Muhammad Muaaz, Sebastian Scholz, Rene Mayrhofer. *Diversity in Locked and Unlocked Mobile Device Usage*. Adjunct Proceedings of ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2014). 379–384, 2014
- 3 Rene Mayrhofer. *Ubiquitous Computing Security: Authenticating Spontaneous Interactions*. Vienna University, 2008

3.16 A multimodal wearable system for logging personal behaviors and interests

Masashi Nakatani (*University of Tokyo, JP*)

License © Creative Commons BY 3.0 Unported license
© Masashi Nakatani

Joint work of Liang, Feng; Miyazaki, Hazuki; Minamizawa, Kouta; Nakatani, Masashi; Tachi, Susumu

URL <http://www.merkel.jp/research>

We propose a wearable/mobile system that can capture our daily life experiences with multimodality (vision, sound, and haptics). This device is consisted of wearable devices that captures ego-centric view through a camera, microphones for audio information and nine-axis inertial motion sensor as haptics information. Collected data is analyzed based on individual interests to the environment, then can be used for predicting users' interest based on statistics data. By combining with the state-of-art smart eyewear technology, we aim at segmenting captured audio-visual scene based on measured dataset (mainly haptics information) as well as personal interest as ground truth. This system would be helpful for providing custom-made service based on personal interests, such as guided tour of the city, hike planning, and trail running etc. This system may also provide benefit for elderly people who may need life support in their daily lives.

3.17 Pupil – accessible open source tools for mobile eye tracking and egocentric vision research

Will Patera (*Pupil Labs – Berlin, DE*) and Moritz Kassner (*Pupil Labs – Berlin, DE*)

License © Creative Commons BY 3.0 Unported license
© Will Patera and Moritz Kassner

Main reference M. Kassner, W. Patera, A. Bulling, “Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction”, in Proc. of the ACM Int'l Joint Conf. on Pervasive and Ubiquitous Computing (UbiComp'14), pp. 1151–1160, ACM, 2014.

URL <http://dx.doi.org/10.1145/2638728.2641695>

URL <https://pupil-labs.com/>

Pupil is an accessible, affordable, and extensible open source platform for eye tracking and egocentric vision research. It comprises a lightweight modular eye tracking and egocentric vision headset as well as an open source software framework for mobile eye tracking. Accuracy of 0.6 degrees and precision 0.08 degrees can be obtained. Slippage compensation is implemented using a temporal 3D model of the eye. Pupil is used by a diverse group of researchers around the world. The primary mission of Pupil is to create an open community around eye tracking methods and tools.

References

- 1 Moritz Kassner, William Patera, Andreas Bulling, *Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction*. Proc. of the ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp), pp. 1151–1160, 2014.

3.18 Smart Eyewear for Cognitive Interaction Technology

Thies Pfeiffer (Universität Bielefeld, DE)

License © Creative Commons BY 3.0 Unported license
© Thies Pfeiffer

Main reference T. Pfeiffer, P. Renner, N. Pfeiffer-Leßmann, “EyeSee3D 2.0: Model-based Real-time Analysis of Mobile Eye-Tracking in Static and Dynamic Three-Dimensional Scenes”, in Proc. of the 9th Biennial ACM Symp. on Eye Tracking Research & Applications, pp. 189–196, ACM, 2016.

URL <http://dx.doi.org/10.1145/2857491.2857532>

We are currently following a line of research in which we focus on assistance systems for humans that are adaptive to the competences and the current cognitive state of the user. In the project ADAMAAS (Adaptive and Mobile Action Assistance in Daily Living Activities), for example, we are addressing the groups of elderly and people with mental handicaps. One main research question is, what kind of assistance to provide in which situation (including competence and cognitive state) to support the person in maintaining an autonomous life. Other scenarios we are targeting with different projects are sports training, assistance in decision tasks and chess playing.

We believe that smart eyewear is a key technology here, as in the targeted activities people have to make use of their hands. Coupled with sensor technologies, such as sensors for egocentric vision or eye tracking, we aim to assess the current cognitive state. Paired with display technologies for augmented reality (visual or acoustic) smart eyewear can provide contextual feedback that is adaptive to the current progress and level of expertise of the wearer.

One method we are applying to follow this goal is virtual reality simulation for a rapid prototyping of different design alternatives for the hardware and the user interface. For this we make use of either a fully immersive CAVE or HMDs, such as Oculus Rift or Samsung Gear VR. In one of the breakout sessions here at Dagstuhl, we were able to discuss merits and challenges of this approach in more details and from different perspectives.

References

- 1 Thies Pfeiffer, Patrick Renner, Nadine Pfeiffer-Leßmann (2016, to appear). *EyeSee3D 2.0: Model-based Real-time Analysis of Mobile Eye-Tracking in Static and Dynamic Three-Dimensional Scenes*. In ETRA’16: 2016 Symposium on Eye Tracking Research and Applications Proceedings. <http://dx.doi.org/10.1145/2857491.2857532>
- 2 Patrick Renner, Thies Pfeiffer (2015). *Online Visual Attention Monitoring for Mobile Assistive Systems*. In SAGA 2015: 2nd International Workshop on Solutions for Automatic Gaze Data Analysis (pp. 14-15). eCollections Bielefeld. <http://biecoll.ub.uni-bielefeld.de/volltexte/2015/5382/>
- 3 Jella Pfeiffer, Thies Pfeiffer, Martin Meißner (2015). *Towards attentive in-store recommender Systems*. In Annals of Information Systems: Vol. 18. Reshaping Society through Analytics, Collaboration, and Decision Support (pp. 161-173). Springer International Publishing.

3.19 Activity Recognition and Eyewear Computing

Philipp M. Scholl (Universität Freiburg, DE)

License © Creative Commons BY 3.0 Unported license
© Philipp M. Scholl

Main reference P. Scholl, K. Van Laerhoven, “Wearable digitization of life science experiments”, in Proc. of the 2014 ACM Int’l Joint Conf. on Pervasive and Ubiquitous Computing: Adjunct Publication, pp. 1381–1388, ACM, 2014.

URL <http://dx.doi.org/10.1145/2638728.2641719>

Combining motion-based Activity Recognition and Eyewear Computing could allow for new ways of documenting manual work. In a wetlab environment, protective garment is necessary to avoid contamination of the experiment subjects, as well as protecting the experimenter from harmful agents. However, the taken procedure needs to minutiously documented. In the sense of Vannevar Bush’s vision[1] of a scientist wearing a head-mounted camera records his experiment. These recordings still need to be indexed to be useful. The idea is to detect well-defined activities from wrist motion, and similar sensors, which can serve as an additional index to these recordings, serving as external memories for scientists. These recordings can either be reviewed post-experiment or accessed implicitly while the experiment is on-going.

References

- 1 Bush, Vannevar. *As we may think*. The atlantic monthly, pp. 101–108, 1945.

3.20 Eyes, heads and hands: The natural statistics of infant visual experience

Linda B. Smith (Indiana University – Bloomington, US)

License © Creative Commons BY 3.0 Unported license
© Linda B. Smith

Main reference S. Jayaraman, C. Fausey, L. B. Smith, “The Faces in Infant-Perspective Scenes Change over the First Year of Life”, PLoS ONE, 10(5):e0123780, 2015.


URL <http://dx.doi.org/10.1371/journal.pone.0123780>

URL <http://www.iub.edu/~cogdev>

The visual objects labeled by common nouns – truck, dog, cup – are so readily recognized and their names so readily generalized by young children that theorists have suggested that these “basic-level” categories “carve nature at its joints.” This idea is at odds with contemporary understanding of visual object recognition, both in human visual science and in computational vision. In these literatures, object recognition is seen as a hard and unsolved. If object categories are “givens” for young perceivers, theorists of human and machine vision do not yet know how they are given. Understanding the properties of infant and toddler egocentric vision – and the coupling of those properties to the changing motor abilities provides potentially transformative new insights into typical and atypical human developmental process, and, perhaps, to machine vision. The core idea is that that the visual regularities that t young perceivers encounter is constrained by the limits of time and place and by the young child’s behavior, including the behavior of eyes, heads and hands. Changes in infant motor abilities gate and order regularities and in a sense, guide the infant through a series of sequential tasks and through a search space for an optimal solution to the problem of recognizing objects under varied and nonoptimal conditions Although the natural statistics of infant experience may not quite “carve nature at its joints,” but we propose they make those joints easier to find. In pursuit of this idea, we have collected and are analyzing a large corpus of infant-perspective scenes, about 500 total hours of video, 54 million images.

3.21 Head-Worn Displays (HWDs) for Everyday Use

Thad Starner (*Georgia Institute of Technology – Atlanta, US*)

License  Creative Commons BY 3.0 Unported license
© Thad Starner

Consumer purchase and use of eyewear is driven more by fashion than just about any other feature. Creating head-worn displays (HWDs) for everyday use must put fashion first. Due to lack of familiarity with the use of the devices, HWD consumers often desire features that are impractical or unnecessary. Transparent, see-through displays require more power, are more expensive, are difficult to see in bright environments, and result in inferior performance on many practical virtual and physical world tasks. Yet consumers will less readily try opaque, see-around displays because they are not aware of the visual illusion that makes these displays appear see-through. Many desire wide field-of-view (FOV) displays. However, the human visual system only sees in high resolution in a small 2 degree area called the fovea. Small FOV consumer displays (i.e., smart phones) have been highly successful for reading books, email, messaging, mobile purchasing, etc. A typical smart phone has a 8.8x15.6 degree FOV. Small FOV displays are much easier to manufacture with fashionable eyewear and will likely be the first successful (> 2m) consumer HWD product. Head weight should be kept under 75g for everyday use. A traveling exhibit of previous HWDs can be found at <http://wcc.gatech.edu/exhibition>

References

- 1 Thad Starner. *The Enigmatic Display*. IEEE Pervasive Computing 2(1):133-135, 2003.

3.22 Unsupervised Discovery of Everyday Activities from Human Visual Behaviour and 3D Gaze Estimation

Julian Steil (*Max-Planck-Institut für Informatik – Saarbrücken, DE*)

License  Creative Commons BY 3.0 Unported license
© Julian Steil

Joint work of Andreas Bulling, Yusuke Sugano, Mohsen Mansouryar

Main reference J. Steil and A. Bulling, “Discovery of Everyday Human Activities From Long-Term Visual Behaviour Using Topic Models”, in Proc. of the 2015 ACM Int’l Joint Conf. on Pervasive and Ubiquitous Computing (UbiComp), pp. 75–85, 2015.

URL <http://dx.doi.org/10.1145/2750858.2807520>

Practically everything that we do in our lives involves our eyes, and the way we move our eyes is closely linked to our goals, tasks, and intentions. Thus, the human visual behaviour has significant potential for activity recognition and computational behaviour analysis. In my talk I briefly discussed the ability of an unsupervised method to discover everyday human activities only using long-term eye movement video data [1]. Moreover, I presented a novel 3D gaze estimation method for monocular head-mounted eye trackers which directly maps 2D pupil positions to 3D gaze directions [2]. My current research is driven by the idea to shed some light on attention allocation in the real world.

References

- 1 Julian Steil and Andreas Bulling. *Discovery of Everyday Human Activities From Long-Term Visual Behaviour Using Topic Models*. Proc. of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp), pp. 75-85, 2015. <http://dx.doi.org/10.1145/2750858.2807520>

- 2 Mohsen Mansouryar and Julian Steil and Yusuke Sugano and Andreas Bulling. *3D Gaze Estimation from 2D Pupil Positions on Monocular Head-Mounted Eye Trackers*. Proc. of the 9th ACM International Symposium on Eye Tracking Research & Applications (ETRA), pp. 197-200, 2016. <http://dx.doi.org/10.1145/2857491.2857530>

3.23 Calibration-Free Gaze Estimation and Gaze-Assisted Computer Vision

Yusuke Sugano (Max-Planck-Institut für Informatik – Saarbrücken, DE)

License © Creative Commons BY 3.0 Unported license

© Yusuke Sugano

Joint work of Bulling, Andreas; Matsushita, Yasuyuki; Sato, Yoichi

Main reference Y. Sugano and A. Bulling, “Self-Calibrating Head-Mounted Eye Trackers Using Egocentric Visual Saliency”, in Proc. of the 28th ACM Symp. on User Interface Software and Technology (UIST’15), pp. 363–372, 2015.

URL <http://dx.doi.org/10.1145/2807442.2807445>

Human gaze can provide a valuable resource for eyewear computing systems to understand both internal states of the users (e.g., their activities) and external environments (e.g., scene categories). In this ignite talk I presented a brief overview of my previous researches from calibration-free gaze estimation to gaze-assisted computer vision techniques. One approach for calibration-free gaze estimation is to use visual saliency maps estimated from the scene video as probabilistic training data [1]. Another approach is to take a purely machine learning-based approach to train an image-based gaze estimator using a large amount of eye images with ground-truth gaze direction labels [2]. If gaze estimation can be naturally integrated into daily-life scenarios with these techniques, collaboration between eyewear computers and human attention will have larger potential for future investigation. For example, gaze can guide computers to find semantically important objects in the images [3], and is expected to provide important information for user- and task-specific image understanding.

References

- 1 Yusuke Sugano and Andreas Bulling, *Self-Calibrating Head-Mounted Eye Trackers Using Egocentric Visual Saliency*, in Proc. of the 28th ACM Symposium on User Interface Software and Technology (UIST), pp. 363-372, 2015.
- 2 Yusuke Sugano, Yasuyuki Matsushita and Yoichi Sato, *Learning-by-Synthesis for Appearance-based 3D Gaze Estimation*, Proc. 27th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2014), June 2014.
- 3 Yusuke Sugano, Yasuyuki Matsushita and Yoichi Sato, *Graph-based Joint Clustering of Fixations and Visual Entities*, ACM Transactions on Applied Perception (TAP), 10(2):1-16, June 2013.

3.24 Wearable barcode scanning

Gábor Sörös (ETH Zürich, CH)

License © Creative Commons BY 3.0 Unported license

© Gábor Sörös

URL <http://people.inf.ethz.ch/soeroesg/>

Ubiquitous visual tags like barcodes and QR codes are the most prevalent links between physical objects and digital information. Technological advancements in wearable computing

and mobile computer vision may radically expand the adoption of visual tags because smart glasses and other wearable devices enable instant scanning on the go. I develop robust and fast methods to overcome limitations and add advanced features that can make wearable barcode scanning an attractive alternative of traditional laser scanners. I present an overview of my research on tag localization, motion blur compensation, and gesture control on resource-constrained wearable computers.

References


- 1 Gábor Sörös, Stephan Semmler, Luc Humair, Otmar Hilliges, *Fast Blur Removal for Wearable QR Code Scanners*, Proc. of the 19th International Symposium on Wearable Computers (ISWC 2015)
- 2 Jie Song, Gábor Sörös, Fabrizio Pece, Sean Fanello, Shahram Izadi, Cem Keskin, Otmar Hilliges, *In-air Gestures Around Unmodified Mobile Devices*, Proc. of the 27th ACM User Interface Software and Technology Symposium (UIST 2014)
- 3 Gábor Sörös, Christian Floerkemeier, *Blur-Resistant Joint 1D and 2D Barcode Localization for Smartphones*, Proc. of the 12th International Conference on Mobile and Ubiquitous Multimedia (MUM 2013)

4 Workshops and Tutorials

4.1 Workshop: PUPIL

Moritz Kassner (*Pupil Labs UG*)

Will Patera (*Pupil Labs UG*)

License  Creative Commons BY 3.0 Unported license
© Moritz Kassner, Will Patera

Introduction

Will Patera and Moritz Kassner (co-founders of Pupil Labs) conducted a workshop using Pupil – eye tracking and egocentric vision research tool with the members of the Dagstuhl seminar. The workshop was organized in four general parts: an overview of to the Pupil platform, setup and demonstration demo, data collection in groups using Pupil, and concluding presentation of results from each group and feedback on the tool.

Hardware

Pupil is a head-mounted video based eye tracking and egocentric vision research tool. The headset material is laser sintered polyamide 12. Pupil attempts to reduce slippage by deforming the geometry of the headset prior to fabrication to make a comfortable and snug fit, and keeping the headset weight low. A binocular configuration with two 120hz eye cameras and one 120hz scene camera weighs 47 grams. The headset is modular and can be set up for egocentric vision, monocular eye tracking, or eye only cameras. The headset connects to a computer via high speed USB 2.0. Cameras used in the Pupil headset are UVC compliant.



■ **Figure 1** In the eye tracking workshop organised by Moritz Kassner and Will Partera (Pupil Labs UG, Berlin), seminar participants obtained hands-on experience with PUPIL, an open source platform for head-mounted eye tracking and egocentric vision research.

Software

The software is open source¹ Striving for concise, readable implementation and a plugin architecture. There are two main software applications/bundles; Pupil Capture for real-time and Pupil Player for offline visualization and analysis.

Working Principles

Pupil is a video based tracker with one camera looking at the scene and another camera looking at your pupil. Dark pupil tracking technique. Eye is illuminated in the infrared. Pupil detection is based on contour ellipse fitting and described in detail here in [1]. Performance was evaluated using the Swirski dataset described in [2]. For gaze mapping we need a transformation to go from pupil space to gaze space. One method is regression based gaze mapping. Mapping parameters are obtained from a 9 point marker calibration. Accuracy is 0.6 deg, precision 0.08 deg are nominal. Pupil dilation and movement of the headset degrades accuracy.

Recent Work

Current trackers assume the headset is "screwed to the head". You need to compensate for the movement of the headset. Additionally, a model-less search for the ellipse yields poor results when the pupil is partially obstructed by reflections or eyelashes or eyelids. With the use of a 3D model of the eye and a pinhole model of the camera based on Swirski's work in [3] we can model the eyeball as a sphere and the pupil as a disk on that sphere. The sphere used is based on an average human eyeball diameter is 24mm. The state of the model is

¹ <https://github.com/pupil-labs/pupil>

the position of the sphere in eye camera space and two rotation vectors that describe the location of the pupil on the sphere.

Using temporal constraints and competing eye models we can detect and compensate for slippage events when 2D pupil evidence is strong. In case of weak 2D evidence we can use constraint from existing models to robustly fit pupils with considerably less evidence than before.

With a 3D location of the eye and 3D vectors of gaze we don't have to rely of polynomials for gaze mapping. Instead we use a geometric gaze mapping approach. We model the world camera as a pinhole camera with distortion, and project pupil line of sight vectors onto the world image. For this we need to know the rotation translation of world and eye camera. This rigid transformation is obtained in a 3 point calibration routine. At the time of writing we simply assume that the targets are equally far away and minimize the distance of the obtained point pairs. This will be extended to infer distances during calibration.

Workshop Groups

At the workshop, participants split into eight groups to explore various topics related to head-mounted eye tracking and collect three small-scale datasets.

- Group 1 – Päivi Majaranta, Julian Steil, Philipp M. Scholl, Sabrina Hoppe – “Mutual Gaze” – used Pupil to study two people playing table tennis and synchronized videos.
- Group 2 – René Mayrhofer, Scott Greenwald – used Pupil to study the iris and see if eye cameras on Pupil could be used for iris detection.
- Group 3 – Thies Pfeiffer, Masashi Nakatani – used fiducial markers on the hands and computer screen and proposed a method to study typing skills on a “new keyboard” (e.g. German language keyboard vs American English language keyboard)
- Group 4 – Thad Starner, Paul Lukowicz, Rita – “Qualitative tests” – tested calibration and stability of the system
- Group 5 – Yusuke Sugano, Walterio Mayol-Cuevas, Gábor Sörös – “Indoor outdoor gaze” – recorded datasets in varying environments (indoor, outdoor, transition between) and varying activities: cycling, walking, vacuuming.
- Group 6 – James M. Rehg, Linda B. Smith, Ozan Cakmakci, Kristen Grauman – “Social Gaze Explorers” recorded gaze data during a 3 speaker interaction and experimented with simple approaches to detect when a wearer switches their attention from one speaker to another.
- Group 7 – Shoya Ishimaru, Koichi Kise, Yuji Uema – JINS Meme + Pupil – This group demonstrated a proof of concept combining JINS MEME EOG eye tracking device with Pupil. Pupil could be used to collect ground truth data for EOG systems like JINS MEME.
- Group 8 – Steven K. Feiner, Kiyoshi Kiyokawa, Masahiko Inami – used pupil while performing everyday tasks like making coffee and having conversations in a group.


References

- 1 Moritz Kassner, Will Patera, Andreas Bulling. *Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction*. Adj. Proc. of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp), pp. 1151-1160, 2014. <http://dx.doi.org/10.1145/2638728.2641695>
- 2 Lech Świrski, Andreas Bulling, Neil Dodgson. *Robust, real-time pupil tracking in highly off-axis images*. Proc. of the 7th International Symposium on Eye Tracking Research and Applications (ETRA), pp. 173-176, 2012. <http://dx.doi.org/10.1145/2168556.2168585>

- 3 Lech Świrski, Neil A. Dodgson. *A fully-automatic, temporal approach to single camera, glint-free 3D eye model fitting*. Proc. of the 3rd International Workshop on Pervasive Eye Tracking and Mobile Eye-Based Interaction (PETMEI), 2013.

4.2 Workshop: Google Cardboard

Scott W. Greenwald (MIT)

License  Creative Commons BY 3.0 Unported license
© Scott W. Greenwald

Scott Greenwald led a workshop on the Google Cardboard. Google Cardboard is a physical enclosure, consisting primarily of cardboard and two plastic lenses, and matching set of software constructs that turns almost any smartphone into a VR headset. Thanks to Ozan (Google Inc), the workshop was able to provide a Google Cardboard device for every participant to use and take home. Scott prepared a website that was used to document and distribute three live examples and a demo. The live examples could be opened directly on the website with the participants own mobile device. The examples and demo are described in the sections that follow.

Example 1: WebVR Boilerplate

When loading example using a web browser on a mobile device, users see an immersive animated video clip. They can pan around using touch gestures or moving their device in space. By pressing a Cardboard icon, they can switch into stereoscopic mode for viewing using Google Cardboard.

The example illustrates the seamless process of distributing cross-platform VR content



■ **Figure 2** Another workshop organised by Scott Greenwald (MIT) enabled participants to experience Google Cardboard using 30 devices kindly donated by Google Inc, US.

viewable on Cardboard. Mobile web browser technology allows this functionality to be built into a webpage and viewed on any sufficiently-equipped mobile device. In addition to being easier to distribute, this also accelerates the development workflow. The next important observation is that the same content was view with and without the VR headset. This is the concept of Responsive WebVR , an important step for making the VR-enabled web accessible on all devices.

The 3D environment the user sees in the example consists of a single sphere. The virtual scene camera through which the viewer is looking is located inside the sphere, and can be rotated using either touch gestures or physically moving the mobile device. The core mobile web technology that makes this possible is the WebGL part of the HTML5 specification. WebGL allows web pages to leverage mobile graphics hardware previously inaccessible inside the browser. Although it is possible to use the browser's built in WebGL language constructs directly, Scott advises using the Three.js framework which raises the level of abstraction and thereby simplifies the process of using the browser's WebGL capabilities.

Example 2: See-through video AR

This Cardboard-only video see-through app allows the user to use a button interaction to place playful markers in field of view that track the environment. It demonstrates a computer vision algorithm running on a live video stream in the mobile web browser. This is made possible by an HTML5 Media Capture API called `getUserMedia`. It allows the browser to access webcam and microphone hardware on the host device. For optical flow tracking, the example uses `jsfeat`, a JavaScript-based computer vision library.

Example 3: Social with WebRTC (Demo)

I demonstrate a two-device configuration where a cardboard user sees a video see-through application, and is able to see visual markers sent by a remote companion on a tablet and desktop system. The important aspect of this system is that it uses an HTML5 real-time communication technology called WebRTC. The example implementation shows the robustness of both `getUserMedia`, also shown in the previous example, and WebRTC for real-time sending of marker coordinates.

Example 4: Neuron Viewer

This example shows a high-resolution 3D model using the Responsive WebVR boilerplate (see Example 1). In contrast to the previous example building responsive WebVR, this example shows a high-resolution 3D geometry. This showcases the capabilities of mobile graphics hardware to not only show photographic content, but also polygon-based geometric content.

Conclusion

Most participants were able to successfully run most of the examples. Some participants' devices, however were insufficient or incompatible with the HTML5 technologies required. This demonstrates that, at this moment in time, the cross-platform vision of the mobile web is not yet fully realized in practice. Every device from the latest generation, as well as some devices several years old, were able to view the content. From this one can conclude that it is only a matter of time before this technology will represent a truly universal method for rapidly prototyping and distributing web-based VR and AR experiences for mobile devices.

4.3 Workshop: J!NS Meme

Shoya Ishimaru (Osaka Prefecture University)

Yuji Uema (JIN Co. Lt.)

Koichi Kise (Osaka Prefecture University)

License © Creative Commons BY 3.0 Unported license
© Shoya Ishimaru, Yuji Uema, Koichi Kise

Introduction of J!NS MEME

J!NS MEME is developed and released November 2015 by Japanese eyewear company, JIN CO.,LTD. Two significant differences between J!NS MEME and other traditional smart eyewear are long battery life and physical appearance. They last for around one day and look very close to normal eye wear.

They can stream sensor data to a second device (e.g. smart phone or computer) using Bluetooth LE. Sensor data includes vertical and horizontal EOG channels and accelerometer/gyroscope data. The runtime of the device is 12 hours enabling long term recording and, more important, long term real-time streaming of eye and head movement.

Applications of J!NS MEME to Reading-Life Log

I introduced our research of reading-life log by using J!NS MEME. The device J!NS MEME is not an eye tracker so that it is not possible for us to get eye gaze information; what we can get is the information about eye movement and eye blinks. Based on this information, we have implemented a wordometer, which counts the number of read words, and reading detector, which detects the period of reading in our daily activities.

These are examples for inspiring participants to think how to use J!NS MEME for their new applications.

Software for J!NS MEME

We introduced software development with J!NS MEME. According to primary asking, we had prepared three types of devices and their sensor logging softwares on several OS platforms (iOS/Windows/Mac). All materials and tutorial are available on <http://shoya.io/dagstuhl/>

Experiences of using J!NS MEME

We had 4 teams that tackled the following research topics with J!NS MEME.

- Head and Eye Gestures with MEME: The team tried several eye gestures and head gestures for user interaction and could present some initial ideas for unobtrusive gestures.
- Combination with Pupil: The team worked on integrating a Pupil eye-tracker with MEME and could show a working prototype.
- Haptic Feedback on EyeWear: The group presented a quick haptic feedback system attached to MEME for unobtrusive feedback.
- Eye movement influence on cognitive states: The group presented ideas on how to use the simple sensors of MEME to detect cognitive states (e.g. fatigue, disorientation, Alzheimer's and drug influence).

4.4 Tutorial: Head-Worn Displays

Kiyoshi Kiyokawa (Osaka University), Thad Starner (Georgia Institute of Technology), Ozan Cakmakci (Google)

License © Creative Commons BY 3.0 Unported license
© Kiyoshi Kiyokawa, Thad Starner, and Ozan Cakmakci

The head-worn display tutorial was organized into three sections: an ideal perspective, a user perspective, and an optical design perspective.

Ideally, head worn displays (HWDs) are expected to produce any visual experience we imagine, however, compared to this ultimate goal, current optical see-through (OST) HWDs are still far from perfect. In the first part of the HWD tutorial, we introduced introduced a number of issues that need to be tackled to make a 'perfect' OST-HWD. They include size, weight, the field of view, resolution, accommodation, occlusion, color purity, and latency. These issues are discussed by taking a number of research prototypes as example solutions. Then another fundamental problem of OST-HWDs is introduced that is, the user's exact view is not accessible as image overlay happens on the user's retina. A new concept of corneal feedback augmented reality (AR) is then introduced as a promising approach to realize a closed feedback loop for better registration, color correction, contrast adjustment, accurate eye tracking and scene understanding, by continuously analysing the reflections in the eye. In the end of the talk, by taking visualization of motion prediction as an example, it is emphasized that sensing is the key to success not only for better visual experience but also for more advanced applications.

Consumer purchase and use of eyewear is driven more by fashion than just about any



■ **Figure 3** The first tutorial provided an introduction to head-mounted displays (Kiyoshi Kiyokawa, Osaka University), insights into the development of Google Glass (Thad Starner, Georgia Institute of Technology), as well as a 101 on the design of optical systems (Ozan Cakmakci, Google).

other feature. Creating head-worn displays (HWDs) for everyday use must put fashion first. Due to lack of familiarity with the use of the devices, HWD consumers often desire features that are impractical or unnecessary. Many desire wide field-of-view (FOV) displays. However, the human visual system only sees in high resolution in a small 2 degree area called the fovea. Small FOV consumer displays (i.e., smart phones) have been highly successful for reading books, email, messaging, mobile purchasing, etc. A typical smart phone has a 8.8x15.6 degree FOV. Small FOV displays are much easier to manufacture with fashionable eyewear and will likely be the first successful (> 2m) consumer HWD product. Head weight should be kept under 75g for everyday use.

From the point of view of optical design, most researchers in the field of eyewear computing rely on existing commercially available head-worn displays to do their research and development. In the last part of the tutorial, we illustrated the optical design process by taking a deep sea diving telemetry display as a case study. The optical design process of a visual instrument involves understanding the design parameters, for example, image quality, field of view, eyebox, wavelength band, and eyerelief. We started the optical design monochromatically with a singlet lens to illustrate the limiting monochromatic aberrations. Next, we evaluated the optical performance with a photopic spectrum, and discussed approaches color correction. The use of transverse ray aberration plots was emphasized as a debugging tool. The main point in the last part of the tutorial was to remind researchers that it is possible to design custom optics at reasonable cost in contrast to relying solely on commercially available displays.

The participants got a chance to experience Google Glass (mirror based magnifier), Optinvent ORA (flat lightguide with collimation optics), Epson BT-200 (flat lightguide with freeform outcoupler), Triplett Visualizeyezer (concept of a beamsplitter), and Rift 1 (single lens based magnifier) during the workshop as part of the tutorial. We used the various hardware to illustrate eyebox, field of view, chromatic aberrations, distortion, and brightness.

4.5 Tutorial: Egocentric Vision

James M. Rehg (Georgia Institute of Technology)

Kristen Grauman (University of Texas)

License  Creative Commons BY 3.0 Unported license
© James M. Rehg, Kristen Grauman

The video produced by a head-worn camera possesses a key property – the motion of the camera through the scene is fundamentally guided by the intentions and goals of the camera-wearer. As a consequence, first person video implicitly contains potent cues for scene understanding that can be leveraged for video analysis. This tutorial consisted of two parts. The first part discussed specific egocentric cues, such as head motion and the locations of the hands, that can be extracted from egocentric video and utilized for predicting the attention and activities of the camera-wearer. The second part discussed representation learning and the inference of attention to drive the summarization of egocentric video.

Egocentric video provides a unique opportunity to capture and analyze the visual experiences of an individual as they go about their daily routines. The first part of the tutorial demonstrated that egocentric cues such as head motion and hand location can be used to predict the attention of the first person as they perform hand-eye coordination tasks such as activities of daily living. Using a dataset of cooking activities (GTEA+), we demonstrated

that accurate predictions of the the user's gaze could be obtained through temporal fusion of egocentric cues. These same cues can also be used as features in performing activity recognition. We showed that egocentric features can be used to complement and improve classical features such as Improved Dense Trajectories (IDT) for an activity recognition task. We presented a method for stabilizing egocentric video to remove the effect of head motion on IDT, leading to increased accuracy in using that feature. We also demonstrated the added benefit of using head motion and hand positions as features, beyond the baseline provided by the standard IDT approach. These findings demonstrate the unique properties of egocentric video and the existence of specific egocentric features which provide an improvement beyond classical video representations.

The fact that the camera wearer is an active participant in the visual observations received has important implications for both (1) representation learning from egocentric video and (2) understanding cues for attention. The second part of the tutorial overviewed work on both of these fronts. In terms of representation learning, we discussed how egomotion that accompanies unlabeled video can be viewed as an implicit and free source of weak supervision. The goal is to learn the connection between how the agent moves and how its visual observations change. Whereas modern visual recognition systems are focused on learning category models from "disembodied" Web photos, often for object class labeling, the next generation of visual recognition algorithms may benefit from an embodied learning paradigm. We presented one such attempt, based on deep learning for representation learning that regularizes the classification task with the requirement that learned features be equivariant. In terms of attention cues, we surveyed ideas for exploiting the information about a camera wearer's gaze, pose, and attention to (a) compute compact summaries from long egocentric videos, (b) estimate when a video frame passively captured on the camera looks like an intentionally framed shot, and (c) detecting moments of heightened engagement where the camera wearer is examining his/her environment to gather more information about something. In all cases, learning based methods are being developed to capture the special informative structure from egocentric video. Key future directions for these lines of work include dealing with multi-modal sensing, transforming the learned view predictive representations to tackle active vision problems, exploring novel forms of visualization of a summary, and identifying scalable methods for quantitative evaluation of video summaries.

5 Challenges

To take advantage of the opportunities for sensing and interaction offered by near eye (or more general head mounted) systems significant computational resources are needed, especially for processing of video signals, advanced information fusion and long term learning. While in some cases processing can be outsourced to remote devices (or the cloud), in others local processing on the device is desirable. Such processing must take into account stringent power, thermal and space constraints specific to the near eye location. Thus, appropriate digital and mixed signals architectures specifically combining special purpose circuits with appropriate, reconfigurable components and low power general purpose processing capabilities need to be developed.

In most worn computing systems, four major inter-related technical challenges must be addressed: power & heat, networking (on & off-body), interface, and privacy. For eyewear computing, interface can be further divided into several sub challenges: head weight, fashion, and attention. Advances in electronics will continue to drive improvements in power &

heat, networking, and head weight. However, there is much work yet to be done in creating low-attention interfaces that are appropriate for use in everyday environments. Eyewear computing interfaces should be designed to be secondary to a user's primary task and distract minimally from it. Similarly, depending on the social situation of use, the appearance of the electronics (fashion) and the use of the electronics (interaction) should not distract bystanders, colleagues, or conversational partners from their primary tasks.

This seminar has suggested possible directions toward improving interactions with eyewear computing. Can context-awareness, leveraging the access to the user's senses afforded by eyewear computing, help deliver content to the user unobtrusively and at the right moment? Can eye, head, and or body movement be used to provide both explicit and implicit input to a eyewear computing agent that can assist the user in day-to-day tasks? These questions hint at another, more ambitious agenda: can we use eyewear computing's unique first person perspective to create an intelligent assistant that learns to live and interact in the human world? Given the recent interest and improvements in techniques that leverage "big data" to gain new insights into human-level problems like speech recognition and parsing images, initiatives leveraging eyewear computing might be well poised to gain levels of understanding of what it means to be human that systems restricted to data from the web can not attain.

6 Breakout Sessions

6.1 Group 1: Multimodal EyeWear Computing

Masashi Nakatani (University of Tokyo, JP)

License  Creative Commons BY 3.0 Unported license
© Masashi Nakatani

Joint work of Gábor Sörös, Moritz Kassner, Ozan Cakmakci, Päivi Majaranta, Sabrina Hoppe, Will Patera, Yuji Uema

Recent advancement of eyewear device and computing allows us to use the equipment in multiple contexts. Sensory substitution and sensuarization is one example, in which the system can make captured infomation into more perceptually detectable [1]. The use of eyewear computing should not be limited to visual information. Recent advancement of multimodal study in perceptual psychology and engineering implentation allow to integrate eyewear into multimodal interface. One representative applications of multimodal integrations is sensory substitution, in which people who has sensory impairments (mostoly vision and auditory sensation).

In addition to multimodal integration as a device, potential collaborative study with cognitive psychology is now getting to be expected. As old syaing states, "Look into my eyes and hear what I'm not saying, for my eyes speak louder than my voice ever will." Based on this intuition, a bridge between eye-wear and cognition attracts attentions from old ages.

Taking into these aspects into account, this section describes potential collaborative research fields in eyewear computing, and discuss possible research directions in the future.

Auditory

Gaze pointing has been proposed to use for producing sound and music composition [3] [4].

In virtual reality application, 3D spatial audio information is important for producing the feeling of immersion and presence of presented graphics. Based on this insight, 3D audio information has been already integrated in the application of Google Cardboard and Oculus Rift, and Apple iOS is plausible platform for including gaze based audio feedback.

Haptic

Touch modality has been considered to be beneficial for hearing impaired people to give feedback through the skin. Haptic feedback is provided based on visual [2] and audio information[5] well as audio feedback[6].

The advantage of using haptic feedback is that this information is relatively private comparing with visual and audio feedback.

Haptic Gaze Interaction is one of promising research direction of eyewear computing with haptic feedback[7]. A series of this project has been initiating basic research on combining of eye-pointing and haptic interaction. There are various open opportunities for natural and efficient human-computer interaction.

Olfactory

Olfactory sensation has a good connection with memory. One person may retrieve his/her memory by smelling certain odor, and neuroscientists are pursuing the neural basis of this phenomenon.

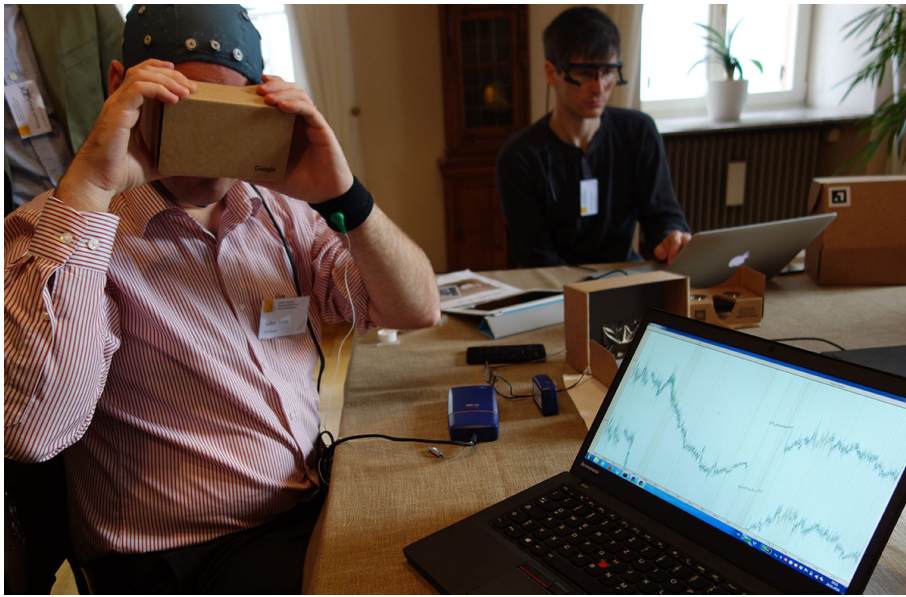
Consideration the combination with eyewear, it is also interesting hypothesis to test whether vision-odor conversion may increase the ability of memorizing certain tasks in the office or in the class. Zoladz and Raudenbush reported that odorant may enhance cognitive processing [8]. This suggests a potential that visual-scene-odor conversion may also beneficial for office workers.

Brain Activity

Electroencephalography (EEG in short) is now popular method in cognitive psychology to access brain activity during daily behaviors. Comparing with other brain imaging techniques (fMRI, MEG, NIRS), participants can relatively freely move during the task. In addition to that wearable EEG is commercially available (cite Emotive) with reasonable price. Exploiting EEG as an even more direct pathway to the brain to create adaptive or assistive user interfaces.

Examining brain activity in virtual reality environment may attract attentions from industrial and academic interests. Not only using conventional VR environment using multiple screens and projectors, but also recent proposal of Google cardboard helps to test the relationship between eye-movements and visual feedback in immersive environment.

One potential disadvantage is that to obtained information from EEG and traditional eye wear sensors may be redundant, so that experimenters need to consider carefully before conducting actual experiment. Eyewear researchers will benefit from collaborative study with cognitive scientists.



■ **Figure 4** Hands-on trial of an EEG measurement while experiencing VR through the Google Cardboard. Seminar participants combined their expertise on site and conducted preliminary experiments.

Facial Expressions

Electromyography enables using facial expressions such as smiling or frowning to select things that are currently under visual focus. Wearing such a head-mounted "face interface" enables interaction based on voluntary gaze direction and muscle activation. For more information, see e.g. [9].

Memory

Eyewear computing has also impacts on the therapy of trauma. Psychotherapy study using Eye Movement Desensitization (EMD) reported that this technique is beneficial for decreasing traumatic experience (reference). EMD is the procedure in which patients access to their traumatic memories in the context of a safe environment, the hypothesis is that information processing is enhanced, with new associations forged between the traumatic memory and more adaptive memories or information. These new associations result in complete information processing, new learning, elimination of emotional distress, and the development of cognitive insights about the memories. Although eyes move for any memory, EMD is still effective for some patients. If eyewear computing research can collaborate with medical and cognitive scientists studying memory and its psychotherapy, both side would get benefits in the advancement of device and human emotions.

Conclusion

There are multiple potentials of eyewear computing by combining with other modality or for practical use in medical science. Close relationship between eyewear computing community

and other research fields would be critical for the prosperity of both side of the research. We ended up the discussion by encouraging researchers to have multiple interests in basic and applied study of eyewear computing.

References

- 1 Péter Galambos, András Róka, Gábor Sörös, Péter Korondi. *Visual feedback techniques for telemanipulation and system status sensualization*, Proc. of IEEE 8th International Symposium on Applied Machine Intelligence and Informatics (SAMI), pp. 145–151, 2010.
- 2 Matthias Berning, Florian Braun, Till Riedel, and Michael Beigl. *ProximityHat: A Head-worn System for Subtle Sensory Augmentation with Tactile Stimulation*. ISWC'15 Proceedings of the 2015 ACM International Symposium on Wearable Computers, pp. 31–38, 2015.
- 3 Hornof, A., Sato, L. (2004, June). EyeMusic: making music with the eyes. In Proceedings of the 2004 conference on New interfaces for musical expression (pp. 185–188). National University of Singapore.
- 4 Hornof, A. J., and Vessey, K. E. (2011, September). The Sound of One Eye Clapping Tapping an Accurate Rhythm With Eye Movements. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (55,1pp. 1225–1229). SAGE Publications.
- 5 Novich S, Eagleman D. *Using space and time to encode vibrotactile information: toward an estimate of the skin's achievable throughput*. Experimental Brain Research. 233(10), pp. 2777-2788, 2015.
- 6 Peter B. L. Meijer. *An Experimental System for Auditory Image Representations*. IEEE Transactions on Biomedical Engineering, vol. 39, no. 2, pp. 112–121, 1992.
- 7 Jari Kangas, Deepak Akkil, Jussi Rantala, Poika Isokoski, Päivi Majaranta, and Roope Raisamo. *Gaze gestures and haptic feedback in mobile devices*. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'14). ACM, New York, NY, USA, pp. 435–438, 2014.
- 8 Phillip R. Zoladz, Bryan Raudenbush, *Cognitive Enhancement Through Stimulation of the Chemical Senses*, North American Journal of Psychology, Vol. 7 Issue 1, pp. 125–140, 2005.
- 9 Tuisku, O., Surakka, V., Vanhala, T., Rantanen, V., and Lekkala, J. (2012). Wireless Face Interface: Using voluntary gaze direction and facial muscle activations for human–computer interaction. *Interacting with Computers*, 24(1), 1–9.

6.2 Group 2: Egocentric Vision

Rita Cucchiara (University of Modena)

Kristen Grauman (University of Texas)

James M. Rehg (Georgia Institute of Technology)

Walterio W. Mayol-Cuevas (University of Bristol)

License © Creative Commons BY 3.0 Unported license

© Rita Cucchiara, Kristen Grauman, James M. Rehg, Walterio W. Mayol-Cuevas

Joint work of Andreas Bulling, Rita Cucchiara, Kristen Grauman, Kiyoshi Kiyokawa, James M. Rehg, Linda B. Smith, Julian Steil, Yusuke Sugano, Walterio W. Mayol-Cuevas, and Masahiko Inami

The ego-centric view is one created by the wearer's own actions and momentary goals. The visual properties of these ego-centric scenes and videos have their own properties (hands, center bias for attended objects, information reduction). They are also inherently connected to the wearer's movements – eye gaze, head movements, hand movements, whole body movements. Egocentric vision is an emerging field of computer vision, specially devoted to defining models, algorithms and techniques dealing with egocentric video, generally captured

by eyewear devices. It exploits geometry, pattern recognition and machine learning paradigms to understand ego-centric scenes, recognizing objects, actions and the interaction between the wearer, the persons and the surrounding environment. It addresses problems of personalized video summarization [19], relevance determination for detecting what is important and how things are used for user guidance [11], and the prediction of attention [20] and activities.

In this breakout session, we discussed (a) the links between movement and visual learning in animals, humans, and machines, (b) its relation to supervised and unsupervised learning, (c) the challenges of egocentric vision in the wild considerable body movements and lighting changes and (d) and relation other frames of reference for capturing the visual information.

Movement and Visual Learning

Visual experience in the context of planned and goal-directed movements are known to enhance learning and change patterns of brain connectivity in animals and humans [1, 2]. There are possible synergies between computational and theoretical approaches cross biological and machine learning. Particularly relevant may be leveraging the structure in multimodal information in goal-directed action. In biological learning, computational models have been commonly based on prediction [3] and re-entrance (see also [5] for recent review).

Different actions and different tasks may prevent different solutions for computer vision systems as they yield different patterns of egocentric vision and different patterns of eye, head, and body movements. The role of hands, object manipulations and different patterns of movements in goal directed tasks of different kinds also merit consideration. When the larger multimodal properties of human are considered in natural real-world tasks, the question arises as to whether foveation and saccades are of central importance or whether these might be considered attributes of human vision specific to specific tasks such as reading but which do not deeply inform human visual behavior in more active contexts nor perhaps computer vision systems ([6, 7]).

The current status of computational vision doesn't speak to how children learn to recognize objects, which is robust at a very early age (see, [8]). Current computational vision with its emphasis on static images may not be very useful for autonomous vehicles or for building learning robots that improve through their own experience. Egocentric vision, by leveraging the properties of goal-directed purposeful action, provides potentially transformative domain for the classic questions of machine vision including object segmentation and object recognition.

Issues in Weakly/Unsupervised Learning

A current trend in computer vision is exploiting machine learning, and in particular Deep Learning, to support the tasks of object and activity recognition. This trend holds for egocentric vision as well. Nevertheless, although recent improvements in recognition performance may seem impressive, these gains have been made in closed world datasets where the space of object labels is fixed. In contrast, when a person is moving in the real world, they continuously encounter novel objects and unique situations, and classifiers trained with closed world datasets cannot be easily adapted. This raises the question of egocentric learning from weak supervision. Large amounts of video can be easily acquired, how can we utilize this data if labels are available only sparsely? For example, can we acquire models of objects

based on observations of how they are used in performing actions? Semi-supervised learning can be used for recognizing action patterns or objects that have very few annotated examples: an example is hand gesture recognition, which is frequently characterized by personal, and often unique, gestures. In this case a robust hand segmentation process can be coupled with a semi-supervised process of recognizing actions from very few examples. [19].

Egocentric Vision and Neuroscience

The neuroscience aims at understanding how information sensed by eyes become vision, how thoughts become memory, how visual behavior comes from biology. This is very important to understand the human vision and how can this be exploited in computer vision too. According with Kandel (Kandel 2012) the cortical view suggests that understanding what is important from an observer is a mix of recognition and tracking. That can be viewed in the manner artists approached the problem of objects' shape, movement and 3D representation in paintings, and is strongly connected with their exploitation of the cortical vision in the "way of what" and the "way of where". Therefore combining semi-supervised recognition, object tracking, action inferring from a first person view is still the big issue.

Egocentric Vision and Other Sensors

Just as humans perceive via multiple modalities, so egocentric perception could benefit from leveraging additional sensing modalities. It is an open issue how to use other sensors and augmentation techniques to provide additional cues to inform image understanding. From power standpoint, vision is a very expensive sensor, and it could be interesting to exploit less expensive sensors to cue vision and reduce the overall power budget. One challenge we face is that while on-body signals from a variety of wearable sensors are likely to be inter-related, the nature of the relationships can change with the task and over time. For example, signals are related as a result of basic physiological processes [21] as well as via the task the subject is performing. In the case of object manipulation tasks, two sources of difficulty are the challenge of reliably estimating motion and the lack of effective object-level representations.

Relationship Between First Person and Third Person Vision

What is the relationship between first person vision and other types of camera positions and movements? We could define three kinds of egocentric vision: Eyeball imaging, Head-mounted imaging, Omnidirectional imaging. This first, still far to be commonly used, will extract the precise information of where the people are fixating the gaze. What are the natural statistics of the visual world via egocentric vision? For example, since cameras move smoothly through the scene there is continuity of space and time which produces a power law distribution of types and tokens in video sequences. So temporal continuity potentially provides a very strong prior that should help to solve the problem. We may need to revisit all of the standard problems in vision from the standpoint of egocentric vision. Segmentation, perception of form, recognition, etc. But there are other applications in eye wear computing such as human augmentation and of accessibility applications such as visually-impaired.

Attention Models

Egocentric vision may require a rethink about our attention models so that what is sensed, recorded or processed is relevant to a task or tasks. Such an *active sensing* approach finds similarities with what can be inferred from the way people make high-level decisions on where to focus visual and processing resources. Eye gaze patterns being a clear example of a gated visual interaction with the world. Current attention models used in the egocentric vision literature are based on eye fixations [10, 11], image features [12] or head motion [13]. However there is evidence in the human vision literature, of tasks where fixations may not be used [14, 6] and in egocentric systems where a broader peripheral sensing may be sufficient for activity detection or navigation tasks [16, 15]. There is thus a need to consider what attention models our eyewear systems will benefit from with the possibility that these are task-driven.

References

- 1 Held, R., Hein, A. (1963). *Movement-produced stimulation in the development of visually guided behavior*. Journal of comparative and physiological psychology, 56(5), 872.
- 2 James, Karin Harman. *Sensori-motor experience leads to changes in visual processing in the developing brain*. Developmental science 13, no. 2 (2010): 279-288.
- 3 Lake, B. M., Salakhutdinov, R., and Tenenbaum, J. B. (2015). *Human-level concept learning through probabilistic program induction*. Science, 350(6266), 1332-1338.
- 4 Sporns, O., Gally, J. A., Reeke, G. N., Edelman, G. M. (1989). *Reentrant signaling among simulated neuronal groups leads to coherency in their oscillatory activity*. Proceedings of the National Academy of Sciences, 86(18), 7265-7269.
- 5 Byrge, L., Sporns, O., Smith, L. B. (2014). *Developmental process emerges from extended brain-body-behavior networks*. Trends in cognitive sciences, 18(8), 395-403.
- 6 Foulsham, T., Walker, E., Kingstone, A. (2011). *The where, what and when of gaze allocation in the lab and the natural environment*. Vision research, 51(17), 1920-1931.
- 7 Kingstone, Alan, Daniel Smilek, Jelena Ristic, Chris Kelland Friesen, and John D. Eastwood. *Attention, researchers! It is time to take a look at the real world*. Current Directions in Psychological Science 12, no. 5 (2003): 176-180.
- 8 Bergelson, E., Swingle, D. (2012). *At 6–9 months, human infants know the meanings of many common nouns*. Proceedings of the National Academy of Sciences, 109(9), 3253-3258.
- 9 Eric Kandel *The Age of Insight*, 2012
- 10 Alireza Fathi, Yin Li, James M. Rehg *Learning to Recognize Daily Actions using Gaze*. ECCV, 2012.
- 11 D Damen, T Leelasawassuk, O Haines, A Calway, W Mayol-Cuevas. *You-Do, I-Learn: Discovering Task Relevant Objects and their Modes of Interaction from Multi-User Egocentric Video*. British Machine Vision Conference (BMVC), Nottingham, UK. 2014.
- 12 B. Xiong and K. Grauman. *Detecting Snap Points in Egocentric Video with a Web Photo Prior*. In Proceedings of the European Conference on Computer Vision (ECCV), Zurich, Switzerland, Sept 2014.
- 13 T Leelasawassuk, D Damen, W Mayol-Cuevas. *Estimating Visual Attention from a Head Mounted IMU*. International Symposium on Wearable Computers (ISWC). 2015.
- 14 Franchak, J. M., Adolph, K. E. (2010). *Visually guided navigation: Head-mounted eye-tracking of natural locomotion in children and adults*. Vision research, 50(24), 2766-2774.
- 15 M Milford, *Visual Route Recognition with a Handful of Bits*. Robotics: Science and Systems, 2012.
- 16 B Clarkson, K Mase, A Pentland, *Recognising User's context from wearable sensors: baseline system*, MIT VisMod Tech Report No 519, March, 2000.

- 17 Baraldi, L., Paci, F., Serra, G., Benini, L. Cucchiara, R. *Gesture Recognition in Ego-Centric Videos using Dense Trajectories and Hand Segmentation*, Proc. of 10th IEEE Embedded Vision Workshop at CVPR14, 2014
- 18 Lee, K.J., Grauman, K., *Predicting important objects for egocentric video summarization*, International Journal of Computer Vision, 1-18n 2015.
- 19 Cucchiara, R., Varini, P., Serra, G. *Personalized Egocentric Video Summarization for Cultural Experience* Proc. of the ACM Int. Conf. on Multimedia Retrieval, 2015
- 20 Li, Y., Fathi, A., and Rehg, J.M. *Learning to Predict Gaze in Egocentric Video*, In Proc. IEEE Intl. Conf. on Computer Vision (ICCV), Sydney, Australia, Dec 2013.
- 21 Hernandez, J., Li, Y., Rehg, J.M., and Picard, R.W. *BioGlass: Physiological Parameter Estimation Using a Head-mounted Wearable Device*, In Proc. Intl. Conf. on Wireless Mobile Communication and Healthcare (MobiHealth), 2014

6.3 Group 3: Security and Privacy

René Mayrhofer (Johannes Kepler University Linz)

License  Creative Commons BY 3.0 Unported license
© René Mayrhofer

As a relatively new scientific area, Eyewear Computing has not yet attracted much attention to security and privacy issues. Nonetheless, even current products and prototypes show some of the challenges that will need to be addressed before wide adoption is possible (or should be aimed for):

Device-to-user authentication counters the threat of devices being physically replaced by similar-looking but modified devices [1]. If users cannot be sure that the device they are about to put on is really their own, then they may divulge information to third parties or be presented with false information. Device-to-user authentication is highly related to device security – the former addresses the threat physical tampering, the latter of software tampering.

User-to-device authentication prevents the opposite threat of other (potentially malicious) people using a device and abusing the credentials associated with it or the data directly stored locally on the device. This can be (mostly) considered a solved problem for smart phones (with fingerprint readers as one reasonable compromise between security and usability, PIN/passwords, unlock patterns, and other methods already being used in practice [2]), but is still an open problem for eyewear devices.

Device-to-device authentication is required as soon as devices communicate wireless with other devices, as wireless radio links are not open to human senses. Proving to human users which other devices their own are communicating with (sometimes referred to as the human-in-the-loop property [3]) is typically achieved with *pairing* of devices, e.g. for Bluetooth connections. Pairing is an appropriate approach to long-term device links (such as eyewear devices to smart phones), but scales poorly for short-term interactions (such as using printers, projectors, etc. in the infrastructure).

Device security itself is an open challenge for many device categories (with smart phones being the most well-known at this time), and any eyewear devices will face the same issues due to complexity of the software stack they run [4]. However, assuming a similar set of sensors, communication interfaces, and platform, the potential implications of eyewear devices being subverted remotely are much more severe: using subliminal signals

on displays that are always visible, the risk for users being susceptible to manipulation is significantly higher.

Privacy of the wearer is an issue for all wearable devices connecting to and transmitting data into cloud services. The solution space is currently being explored for smart phones and will probably apply to eyewear display when considering that some additional sensors may be continuously recording.

Privacy of others is more difficult to guarantee, given the typical integration of cameras and microphones. Depending on the relevant legal system, it may not even be permissible to use standard eyewear devices with currently missing privacy safeguards in some settings. New technical approaches such as privacy-preserving filters close to the respective sensor may be required.

Although there are many unsolved issues, there is also the potential for eyewear devices to make some security issues easier to address. By relying on the fact that these devices are typically very close to the body during the day, continuous cross-device authentication may be used to more easily and more quickly unlock other devices of the same user [5]. By utilizing the camera, microphone, or other built-in sensors, eyewear devices could become proxies for spontaneous authentication to infrastructure services. By giving feedback on other devices and services to their users, they could improve awareness of potential security and privacy issues and provide improved transparency. Summarizing, security and privacy challenges are still largely unexplored in the specific context of eyewear computing, and novel solutions may be significantly different from current approaches to smart phone security.

References

- 1 Rainhard D. Findling, Rene Mayrhofer. *Towards Device-to-User Authentication: Protecting Against Phishing Hardware by Ensuring Mobile Device Authenticity using Vibration Patterns*. 14th International Conference on Mobile and Ubiquitous Multimedia (MUM'15), 2015
- 2 Daniel Hintze, Rainhard D. Findling, Muhammad Muaaz, Sebastian Scholz, Rene Mayrhofer. *Diversity in Locked and Unlocked Mobile Device Usage*. Adjunct Proceedings of ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2014). 379-384, 2014
- 3 Rene Mayrhofer. *Ubiquitous Computing Security: Authenticating Spontaneous Interactions*. Vienna University, 2008
- 4 Rene Mayrhofer. *An Architecture for Secure Mobile Devices*. Security and Communication Networks, 2014
- 5 Daniel Hintze, Rainhard D. Findling, Muhammad Muaaz, Eckhard Koch, Rene Mayrhofer. *Cormorant: towards continuous risk-aware multi-modal cross-device authentication*. Adjunct Proceedings of ACM International Joint Conference on Pervasive and Ubiquitous Computing and Symposium on Wearable Computers. 169-172, 2015

6.4 Group 4: Eyewear Computing for Skill Augmentation and Task Guidance

Thies Pfeiffer (Bielefeld University)

Steven K. Feiner (Columbia University)

Walterio W. Mayol-Cuevas (University of Bristol)

License © Creative Commons BY 3.0 Unported license

© Thies Pfeiffer, Steven K. Feiner, Walterio W. Mayol-Cuevas

Joint work of Steven K. Feiner, Scott Greenwald, Shoya Ishimaru, Koichi Kise, Kai Kunze, Walterio W. Mayol-Cuevas, René Mayrhofer, Masashi Nakatani, Thies Pfeiffer, Philipp M. Scholl, Gábor Sörös

Definitions/Distinctions

Assistance provided by eyewear computing can be classified as either skill extension or skill training. In *skill extension*, smart glasses provide assistance that goes beyond human capabilities, essentially offering super-human abilities. Examples include see-through views, such as now featured for some cars, and automatic language translation. On the other hand, in *skill training* or *skill support*, users are helped to be better at their normal skills (e.g., in terms of accuracy, speed, or quality).

A second important distinction is whether assistance is provided before, during, or after task performance. Of these possibilities, providing assistance during task performance places real-time constraints on the process. Assistance provided in advance can prepare and train the user, while assistance provided after the task, can help the user evaluate how well they did.

Assistance can also be categorized by the degree to which it is responsive to the user's actions. For example, eyewear might passively overlay relevant information on the user's environment, such as the unchanging 3×3 grid that many camera user interfaces can display to encourage the user to use the “rule of thirds” in composing an image. In contrast, eyewear could actively guide the user, in this case by steering the user to achieve an image composition based on this heuristic [14].

Application Areas for Eyewear Computing Related to Skills

- **Learning / Teaching.** Many countries have identified digital learning and online learning as an important component of the current and future educational system. Eyewear computing could support adaptive, personalized learning experiences in contexts beyond the desktop (i.e., beyond the classroom or the office). Learning material could be restructured and paced according to the personal needs, learning progress, and competences of the learner, even in group learning situations. One important motivational aspect could be gamification. Eyewear could not only educate the learner, but also provide valuable feedback to the teacher. There is a chance that personalized learning materials could also support better social integration by balancing expertise effects in learning situations by presenting different levels of assistance to individual learners. For example, if the teacher raises a question, advanced learners could be presented with four answers that are difficult to decide upon, whereas others could have easier choices. While the number of the correct answer could be the same for each student, the decision task could thus be adapted to their individual competences. The key point is that the learners would not be directly aware of that and thus social implications could be reduced. This might motivate weaker pupils to participate better during class.
- **Music / Physical Skill Training.** Training physical skills is a different challenge. How, for example, could eyewear improve the skills of a pianist? The answer is not straightforward,

but we discussed several possible approaches: visualizing correct hand postures, visualizing target keys (how useful this would be is questionable, as early hardware versions have been employed since the mid 20th century – e.g., Thomas Organ Color-Glo, https://en.wikipedia.org/wiki/Thomas_Organ_Company), or having an augmented score that adapts to the current level of expertise of the pianist (e.g., automatically updating so that problematic passages are repeated more often or are varied according to a training program).

- **Privacy.** There is a common conception that smart glasses can compromise privacy; however, there are some cases in which they might instead make it possible to maintain privacy. For example, in a future scenario in which smart machines adapt to a user’s level of expertise, their reactions toward a specific user could reveal the user’s competences to all onlookers (e.g., has difficulty using a coffee maker, confuses left and right, or forgets account information). Similarly, personalized advertising that is viewable by others could be embarrassing. Smart glasses could instead keep this information private. This could be especially attractive to professionals who would like to look up information on the fly while interacting with others. For example, a physician might not want to demonstrate their lack of knowledge about a recent study when talking with a patient who has just heard a news story about that study. Smart glasses could make it possible for the physician to find relevant information during the conversation; with the right user interface that could be done surreptitiously.
- **Communication Training / Therapy.** Communication training has been successfully used for children with autism using VR [11]. VR training applications for public speaking have long been developed and evaluated [10, 2] and are now available for consumer devices, such as Google Cardboard and Samsung Gear VR [13].
- **Task Motivation.** Gamification and other concepts are not directly affecting skill training, but help to maintain or create a necessary level of motivation. One important challenge is to design a system that helps the user to maintain a high level of motivation in the long run. In summary, that depends on the feedback and the intrinsic motivation of the users. This would add a strong A.I. component to eyewear computing.
- **Digital Memory.** A remembrance aid could use face and name recognition, and remind the user of their previous interactions with someone whom they may have forgotten.
- **Replacing/Augmenting/Assisting Senses.** Possibilities include improved hearing with noise cancellation and diminished reality [8] to suppress parts of the environment that the user wishes to avoid (e.g., deleting advertisements). But, note that expurgating things that the user wishes to avoid may keep important problems from being addressed.
- **Behavior Change.** Some aspects of interacting with a machine might increase the openness of persons in comparison with how they talk with other people [4].
- **Surveillance.** Eyewear can make surveillance possible without the need to consciously attend to capturing video. For example in a disaster scenario, such as an earthquake or nuclear plant accident, users need to concentrate on their own safety and that of others.
- **User Behavior Studies.** In particular in working contexts. For example in ISS (even though most of behaviors are already captured in current system)

Key “Selling Points” of Eyewear Computing for Augmentation/Guidance

This is a non-exhaustive list of selling points for eyewear computing that were collected from comments during the discussion.

- **Embedded display.** AR allows us to present information directly embedded in the context of real-world action, eliminating the need for the user to look back and forth between the real world action and a separate information source [5]. This can be accomplished by no other technology.
- **Subtle cueing.** Sometimes we do not need a fully featured AR, but subtle cueing, such as stimulating a rhythm while doing physical exercises or displaying subtle, possibly imperceptible, visual prompts to direct a user's attention [12, 6].
- **Shared attention.** Eyewear shares the wearer's attention and is thus very close by to the current interaction context.
- **Combined sensing and action/display.** Eyewear provides sensing and contextual presentation in one device.

Challenges

- Knowledge about multimodal communication/feedback channels between participants, in particular in a learning environment with teacher–pupil interactions, is essential for creating supportive eyewear applications for education.
- Context awareness could be a key feature of AR and smart glasses, yet to identify the state (emotional, cognitive, physiological) and the competences of the individual user (wearer) or interaction partner is a huge challenge. This needs to be tackled to be able to provide the right feedback to the user.
- Haptics could be important (e.g., for skill training), but it is an open question how to provide haptic feedback in a generalized way.
- Regarding learning, there appears to be more work on teacher–pupil interaction than on the effects of the peer group. Would this be helpful to consider in the future? We see many ways to add that to AR and VR simulations (This was followed by a discussion of Oculus Social Alpha, the peer-couchsurfing app).
- Authoring of applications for smart glasses, in particular for training, is a big issue that comes with its own problems.
- Measuring user interactions.
- Social acceptance: Will people want their use of mental/skill augmentation to be private? That is, will it be embarrassing to be seen using these augmentations? Compare to hearing aids, where only the most expensive ones are currently unobtrusive. Is it going to be scary that some people have skill augmentation? That is, will others feel threatened by losing out on a perceived (or very real) advantage? The core of this issue is the feeling of access to the technology. If someone feels they can have access to the augmentation and chooses not to use it, it is not a social issue anymore. For example, a photographer using a film camera has chosen not to use the functionality provided by a digital camera, and is unlikely to feel threatened or minimized by a photographer using a digital camera, as s/he could choose to use one if desired. (Indeed, the film photographer might feel superior to the digital photographer.) What if a company were to make and distribute for free to anyone interested a state-of-the-art wearable, in return for access to its data? Would this be any more privacy-compromising than providing state-of-the-art search facilities?
- We need a more sophisticated and nuanced understanding of the task so that there can be a more informed decision-making process about the type of guidance (3D, 2D, audio, tactile, etc.) that is most effective for the user and situation. This decision-making process is currently arbitrary at all levels: which hardware to use, the data-processing strategy, and the format of the augmentation. A better understanding of what is most suitable can lead to less expensive and more widely acceptable eyewear.

Questions

- What is the most effective tool for skill augmentation and task guidance?
- Do the same principles that can be thought of as relevant for training and supporting physical skills also hold for mental/cognitive skills?

Low-Hanging Fruit Applications

Which applications would seem to clearly benefit from being included with eyewear computing for augmenting skills based on existing applications/frameworks/tools? Some could be prototyped as undergrad \leq 4-week projects, if given access to existing tools, but could be much harder to do well.

- Checklists for to-do tasks
- Shopping lists
- Live translation (especially with a low-cognitive load format: <http://spritzinc.com/>, teleprompter)
- Context-sensitive calendar view
- Remembrance agent. Vannevar Bush [3] proposed a stereoscopic head-worn camera: “As the scientist of the future moves about the laboratory or the field, every time he looks at something worthy of the record, he trips the shutter and in it goes, without even an audible click.” This could be augmented with a display and search capabilities. Rhodes [9] describes an early text-based implementation.
- Augmenting Conversations Using Dual-Purpose Speech [7, 1]

References

- 1 Ackerman JM, Nocera CC, Bargh JA. *Incidental haptic sensations influence social judgments and decisions*, *Science*, 328(5986):1712–1715, 2010.
- 2 Page Anderson, Elana Zimand, Larry F. Hodges, and Barbara O. Rothbaum. Cognitive behavioral therapy for public-speaking anxiety using virtual reality for exposure. *Depression and Anxiety*, 22(3), 156–158, 2005.
- 3 V Bush, *As We may Think*. The Atlantic Magazine. July 1945. <http://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/>
- 4 Jonathan Gratch, Gale M. Lucas, Aisha Aisha King, and Louis-Philippe Morency. It’s only a computer: the impact of human-agent interaction in clinical interviews. *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS’14)*. 85–92, 2014.
- 5 Steven Henderson and Steven Feiner, Augmented reality in the psychomotor phase of a procedural task, *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, Basel, Switzerland, 191–200. 2011. <http://dx.doi.org/10.1109/ISMAR.2011.6092386>
- 6 Weiquan Lu, Henry B.-L. Duh, Steven Feiner, and Qi Zhao. Attributes of subtle cues for facilitating visual search in augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 20(3), 404–412. March 2014. <http://dx.doi.org/10.1109/TVCG.2013.241>
- 7 Kent Lyons, Christopher Skeels, Thad Starner, Cornelis M. Snoeck, Benjamin A. Wong, and Daniel Ashbrook. *Augmenting conversations using dual-purpose speech*. Proceedings of the 17th annual ACM symposium on User Interface Software and Technology (UIST), pp. 237–246, 2004.
- 8 S. Mann and J. Fung. EyeTap devices for augmented, deliberately diminished, or otherwise altered visual perception of rigid planar patches of real-world scenes, *Presence*, 11(2):158–175, April 2002. <http://dx.doi.org/10.1162/1054746021470603>

- 9 Bradley J. Rhodes. The wearable remembrance agent: A system for augmented memory. *Personal Technologies*, 1(4), 218–224. December 1997.
- 10 M. Slater, D.P. Pertaub and A. Steed, Public speaking in virtual reality: facing an audience of avatars. *IEEE Computer Graphics and Applications*, 19(2), 6–9, March/April 1999. <http://dx.doi.org/10.1109/38.749116>
- 11 Penny J. Standen and David J. Brown. Virtual reality in the rehabilitation of people with intellectual disabilities: Review. *CyberPsychology & Behavior*, 8(3), 272–282, June 2005. <http://dx.doi.org/10.1089/cpb.2005.8.272>
- 12 Eduardo E. Veas, Erick Mendez, Steven K. Feiner, and Dieter Schmalstieg. Directing attention and influencing memory with visual saliency modulation. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'11)*. 1471–1480. 2011. <http://dx.doi.org/10.1145/1978942.1979158>
- 13 VirtualSpeech application. <http://virtualspeech.co.uk/2016>
- 14 Yan Xu, Joshua Ratcliff, James Scovell, Gheric Speiginer, and Ronald Azuma. Real-time Guidance Camera Interface to Enhance Photo Aesthetic Quality. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI'15)*. ACM, New York, NY, USA, pp. 1183–1186. <http://dx.doi.org/10.1145/2702123.2702418>

6.5 Group 5: EyeWear Computing for Gaming

Thad Starner (Georgia Institute of Technology – Atlanta, US)

License © Creative Commons BY 3.0 Unported license
© Thad Starner

Joint work of Sabrina Hoppe, Masahiko Inami, Moritz Kassner, Päivi Majaranta, Will Patera, Thad Starner, Julian Steil, Yusuke Sugano

Gaming seems an ideal platform for demonstrating and experimenting with EyeWear. Short, snack-style games can be used as probes for new interaction techniques, while longer form games might be used to create naturalistic and controlled scenarios for basic gaze research. Access to eye, face, and head motion enables new gaming mechanisms. For example, an advanced form of the popular game “Fruit Ninja” might require players to use relative eye motion to select a target from a group of distractors. Horror games may use knowledge about the user’s area of focus to guarantee that a new monster is rendered in the user’s peripheral field of view, increasing the level of surprise and startlement. Similarly, timing changes in the interface to correspond with eye blinks and saccadic blindness might enable a higher or lower level of stress in the game depending on the maker’s intention. Staring and squinting could be coupled with mechanisms of selection or zooming, and games might require the user make certain facial expressions to encourage certain moods in the game (e.g., snarling at an opponent or smiling at a dog that the player is trying to befriend). Using EyeWear to support gamification of everyday activities, like what Fitbit does for walking, might encourage healthy or desired behaviors. Gamification approaches can be also useful to collect large-scale data required for other related research areas. Everyday-use EyeWear games could also be used for studies, skill creation and rehearsal, persuasive interfaces, physical therapy, physical conditioning, meditation, or just relaxation. In short, EyeWear games seems a promising and enjoyable approach for rapid iteration of eye and head interfaces that might be adopted later to other applications.

6.6 Group 6: Prototyping of AR Applications using VR Technology

Scott W. Greenwald (MIT Media Lab – Cambridge, MA, USA)

License  Creative Commons BY 3.0 Unported license
© Scott W. Greenwald

Joint work of Rita Cucchiara, Ozan Cakmakci, Thies Pfeiffer

We discussed a methodology for iteratively designing and building eyewear applications using a sequence of mixed-reality prototype systems that incrementally approach the final user experience. The methodology allows researchers to derive and benefit from user-centered design insights without needing a complete prototype. Although it is already common to use paper-prototyping and “wizard of oz” techniques to pilot candidate designs, we are observing that in many cases there should be more intermediate steps between the paper prototype and the system based on the final hardware configuration. That is, many augmented reality user interactions and system affordances can be simulated with varying fidelities that are useful but cheaper or more rapid to implement. One example would be simulating an optical see-through eyewear system using a video see-through system. We also observe that small prototyping steps can be made within each iteration of the system. For example, in a system where the world is simulated using a “cave” surround projection system, one can move from simulating the use of a mobile device using the projection system to using a physical mobile device before moving completely out of the cave into the physical world. Furthermore, the intermediate design steps put bounds on hardware design requirements (such as field of view on a head-worn display). This approach aims at minimizing unnecessary hardware builds, which accelerates the hardware design cycle while reducing surprises at the end. A concrete outcome from this working group is to expand these ideas into a position paper on this subject.

7 Community Support

Another reoccurring topic discussed at the seminar was community support, i.e. how can we share tools, datasets and practices used in the different research communities present at the seminar. The following list of datasets and tools is the result of these discussions.

7.1 Datasets

- **Bristol egocentric object interactions.** 6 activities, 5 people, mobile eye tracker and egocentric camera. [2]
- **CAMSynthesEyes.** The dataset contains 11,382 synthesized close-up images of eyes. [3]
- **Databrary.** Databrary is a video data library for developmental science. [4]
- **MPI long-term visual behaviour.** Over 80 hours of visual behaviour data in everyday settings. [5]
- **Georgia tech egocentric datasets.** List of datasets on egocentric vision. [6]
- **Unimore Egocentric Vision.** Egocentric vision data sets focusing on social relationships. [8]
- **Swirski Dataset.** Pupil detection data set. [1]
- **Labelled Pupils in the Wild (LPW).** Pupil detection data set. [7]

References

- 1 <https://www.cl.cam.ac.uk/research/rainbow/projects/pupiltracking/datasets/>
- 2 Bristol Egocentric Object Interactions. <https://www.cs.bris.ac.uk/~damen/BEOID/>
- 3 AM SynthesEyes. <https://www.cl.cam.ac.uk/research/rainbow/projects/syntheseeyes/>
- 4 Databrary <https://nyu.databrary.org>
- 5 Discovery of Everyday Human Activities From Long-term Visual Behaviour. <https://www.mpi-inf.mpg.de/departments/computer-vision-and-multimodal-computing/research/human-activity-recognition/discovery-of-everyday-human-activities-from-long-term-visual-behaviour-using-topic-models/>
- 6 Georgia Tech Egocentric Vision <http://cbi.gatech.edu/egocentric/datasets.htm>
- 7 MPII Labelled pupils in the wild (LPW) <http://mpii.de/LPW>
- 8 UNIMORE Egocentric vision for social relationship <http://imabelab.ing.unimore.it/imabelab/researchactivity.asp?idAttivita=23>

7.2 Tools

In the following, we enumerate some of the tools used by the communities present at the seminar.

- **WearScript.** Can be used to capture sensor data from Android, including camera frames (useful for Glass) <http://www.wearscript.com/>
- **Pupil head-mounted eye tracking.** <https://pupil-labs.com/pupil>
- **J!NS MEME Logger.** Records data from developer version of J!NS MEME on iOS. <https://github.com/shoya140/MEMELogger-iOS-developers>
- **Google Glass Logger.** Records all sensor data from glass. <https://github.com/shoya140/GlassLogger>
- Software to render ground truth annotated eye images for pupil detection and tracking / gaze estimation. <https://www.cl.cam.ac.uk/research/rainbow/projects/eyerender/>
- Caffe code for unsupervised feature learning with unlabeled video accompanied by egomotion sensor data. <http://www.cs.utexas.edu/~dineshj/projects/4-egoEquiv/>
- **TraQuMe.** Tool for checking the quality of tracking – this tool supports several trackers via COGAIN EtuDriver. <http://www.uta.fi/sis/tauchi/virg/traqume.html>
- **Snap point detection code.** http://vision.cs.utexas.edu/projects/ego_snappoints/
- Logging multimodal information for cognitive psychology – collecting applied force to the fingerpad. <http://www.tecgihan.co.jp/english/p2.htm>
- **Binaural microphone for collecting soundscapes.** Roland CS-10EM In-Ear Monitors. <http://www.amazon.co.jp/dp/B003QGPCTE>
- **iMotions Biometric Research Platform** Record physiological signals synchronized with stimuli and videos of subjects. <http://imotions.com/>
- **Rapid Gesture Recognition Toolkit.** A Unix command line utility for doing quick evaluations of machine learning algorithms for gesture recognition. <http://gt2k.cc.gatech.edu/>

Participants

- Andreas Bulling
Max-Planck-Institut für
Informatik – Saarbrücken, DE
- Ozan Cakmakci
Google Inc. –
Mountain View, US
- Rita Cucchiara
University of Modena, IT
- Steven K. Feiner
Columbia University, US
- Kristen Grauman
University of Texas – Austin, US
- Scott Greenwald
MIT – Cambridge, US
- Sabrina Hoppe
Max-Planck-Institut für
Informatik – Saarbrücken, DE
- Masahiko Inami
Keio University – Yokohama, JP
- Shoya Ishimaru
Osaka Prefecture University, JP
- Moritz Kassner
Pupil Labs – Berlin, DE
- Koichi Kise
Osaka Prefecture University, JP
- Kiyoshi Kiyokawa
Osaka University – Osaka, JP
- Kai Kunze
Keio University – Yokohama, JP
- Yin Li
Georgia Institute of Technology –
Atlanta, US
- Paul Lukowicz
DFKI – Kaiserslautern, DE
- Päivi Majaranta
University of Tampere, FI
- Walterio W. Mayol-Cuevas
University of Bristol, GB
- René Mayrhofer
Universität Linz, AT
- Masashi Nakatani
University of Tokyo, JP
- Will Patera
Pupil Labs – Berlin, DE
- Thies Pfeiffer
Universität Bielefeld, DE
- James M. Rehg
Georgia Institute of Technology –
Atlanta, US
- Philipp M. Scholl
Universität Freiburg, DE
- Linda B. Smith
Indiana University –
Bloomington, US
- Gábor Sörös
ETH Zurich, CH
- Thad Starner
Georgia Institute of Technology –
Atlanta, US
- Julian Steil
Max-Planck-Institut für
Informatik – Saarbrücken, DE
- Yusuke Sugano
Max-Planck-Institut für
Informatik – Saarbrücken, DE
- Yuji Uema
J!NS – Tokyo, JP

