Report from Dagstuhl Seminar 18252

Ubiquitous Gaze Sensing and Interaction

Edited by

Lewis Chuang[1], Andrew Duchowski[2], Pernilla Qvarfordt[3], and Daniel Weiskopf[4]

1 **LMU München, DE,** lewis.chuang@um.ifi.lmu.de
2 **Clemson University, US,** duchowski@clemson.edu
3 **FX Palo Alto Laboratory, US,** pernilla.qvarfordt@gmail.com
4 **Universität Stuttgart, DE,** daniel.weiskopf@visus.uni-stuttgart.de

Abstract

This report documents the program and outcomes of the three day Dagstuhl Seminar 18252 "Ubiquitous Gaze Sensing and Interaction". The miniaturization of optical devices and advances in computer vision, as well as a lower cost point, have led to an increased integration of gaze sensing capabilities in computing systems. Eye tracking is no longer restricted to a well controlled laboratory setting, but moving into everyday settings. Therefore, this Dagstuhl Seminar brought together experts in computer graphics, signal processing, visualization, human-computer interaction, data analytics, pattern analysis and classification along with researchers who employ eye tracking across a diverse set of disciplines: geo-information systems, medicine, aviation, psychology, and neuroscience, to explore future applications and to identify requirements for reliable gaze sensing technology. This fostered a dialog and allowed: (1) computing scientists to understand the problems that are faced in recording and interpreting gaze data; (2) gaze researchers to consider how modern computing techniques could potentially advance their research. Other issues concerning the ubiquitous deployment of gaze sensing and interaction were also discussed, such ethical and privacy concerns when deploying gaze monitoring devices in everyday settings.

Seminar June 18–21, 2018 – http://www.dagstuhl.de/18252

2012 ACM Subject Classification Human-centered computing → Ubiquitous and mobile computing, Human-centered computing → User models, Human-centered computing → Visualization, Applied computing → Law, social and behavioral sciences, Computing methodologies → Computer vision, Theory of computation → Pattern matching

Keywords and phrases eye tracking, computer vision, pattern analysis, ubiquitous computing, user modeling

Digital Object Identifier 10.4230/DagRep.8.6.77

Edited in cooperation with Tanja Blascheck

Ubiquitous Gaze Sensing and Interaction, *Dagstuhl Reports*, Vol. 8, Issue 06, pp. 77–148
Editors: Lewis Chuang, Andrew Duchowski, Pernilla Qvarfordt, and Daniel Weiskopf
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Executive Summary

Tanja Blascheck (INRIA Saclay, FR)
Lewis Chuang (LMU München, DE)
Andrew Duchowski (Clemson University, US)
Pernilla Qvarfordt (FX Palo Alto Laboratory, US)
Daniel Weiskopf (Universität Stuttgart, DE)

The miniaturization of optical devices and advances in computer vision, as well as a lower cost point, have led to an increased integration of gaze sensing capabilities in computing systems, from desktop computing to mobile devices and wearables. With these advances in technology, new application areas for gaze sensing are emerging. Eye tracking is no longer restricted to a well-controlled laboratory setting, but moving into everyday settings. When technology makes forays into new environments, there are many questions to be resolved and challenges to be met, from computational to applications and interaction. Ubiquitous gaze sensing and interaction require a framework that can accommodate compatible solutions from data acquisition to signal processing to pattern classification and computer vision to visualization and analytics. Including gaze data into interactive applications requires knowledge of natural gaze behaviors as well as how gaze is coordinate with other modalities and actions.

Therefore, this Dagstuhl Seminar brought together computer scientists and gaze researchers to explore future ubiquitous applications and to identify requirements for reliable gaze sensing technology. Ubiquitous gaze sensing and interaction cannot be achieved by research discipline, but require knowledge and scientific advancement in multiple fields. And, of utmost importance is that researchers from different disciplines meet, interact, and address their common challenges. For this reason, experts in computer graphics, signal processing, visualization, human-computer interaction, data analytics, pattern analysis and classification along with researchers who employ gaze tracking across diverse disciplines attended: geo-information systems, medicine, aviation, psychology, neuroscience, etc.This fostered a dialogue and allowed: (1) computing scientists to understand the problems that are faced in recording and interpreting gaze data, (2) gaze researchers to consider how modern computing techniques could potentially advance their research. In addition, we discussed the ethical and privacy concerns of deploying gaze monitoring devices in everyday scenarios.

The workshop was organized to identify identifying possible **scenarios** and pinpointing the associated **challenges** of developing and deploying ubiquitous gaze sensing during the first day. Challenges identified by multiple scenarios, or the ones that were considered to be significant were the focus of in-depth cross-disciplinary groups. These challenges were discussed on the second day. In three sessions taking place during the day, five challenges were debated. "Data Privacy" and "Gaze + X" were two of the most important topics and received multiple dedicated sessions of discussion due to the high interest of the participants.

On the third day the Dagstuhl Seminar finally discussed future work and how to get the research community engaged in researching the various interesting topics covered. Some of the suggestions were to organize workshops at conferences and organizing a special issue focused on ubiquitous gaze sensing. Several of the discussion groups started brainstorming on papers covering the important topics raised at the workshop.

2 Table of Contents

3 Scenarios

All scenarios were grounded on the assumption that gaze sensing technology (e.g., eye tracking) were available and working reliably everywhere. The workshop participants brainstormed a large number of scenarios where ubiquitous gaze sensing could be used for the study of human-behavior or to enhance and enrich interaction with computing systems. The participants pitched scenarios to each other in a speed-dating pitch. The different scenarios were consolidated and voted on to extract scenarios that well exemplified ubiquitous gaze sensing in action. In the end, the workshop attendees selected four scenarios: Going Places, Healthcare, Work & Play, and Everyday Use of Wearable Gaze Trackers, to flesh out opportunities and challenges for realizing the scenarios.

Each scenario was discussed in smaller groups, with a focus on describing the scenario in a future setting, identifying relevant research questions, and determining assumptions on technology advancement within the scenarios.

3.1 Everyday Use of Wearable Gaze Trackers

Andreas Bulling (MPI für Informatik – Saarbrücken, DE), Lewis Chuang (LMU München, DE), Kenneth Holmqvist (Universität Regensburg, DE), Radu Jianu (City – University of London, GB), David P. Luebke (NVIDIA – Charlottesville, US), Diako Mardanbegi (Lancaster University, GB), Thies Pfeiffer (Universität Bielefeld, DE), and Laura Trutoiu (Magic Leap – Seattle, US)

Motivation: Wearable gaze tracking is on the verge of being adopted by the average consumer, but to be fully adopted in everyday use it needs to enables unique and desired services. When working on this scenario, we discussed: (1) what such services might be, (2) how such services could be implemented with a mobile computing system that has access to the user's gaze, and (3) the technical requirements of such an envisioned system.

Scenarios: We envisioned a daily scenario from waking up to going to bed whereby gaze tracking could facilitate personalized computing services.

- *Gaze interactive media*: the user might be presented with news stories on a smart mirror that scrolls according to the user's gaze.
- *Breakfast preparation*: the user would be prompted if it is noted that the user has missed an ingredient.
- *Search suggestions*: when a user displays visual search behavior prior to leaving the house, the system could suggest potential search locations.
- *Public displays*: when gazing at a remote public display (e.g., arrival times of the bus), the relevant information could be delivered directly to one's personal display.
- *Shopping*: when in a shopping mall, a personal computing system could identify what one is interested in buying to make store recommendations.
- *Task scheduling*: a personal computing system could prompt simple tasks upon noticing, from the user's gaze, that the user is available to perform them, such as checking emails while waiting in a queue.
- *Adaptive environment*: ambient lighting, e.g., blinds could be adjusted in accordance to pupil dilation.

- *Social interactions*: gaze paired with face recognition could assist the user in recognizing a familiar acquaintance as well as provide additional information, e.g., name of spouse.
- *Journaling*: a recollection of one's daily events and interactions could be presented at the end of the day to trigger the user's memories during journaling.

Research Questions: What are the everyday functions that wearable gaze tracking could serve? How is gaze sensing and analysis location- and context-dependent? How do we integrate analytics from gaze sensing with other services and computing systems? What implications for does ubiquitous gaze sensing have for data privacy and security?

Assumptions: To realize the scenarios, we identified the following challenges:

- *Computer vision*: context-dependent applications will depend heavily on computer vision, i.e., object and scene recognition.
- *Reliability*: a personal computing system will have to be cognizant of the precision of current gaze estimates, given environmental luminance and other related factors, prior to making a recommendation.
- *User acceptance*: there will be concerns related to privacy, utility, as well as form factors.
- *Form factors*: gaze tracking should be lightweight, non-obtrusive, and does not obscure field-of-view.
- *Power consumption*: the device should not require more than one charge per day.
- *Multi-modal interaction*: gaze input should be coupled with other inputs to ensure robust inferences.
- *Centralized/distributed computing*: there will be a need for a computing infrastructure that allows for secure interaction between one's personal computing device and others.

3.2 Going Places

Amy Alberts (Tableau Software – Seattle, US), Hans-Joachim Bieg (Robert Bosch GmbH – Stuttgart, DE), Tanja Blascheck (INRIA Saclay, FR), Sara Irina Fabrikant (Universität Zürich, CH), Enkelejda Kasneci (Universität Tübingen, DE), Peter Kiefer (ETH Zürich, CH), Michael Raschke (Blickshift GmbH – Stuttgart, DE), Martin Raubal (ETH Zürich, CH), and Daniel Weiskopf (Universität Stuttgart, DE)

Motivation: Ubiquitous gaze sensing and interaction will have a major impact on future mobility. Eye tracking devices will enable pedestrians, cyclists, car drivers, etc.to enhance their skills through training, for localization, or performance improvement, e.g., based on where a person is looking additional information could be depicted. In addition, collected data from a crowd of people can help shape future cities by integrating gaze information while planning urban projects.

Scenarios: The following scenarios were discussed to illustrate ubiquitous gaze sensing for going places:

- Usage of ubiquitous gaze sensing when the car is the main means of transportation: training, spatial cognition, self localization / memory, performance improvement (e.g., Formula 1, Uber, taxis).
- Usage of ubiquitous gaze sensing to ensure or enhance safety, e.g., monitoring/vigilance (sleeping driver), health, advertisement, elderly.

- Usage of ubiquitous gaze sensing for urban planning, e.g., diagnostics, managing traffic.
- Usage of ubiquitous gaze sensing with autonomous cars: e.g., using gaze as interaction; looking outside and the car knows what you are looking at (restaurant).

Research Questions: How can eye tracking assist traffic participants (e.g., pedestrians, cyclists, car drivers) in the future?

Assumptions: The following assumptions are made that have to be fulfilled for this scenario:

- Robust gaze tracking in the car and while cycling.
- Outdoor conditions do not cause problems (e.g., sunlight, glasses, calibration).
- Problem-free integration of many different sensors (e.g., GSR, EEG, head orientation, vehicle sensors).
- Adaptable to multiple environments (e.g., urban, city, highway, forest).

3.3 Healthcare

M. Stella Atkins (Simon Fraser University – Burnaby, CA), Roman Bednarik (University of Eastern Finland – Joensuu, FI), Leslie Blaha (Pacific Northwest National Lab. – Richland, US), Nina Gehrer (Universität Tübingen, DE), and Eakta Jain (University of Florida – Gainesville, US)

Motivation: Ubiquitous gaze sensing and interaction have the potential to transform medical practices in a number of ways. Gaze patterns are known to change depending on physical and mental conditions, hence gaze sensing can provide diagnostic information not available to health professionals today. Beyond diagnostics, healthcare professionals are engaged in a number of different tasks where gaze plays an important role. The medical setting, however, is quite unique so applications need to be specially targeted to be successful.

Scenarios: The following scenarios were discussed to illustrate this scenario:

- Passive monitoring of patients enables continuous monitoring and longitudinal data for diagnostics of health status and evaluations of treatment efficacy. Analysis pushed to the sensors provides continuous analysis, not just continuous data collection. Personalized analytics might enable feedback directly to the patients. This could be done in healthcare establishments (e.g., hospitals, nursing facilities) as well as home and work environments.
- Virtual doctors with realistic and expressive gaze behaviors will be available for mental health evaluation and therapy sessions. Generating high-fidelity simulated behavior is important for garnering patient trust and providing effective feedback. Avatars may be customizable to specific populations by providing appropriate affective and conversational cues.
- Ubiquitous gaze sensing of doctors and health providers provides continuous monitoring of performance. This can be used as a data stream for decision support systems. It provides a record of a provider's observations which can be leveraged for second opinions and record keeping. Expert behaviors can be captured and leveraged in case evaluation and in teaching of other providers.
- Teams of healthcare providers are provided new awareness of each other's activities through gaze sensing. For highly coordinated situations, like emergency triage or surgery, gaze data provides information about more of the situation to providers who

need to coordinate care. Gaze-based interactions provide another method of inputting information to a provider system or record of notes, allowing providers to keep their hands on the patients.
- Virtual health collaboration, or leveraging of mixed reality, will become a possibility. Remote expertise might be brought in to assist. information from the primary surgeon's gaze can be sent to the remote expert. Gaze-based interactions for the remote expert can control the view or cameras, providing needed information.

Research Questions: How will ubiquitous gaze sensing and interaction play into future medical domain applications? Within the breakout group, we discussed the medical domain from three different perspectives:

1. How will ubiquitous gaze sensing impact patients?
2. How will ubiquitous gaze sensing impact care givers?
3. How will ubiquitous gaze sensing change medical care practices or training?

Assumptions: The following assumptions are made that have to be fulfilled to support pervasive future healthcare applications:

- We will have an established legal framework that addresses privacy, especially compliance with medical privacy regulations (e.g., the Directive on Data Protection in the European Union or the Health Information Portability and Accessibility Act (HIPAA) in the USA).
- We have an established the diagnosticity of gaze data for intended medical applications.
- We have established models for gaze metrics related to diagnostic tests and treatments.
- Diagnostics are robust independent of data collection methods (e.g., wearable eye tracker, desktop cameras, cameras integrated into environment or toys/objects).
- When virtual (avatar) healthcare providers are involved, they behave in a believable and trustworthy manner (e.g., make realistic eye movements).
- Settings and variable conditions (e.g., changes in lighting, glasses, calibration) do not include gaze sensing performance.
- Gaze sensing methods are adaptable to different age ranges and health conditions (e.g., children, adults, elderly; mobile and bed-ridden patients).
- Technology is available for home use and clinical use.
- Technology is user-friendly, requiring minimal setup and maintenance from patients, and no calibration.
- Bandwidth, battery power, and data storage issues are solved.
- The speed of analytics is fast enough to move beyond gaze position inferences.
- Multi-scale, multi-resolution eye tracking is possible, and adaptable according to diagnostics needed.

3.4 Playing & Learning

Maria Bielikova (STU – Bratislava, SK), Andrew Duchowski (Clemson University, US), Hans Gellersen (Lancaster University, GB), Krzysztof Krejtz (SWPS University of Social Sciences and Humanities, PL), Kuno Kurzhals (Universität Stuttgart, DE), Radoslaw Mantiuk (West Pomeranian Univ. of Technology – Szczecin, PL), and Pernilla Qvarfordt (FX Palo Alto Laboratory, US)

Motivation: Edutainment is an interesting setting in that it allows participants to learn while playing an engaging game. Today these games are limited how they can model users' understanding and level of learning. Gaze has potential in revealing both users' attention and cognitive processes that can be used to improve models of understanding and learnings.

Scenarios: Suppose we have a multi-party game, e.g., playing a problem-solving gaze, where some participants may be playing from a remote location while other play together in the same room. The game could be projected on a shared surface, or represented in VR or traditional displays. The students' gaze is tracked to help communicate with other students or with remote teachers. The system monitors comprehension, gives advice if needed, or calls on the teachers' attention to help the students when they are stuck. The system can detect fatigue, intellectual helplessness, confusion, and tasks adjusted to educational level. It could provide teacher and students with replay with analytics of the learning session so that they can review, discuss, and learn how to improve their performance.

Research Questions: How can gaze be used in multi-party scenarios such as (VR) gaming and/or education? When the game is aware of everyone's gaze, how can this be exploited for the benefit of the players in terms of entertainment and learning? When a player's gaze is monitored and visualized, in real-time or as a kind of brief historical scanpath, how could other players or a remote teacher make use of this? How can we model learning from gaze and other modalities?

Assumptions: The following assumptions are made that have to be fulfilled for this scenario:
- Recognition and modeling of student/player cognitive state via gaze, actions and visual context, e.g., real-time analysis of comprehension, is solved.
- Real-time detection of mindless gaze as an indication of cognitive fatigue.
- Gaze visualization of multiple people for interpersonal communication.
- Social presence by gaze, may need to learn "gaze language".
- Joint attention is easy to represent/visualize.
- Gaze as additional channel of information is understood.
- Ethical issues non-existent.
- Skill assessment via gaze and eye-hand coordination is understood.
- Learning disability detection (autism, ADHD) is doable.
- Detection of cheating is doable.

4 Challenges

Based on the assumptions identified in the scenarios, the workshop set forth to find challenges that cross multiple scenarios. These challenges were selected for the next set of discussions.

4.1 Gaze + X

Amy Alberts (Tableau Software – Seattle, US), M. Stella Atkins (Simon Fraser University – Burnaby, CA), Hans-Joachim Bieg (Robert Bosch GmbH – Stuttgart, DE), Leslie Blaha (Pacific Northwest National Lab. – Richland, US), Lewis Chuang (LMU München, DE), Andrew Duchowski (Clemson University, US), Nina Gehrer (Universität Tübingen, DE), Hans Gellersen (Lancaster University, GB), Kenneth Holmqvist (Universität Regensburg, DE), Eakta Jain (University of Florida – Gainesville, US), Radu Jianu (City – University of London, GB), Krzysztof Krejtz (SWPS University of Social Sciences and Humanities, PL), David P. Luebke (NVIDIA – Charlottesville, US), Radoslaw Mantiuk (West Pomeranian Univ. of Technology – Szczecin, PL), Thies Pfeiffer (Universität Bielefeld, DE), Pernilla Qvarfordt (FX Palo Alto Laboratory, US), Martin Raubal (ETH Zürich, CH), and Laura Trutoiu (Magic Leap – Seattle, US)

Research Question: If we combine gaze with other information, i.e., about context, user's reaction, actions, and tasks, as well as other sensor information, what extra power in the interpretation of the "gaze + X" do we get?

Description: What is the "X"? In our discussion, "X" could be other signals collected in connection with the gaze, such as pupil size, or by other sensors that detect users action, bio-signals, etc.(gaze + sensors). Stepping away from the user, X could also be context defined by the user's attention. We could, for instance, detect objects, actions, sounds, and speech in the environment and analyze in relation with the gaze (gaze+context). When applying analysis of gaze, it can transform for being an indication of attention at one point in time, to show a fluid behavior. One such example was how a "glance" can be interpreted. Current eye movement classification schemes do not do this very well, yet this information could improve our understanding of user's understanding and interpretation. Utilizing multiple analytics from the same gaze sensing device could hence been seen as another "X" (gaze + machine learning). When achieving a more complete understanding of gaze in relation to other information sources, we can develop a framework for design interactive system or for creating improved models of human cognition.

4.2 Intent and Prediction

Amy Alberts (Tableau Software – Seattle, US), M. Stella Atkins (Simon Fraser University – Burnaby, CA), Roman Bednarik (University of Eastern Finland – Joensuu, FI), Andreas Bulling (MPI für Informatik – Saarbrücken, DE), Andrew Duchowski (Clemson University, US), Sara Irina Fabrikant (Universität Zürich, CH), Nina Gehrer (Universität Tübingen, DE), Eakta Jain (University of Florida – Gainesville, US), Peter Kiefer (ETH Zürich, CH), and Daniel Weiskopf (Universität Stuttgart, DE)

Research Question: The related questions of intent and prediction explored whether gaze could be used to predict a user's action. A particularly interesting qualification to these questions was that of temporal scale. That is, how far into the future could we predict a person's intent, if at all?

Description: As an example of a fairly straightforward proof-of-concept in this scenario was the extreme short-term prediction of saccade landing position, which has already been demonstrated to a certain extent. The greater challenge is in the longer timeframe: could we predict the user's intent on the order of seconds, minutes, hours, or even days? Doing so would require collecting gaze for longer historical periods and clever algorithms for divining intent-based on observed gaze.

4.3 Novel Interaction Paradigms

Amy Alberts (Tableau Software – Seattle, US), Hans Gellersen (Lancaster University, GB), Kenneth Holmqvist (Universität Regensburg, DE), Krzysztof Krejtz (SWPS University of Social Sciences and Humanities, PL), David P. Luebke (NVIDIA – Charlottesville, US), Diako Mardanbegi (Lancaster University, GB), and Laura Trutoiu (Magic Leap – Seattle, US)

Research Question: For novel interaction paradigms, the interesting question is how do we move beyond basic gaze-based selection and possibly the use of gaze gestures?

Description: To a large extent, coming up with novel interaction paradigms depends on the contextual of gaze, e.g., is it in AR or VR, is it looking at a display, or rather is it in the ubiquitous sense where objects have the "power" of detecting gaze (i.e., at them). The latter was a particularly interesting concept termed the Internet of Seeing Things, or IOST. Other scenarios tended to consider head-mounted tracking as in VR or perhaps AR contexts, e.g., how can we use gaze directed at other individuals? Could we also mix in the concept of "Gaze + X" here, as in, when looking at another individual, could gaze direct face recognition modules to identify the other person and then trigger contextual information such as their name, birthday, and other related pieces of information (how many children do they have, if any), etc.

4.4 Data Privacy

Roman Bednarik (University of Eastern Finland – Joensuu, FI), Maria Bielikova (STU – Bratislava, SK), Tanja Blascheck (INRIA Saclay, FR), Andreas Bulling (MPI für Informatik – Saarbrücken, DE), Sara Irina Fabrikant (Universität Zürich, CH), Eakta Jain (University of Florida – Gainesville, US), Peter Kiefer (ETH Zürich, CH), Kuno Kurzhals (Universität Stuttgart, DE), David P. Luebke (NVIDIA – Charlottesville, US), Diako Mardanbegi (Lancaster University, GB), Michael Raschke (Blickshift GmbH – Stuttgart, DE), and Daniel Weiskopf (Universität Stuttgart, DE)

Research Question: How can we ensure privacy once gaze sensing becomes pervasive?

Description: If people wear gaze sensing technology, we have to ensure their privacy and the privacy of others. Privacy is critical because gaze data can reveal much and highly personal information about the person being tracked, including information about personality and potential medical issues. How can user models obtained from analyzing such data be protected? In addition, we have to educate people how to control the privacy of their gaze data, understand the implications of different levels of privacy protection, and make sure that the underlying models do not have negative implications such as preventing us from looking someplace (e.g., 'don't look there'). Furthermore, privacy issues are not restricted to the person wearing a pervasive gaze-sensing device but may include the person's environment, in particular, other people with whom we are interacting and who might be recorded by the sensing device.

4.5 Ubiquitous Gaze-based Guidance and Recommendation Systems

Maria Bielikova (STU – Bratislava, SK), Leslie Blaha (Pacific Northwest National Lab. – Richland, US), Tanja Blascheck (INRIA Saclay, FR), Radu Jianu (City – University of London, GB), Kuno Kurzhals (Universität Stuttgart, DE), Thies Pfeiffer (Universität Bielefeld, DE), Michael Raschke (Blickshift GmbH – Stuttgart, DE), and Martin Raubal (ETH Zürich, CH)

Research Question: How can gaze be used for ubiquitous gaze-based guidance and recommendation systems?

Description: Gaze-based recommendation system use data collected from eye trackers to make recommendations to people. For example, by detecting a persons familiarity or expertise with a new tool or device they can be aided when performing a task. In addition, gaze can be used to detect engagement, activity changes, or context-switching to give the next input while performing a task. These examples also require proper guidance of a person, for example, gaze-guided storytelling techniques. For this, a taxonomy of different scenarios, methods, drawbacks, and benefits as well as the creation of a design space for ubiquitous gaze-based guidance and recommendation systems is required.

5 Open problems

5.1 Toward a Ubiquitous Gaze-based Interaction Model

Amy Alberts (Tableau Software – Seattle, US)

In 2002 a tech article in the New York Times touted the innovation of a small company named FingerWorks. They imagined pieces of glass that could display a keyboard and be controlled by a fingertip. They promised you could "spill your coffee on it" and they keyboard would still work. FingerWorks was eventually bought by Apple and their TouchStream Interaction Model is how we all interact with touch-enabled devices today. The success of the TouchStream model came from solving many human factors, user interaction, and design problems that came with traditional indirection manipulation devices. We are now on the horizon of a new technological breakthrough, that will address the user and design problems we have with touch-enabled devices. Gazed-based user interaction is imminent. However, like touch in 2002, the base interaction model (e.g., select/dismiss, scroll, etc.) for gaze is not known. This paper explores how we might achieve an equally ubiquitous model for gaze-enabled systems of the future.

The most commercially practical application of a gaze-based user interface will be the delivery of information about items in your visual field. Imagine early versions of this where a customer (Alice) walks into a retailer like Nordstrom's or Whole Foods. Alice is given a pair of glasses to wear when she enters the store. These glasses are computer-vision and gaze-tracking enabled. As she shops around the store, she can see digital indicators of 'more information' about items in her field of view. Alice can visually select an item on a shelf (a loaf of artisan bread). The information she sees tells her the price of the loaf, how long it's been on the shelf, and other infographics about its ingredients and caloric composition. Alice grabs the loaf which dismisses the information she's seeing and she moves onto the next item in her list.

Alice's partner is allergic to tree nuts. Alice has the Whole Foods app on her phone in which she indicated this allergy. They are having friends over for dinner and she wants to get a fresh cake for dessert. She approaches the bakery counter and looks at the different cakes. There's a small indicator drawn over the cakes that she should avoid because those cakes include tree nuts.

Alice is about to check out and she remembers she needs to get some ground coffee. She's unsure of where coffee is in this store, so she asks (out loud) "where's the coffee?" She sees an overlay of arrows that indicate the route to the coffee. These arrows adjust and change as she moves through the store.

Core interaction model questions must be addressed to enable the scenarios described here. This interaction model must address human factors considerations that ensure low impact on the human – especially for high volume gestures (e.g., selection). The definition of scenario categories, variables that will affect the reliability of eye tracking technology, and architecture of the software stack will need exploration.

A systematic approach to the gaze-based interaction model should include (but is not limited to) the following

- Comprehensive literature review of the development, consideration, and limitations of existing interaction models (GUI, Touch, Haptics, etc.).

- Identify and develop methodologies to establish human factors, cognitive, and visual system principles that are relevant to a gaze-based interaction model.
- Identify a set of core gestures that must be supported by the gaze-based interaction model (e.g., select/dismiss, scroll, etc.).
- Build acceptance thresholds for successful gaze interaction gestures.
- Build and test a variety of interaction model options to test against the acceptance thresholds.
- Propose a core set of ubiquitous gaze-based gestures.

5.2 Eyegaze Tracking in Medicine

M. Stella Atkins (Simon Fraser University – Burnaby, CA)

Main reference Benjamin Law, M. Stella Atkins, Arthur E. Kirkpatrick, Alan J. Lomax: "Eye gaze patterns differentiate novice and experts in a virtual laparoscopic surgery training environment", in Proc. of the Eye Tracking Research & Application Symposium, ETRA 2004, San Antonio, Texas, USA, March 22-24, 2004, pp. 41–48, ACM, 2004.
URL http://dx.doi.org/10.1145/968363.968370

Eye trackers are used for medical image perception studying how radiologists make diagnoses in medical images such as CT, MRI, and mammograms. The eye tracking data provides understanding of the visual search process and why errors occur. Eye trackers also are used in surgery, especially for minimally invasive and robot-assisted surgical training where eye-hand coordination is a key factor for good performance. Eye tracking gives insight into how experts differ from novices, and how to improve medical training and monitoring methods. Emerging applications are being developed to integrate eye tracking information towards developing eyegaze-driven decision support systems and to provide gaze contingent control in surgery.

5.2.1 Introduction

Developing expertise in radiology and in other clinical visual tasks such as examining a patient to diagnose skin problems, is an important domain where eye tracking can provide valuable information to suggest methods for training and to form effective decision support systems for medical diagnosis. Eye trackers for medical image perception in radiology were pioneered by Drs. Kundel and Nodine [1, 2], where the ultimate aim was to improve diagnosis and reduce the error rates. A recent review details some history and progress [3], concluding that eye tracking can assist in the assessment of expertise, as well as address human errors in visually-based medical decision-making.

Eye tracking research is often performed in the field of minimally invasive surgery (MIS), as MIS is technically much more demanding than open surgery due to the remote interface of the technique with little tactile feedback [4]. MIS training involves practicing simulated surgery tasks such as grasping and reaching objects, using computer simulators in 3D or a physical training box. Eye tracking has revealed differences in the visual behavior between novices and experts performing the same simulated laparoscopic task [5]; experts kept their gaze on the surgical target whereas novices tracked the tool tip. Such knowledge is key to the understanding of how the motor learning process occurs and it elucidates the role of the human visual system on this process. Training also includes novices watching surgery videos, where eye tracking reveals there is a difference between "watching" and "doing" [6]. Other research addresses gaze during delicate neuro-surgery applications requiring the use of a microscope, to which eye trackers can be attached and used to predict the surgeon's intent [7].

5.2.2 Envisioned Challenges and Solutions

Acquiring quality eyegaze data with experts is a huge challenge, but in 2016, an "Image Perception Lab" was held at the Radiological Society of North America annual conference, whereby the lab invited attendees to volunteer their own time reading data while being eye tracked in an interactive session. This initiative is ongoing, and enables much important data to be collected. Other challenges include developing appropriate models of image perception and intent; at the Dagstuhl seminar it was stimulating to discuss these issues with psychologists and eyegaze practitioners, and consider new psychological models to improve diagnostic performance.

Developing effective visual training for surgery is challenging because of the difficulty of intent prediction. With machine learning we can identify key points in the scene videos for further detailed investigation, and infer cognitive state through eye parameters such as pupil size, and ultimately, how we can use this data to train novices where to look, for improved performance. For gaze contingent control in surgery, we need to take advantage of the surgeons' 3D vision through eye tracking through microscopes or through binocular vision eg of the Da Vinci robot. Camera control is a problem in many surgeries, as it's very difficult to synchronize with the surgeon's movements. Gaze-based camera control is an encouraging approach in robotic surgery [8], which may also be suitable for MIS surgery. As a result of the Dagstuhl seminar, I will be contributing a review section on eyegaze training in medicine, part of a "Gaze-based User Intent" review document.

References

1 Kundel, H. L., Nodine, C. F., Carmody, D.: Visual scanning, pattern recognition and decision-making in pulmonary nodule detection. Investigative Radiology, 13(3):175–181, 1978.
2 Kundel, H. L.: Reader error, object recognition, and visual search. Keynote speech, Proc SPIE Medical Imaging, 5372:1–11, 2004.
3 Fox, S., Faulkner-Jones, B. E. Eye-Tracking in the Study of Visual Expertise: Methodology and Approaches in Medicine. Frontline Learning Research, 5(3):43–54, 2017.
4 Gallagher AG, Smith CD, Bowers SP, Seymour NE, Pearson A, McNatt S : Psychomotor skills assessment in practicing surgeons experienced in performing advanced laparoscopic procedures. J Am Coll Surg, 197:479–488, 2003.
5 Law, B., Atkins, M. S., Kirkpatrick, A. E., Lomax, A. and MacKenzie, C. L. 2004. Eye gaze patterns differentiate skill in a virtual laparoscopic training environment.Proceedings of the Symposium on Eye Tracking Research & Applications, ACM:41–47, 2004.
6 Zheng, B., Jiang, X., Bednarik, R., Atkins, M. S. Gaze Characteristics of Video Watching in a Surgical Setting. Proceedings 2nd Workshop on Eye Tracking and Visualization, 11-15, 2016.
7 Eivazi, S., Hafez, A., Fuhl, W., Afkari, H., Kasneci, E., Lehecka, M., Bednarik, R.: Optimal eye movement strategies: a comparison of neurosurgeons gaze patterns when using a surgical microscope. Acta Neurochirugica, 159(6):959–966, 2017.
8 Fujii, K., Gras, G., Salerno, A., Yang, GZ. Gaze gesture based human robot interaction for laparoscopic surgery. Medical Image Analysis, 44:196–214, 2018.

5.3 Gaze-based Attention and Intention Recognition: Potentials and Challenges

Roman Bednarik (University of Eastern Finland – Joensuu, FI)

Main reference Roman Bednarik, Hana Vrzakova, Michal Hradis: "What do you want to do next: a novel approach for intent prediction in gaze-based interaction", in Proc. of the 2012 Symposium on Eye-Tracking Research and Applications, ETRA 2012, Santa Barbara, CA, USA, March 28-30, 2012, pp. 83–90, ACM, 2012.
URL http://dx.doi.org/10.1145/2168556.2168569

5.3.1 Introduction

Eye gaze is central in social aspects of life such as communication between people. We use gaze both for directing our own attention in interactions with others, and we employ it for social signaling during conversations. Gaze also has a central role in learning from others. For example in early language learning [9], toddlers employ following of speaker's gaze to obtain cues to resolve ambiguity.

We know that face is a central source of cues to intention recognition [2]. We employ gaze following as a cue to predict the actions of others [6]. Through mapping of the observed action on the internal motor representations of the action, we, effortlessly, initiate motor programs that allow us to direct our gaze to the action in a proactive way.

Finally, as known for instance from competitive games, people can actively jam signals that can be inferred from their own gaze to deceive the opponents and avoid predictability of their intentions.

In communication with interactive systems, one can envision and imagine intelligent architectures capable of similar feats. Such functionality would allow several breakthroughs, in particular, proactive and intelligent interactive systems. Current interfaces still only react to the actions of the users, because miss the predictive capacity that people normally employ in interactions.

Earlier, we designed and implemented a set of studies to systematically verify, whether intention to act can be detected and predicted from gaze [3]. Using a discriminative ML approach at that time, the performance of such system reached about 80% accuracy in an offline mode.

5.3.2 Potentials

There are many instances where a successful prediction of human action would provide tremendous advances. Not only computational agents could be informed by knowing what is the concern of the upcoming action and where it is (i.e., attention and focus prediction), but also what the action related to the object of concern is going to be.

We then will be able to create systems of early warnings, systems capable of predicting and correcting errors, mechanisms for computational resources optimization including foveated displays, proactive agents and assistive technologies.

Some of these technologies have already been introduced. Driver assistance systems are a paramount example. Earlier research focused on employing EEG signals for automotive applications of intention prediction (e.g., [7]), and recently modern computational architectures for early prediction of intention to maneuver have been employed [10]. The future vehicles will benefit from the predictions of the driver's intentions along with attention to engage assistance systems.

Grasping and reaching for objects is one of the most ubiquitous actions people perform. Prototypes of grasping assistance systems with gaze input both for attention and intention are already being developed [8]; in future similar systems will seamlessly provide support for users in performing both everyday actions, and in specialized critical-domains such as surgical tasks [1] and their training.

5.3.3 Challenges and Questions

Challenges are many. Assuming a technical maturity of gaze tracking systems, one of the main challenges related to accurate intent prediction from gaze lies in computation modeling: what representations of gaze are efficacious such that actionable responses can be performed by an artificial intelligence? While there is little doubt that eye-movement data indeed contains intention-related information, currently we have very little understanding what combination of features carries this information, whether and how much these are user and task specific, and what other variables may be at play.

Another challenge that the gaze-based computational intent modeling community will need to address is the fusion of gaze, other signals and contextual information, across multiple time scales. For example, it has been found that the lane-change intention needs to be interpreted in regard with the driving situation [4]. Again in driving scenarios, head-pose have been found as more reliable source of intentions than gaze as an early signal [5]. Therefore, finding the combination of various user-based signals at different epochs preceding the action will be crucial.

Interplay between attention and intention, for example to help disambiguate the target of the upcoming action, will also need to be modeled. This in turn implies reliable computer vision insights into the intention modeling. Ergo, the domain will need to be able to embrace and model multiple sources of data to help in intention detection and prediction.

Aside from the modeling and computational challenges, the very definition of intent differs across studies. Previous works seem to implicitly approach intent as the period before the action, during which the action is planned. I believe that we need a further distinction, mainly in the terms of granularity, to allow for deeper understanding and consequently efficient modeling.

On the way towards creation of these architectures, we will need to establish large community-built datasets with accurate annotations. These would optimally include contextual and other physiological signals. Once available, benchmarking challenges can be organized to advance the research and development.

Solutions to these problems will help in answering the questions related to the trade-off between accurate and timely predictions: How early we can predict an upcoming action (through long- and short-term intention recognition) with a reliable accuracy?

5.3.4 Concluding Remarks

We only begin to understand the complexity of human intention forming and its automatic recognition using gaze. When sensing advances will provide robust data in a non-intrusive way, the computational part of the domain need to be ready for the technology. The recent advances in machine learning techniques help to overcome past feature engineering burdens and promise to propel the research of gaze-based intention recognition.

References

1 Atkins, M. S., Tien, G., Khan, R. S., Meneghetti, A., Zheng, B.: What do surgeons see: capturing and synchronizing eye gaze for surgery applications. Surgical Innovation, 20(3):241–248, 2013.
2 Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., Plumb, I.: The Reading the mind in the eyes test revised version: a study with normal adults, and adults with asperger syndrome or high-functioning autism. The Journal of Child Psychology and Psychiatry and Allied Disciplines 42(2):241–251, 2001.
3 Bednarik, R., Vrzakova, H., Hradis, M.: What do you want to do next: a novel approach for intent prediction in gaze-based interaction. In: Proceedings of the Symposium on Eye Tracking Research & Applications. ACM:83–90, 2012.
4 Beggiato, M., Pech, T., Leonhardt, V., Lindner, P., Wanielik, G., Bullinger-Hoffmann, A., Krems, J.: Lane change prediction: From driver characteristics, manoeuvre types and glance behaviour to a real-time prediction algorithm. In: UR: BAN Human Factors in Traffic, Springer:205–221, 2018.
5 Doshi, A., Trivedi, M. M.: On the roles of eye gaze and head dynamics in predicting driver's intent to change lanes. IEEE Transactions on Intelligent Transportation Systems, 10(3):453–462, 2009.
6 Flanagan, J.R., Johansson, R. S.: Action plans used in action observation. Nature, 424(6950):769–771, 2003.
7 Haufe, S., Treder, M. S., Gugler, M.F., Sagebaum, M., Curio, G., Blankertz, B.: EEG potentials predict upcoming emergency brakings during simulated driving. Journal of Neural Engineering, 8(5):1–11, 2011.
8 Orlov, P., Shafti, A., Auepanwiriyakul, C., Songur, N., Faisal, A. A.: A gaze-contingent intention decoding engine for human augmentation. In: Proceedings of the Symposium on Eye Tracking Research & Applications, ACM:91:1–91:3, 2018.
9 Tomasello, M., Barton, M. E.: Learning words in nonostensive contexts. Developmental Psychology, 30(5):639-650, 1994.
10 Zyner, A., Worrall, S., Ward, J., Nebot, E.: Long short term memory for driver intent prediction. In: IEEE International Conference on Intelligent Vehicles Symposium, IEEE:1484–1489, 2017.

5.4 Gaze Sensing in (Automated) Vehicles

Hans-Joachim Bieg (Robert Bosch GmbH – Stuttgart, DE)

5.4.1 Abstract

Following decades of research, gaze tracking in vehicles is slowly becoming a reality, potentially spearheading a more widespread application of gaze sensing for everyday purposes. The present article summarizes the application of gaze sensing for driver state monitoring and highlights some of the key challenges in this application domain.

5.4.2 Driver Inattention Monitoring

Online driver state assessment using eye trackers in cars or commercial vehicles has been a long-standing research topic (reviewed by [2]), primarily motivated by potential safety benefits. *Driver inattention* due to distraction has been identified as an important crash risk factor (e.g., [11], see [3] for discussion of constructs such as inattention or distraction).

Eye tracking, and gaze sensing in particular, is useful in identifying driver distraction. In the most basic case, methods for online estimation of driver distraction compare a driver's gaze to the most general situational requirement in driving: keeping one's eyes on the road, i.e., by assessing the frequency of glances through a region of interest that resembles the location of the road [14].

More advanced approaches, offering the potential for a more fine-grained assessment of driver inattention, are conceivable by considering information about the specific driving situation that may already be available in modern vehicles from on-board sensors. Examples include adaptation to the vehicle's velocity and associated steering demand [13] or location of objects in the vicinity of the car to assess whether the driver pays attention to these objects or not [4].

In contrast to approaches where driver distraction is directly assessed through observation of the driver's gaze behavior, another strand of research focuses on estimating the driver's secondary task [1]. Based on this, compatibility with the driving task can be assessed, e.g., from expert ratings or crash risk estimates for the specific task.

With the surge in research on automated vehicles, video-based driver state assessment in general and gaze tracking in particular is explored to assure that the driver conforms to the vigilance requirements of automated driving systems. For example, in partially automated systems (SAE level 2; [12]), drivers are obliged to monitor the automated system, i.e., deviations from lateral and longitudinal control much like in manual driving. In higher automation levels, drivers are freed from this task, but still need to display appropriate attentional behavior when taking over control from the automated system [10].

5.4.3 Challenges and Approaches

Video-based interior sensing systems now slowly make their way into the automobile, primarily driven by the demands of automated driving. Gaze tracking enables a range of driver state assessment methods (see previous section) but at the same time poses challenges in regard to robustness and availability.

For example, precise gaze tracking information may not be available at all or very sparsely for some users, due to vision aids or oculomotor limitations (e.g., strabism). This would lock out users from system functions such as the vehicle automation described above. Eye tracking technology has been used as a formidable interaction aid for users with limited manual motor capabilities. It would be ironic if the same technology would establish a new technological obstacle for other users, as applications incorporating gaze information become more ubiquitous. In addition, with myopia on the rise [7], the refinement of methods for increasing tracking performance when tracking users with glasses or contact lenses is crucial for widespread integration of eye tracking in vehicles and other everyday applications.

Apart from groups of users that may be completely excluded from the benefits of eye tracking technology, various personal or situational factors may lead to temporary performance decrements. Tracking may become disrupted by the rims of glasses, hair, clothing, harsh lighting conditions, or behaviors that lead to difficulties in face tracking or extraction of ocular features (e.g., conversing, squinting while smiling).

These issues may be addressed on the level of system and function design, weighing performance against availability and present opportunities for more comprehensive modeling of user attention: Precision requirements can be relaxed to achieve comparable performance for a large variability of users and conditions. For example, systems may content themselves with coarser estimates of driver visual attention, e.g., from head pose information [8] – at the expense of the system's primary detection performance.

In contrast, graded approaches may complement transiently unavailable precise gaze tracking information with information from coarser but more robust inference sources, such as head pose information. Conversely, sparsely available precise gaze tracking information may enhance the ability to interpret head movement behavior by modeling the individual propensity to perform visual orienting either by head or solely by eye [5].

Finally, knowledge about the driving situation as well as other, multi-modal driver information (e.g., vehicle, in-vehicle information system state, or other interior tracking techniques) may be used to assess the driver's state [9]. Driver gaze information may constitute an important but not the only building block, in a more comprehensive assessment.

References

1 Braunagel, C., Stolzmann, W., Kasneci, E., and Rosenstiel, W. Driver-activity recognition in the context of conditionally autonomous driving. In *IEEE International Conference on Intelligent Transportation Systems,* IEEE:1652–1657, 2015.

2 Dong, Y., Hu, Z., Uchimura, K., and Murayama, N. Driver inattention monitoring system for intelligent vehicles: A review. *IEEE Transactions on Intelligent Transportation Systems,* 12(2):596–614, 2011.

3 Engström, J., Monk, C. A., and US-EU Driver Distraction and HMI Working Group. A conceptual framework and taxonomy for understanding and categorizing driver inattention, 2013.

4 Fletcher, L., Loy, G., Barnes, N., and Zelinsky, A. Correlating driver gaze with the road scene for driver assistance systems. *Robotics and Autonomous Systems,* 52(1):71–84, 2005.

5 Fridman, L., Lee, J., Reimer, B., and Victor, T. Owl and lizard: Patterns of head pose and eye pose in driver gaze classification. *IET Computer Vision,* 10(4):308–313, 2016.

6 Hansen, D. W., and Ji, Q. In the eye of the beholder: A survey of models for eyes and gaze. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 32(3):478–500, 2010.

7 Holden, B. A., Fricke, T. R., Wilson, D. A., Jong, M., Naidoo, K. S., Sankaridurg, P., Wong, T. Y., Naduvilath, T. J., and Resnikoff, S. Global prevalence of myopia and high myopia and temporal trends from 2000 through 2050. *Ophthalmology,* 123(5):1036–1042, 2016.

8 Jha, S., and Busso, C. Analyzing the relationship between head pose and gaze to model driver visual attention. In *IEEE International Conference on Intelligent Transportation Systems,* IEEE:2157–2162, 2016.

9 Li, N., and Busso, C. Predicting perceived visual and cognitive distractions of drivers with multimodal features. In *IEEE Transactions on Intelligent Transportation Systems,* 16(1): 51-65, 2015.

10 Marberger, C., Mielenz, H., Naujoks, F., Radlmayr, J., Bengler, K., and Wandtner, B. Understanding and applying the concept of "driver availability" in automated driving. In *Advances in Human Aspects of Transportation,* N. A. Stanton, Ed., Springer:595–605, 2015.

11 Olson, R. L., Hanowsk, R. J., Hickman, J. S., and Bocanegra, J. Driver distraction in commercial operations. Tech. Rep. FMCSA-RRR-09-042, FMCSA, 2009.

12 SAE. Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems. SAE J3016, 2016.

13 Schmitt, F., Bieg, H.-J., Manstetten, D., Herman, M., and Stiefelhagen, R. Predicting lane keeping behavior of visually distracted drivers using inverse suboptimal control. In *IEEE Intelligent Vehicles Symposium,* IEEE:412–418, 2016.

14 Victor, T. W., Harbluk, J. L., and Engström, J. a. Sensitivity of eye-movement measures to in-vehicle task difficulty. *Accident Analysis and Prevention,* 8(2):167–190, 2005.

5.5 Utilizing Eye Tracking Data for User Modeling in Personalized Recommendation

Maria Bielikova (STU – Bratislava, SK)

Main reference Maria Bielikova, Martin Konopka, Jakub Simko, Robert Moro, Jozef Tvarozek, Patrik Hlavac, Eduard Kuric: "Eye-tracking en masse: Group user studies, lab infrastructure, and practices," Journal of Eye Movement Research, 11(3):6:1–6:15, 2018.
URL http://dx.doi.org/10.16910/jemr.11.3.6

5.5.1 Abstract

Although a lot of attention has been dedicated towards user modeling for personalized recommendation, user model representations and its exposing in the recommendation algorithms, there is still open space on inputs to the user modeling process. Traditionally just mouse, gestures and keyboard inputs are considered. However, gaze presents more detailed and accurate information on the user current activity. It enables to acquire an instant stream of data on the user perception of items being recommended. Moreover, utilizing eye tracking data enables to acquire other important features such as pupil size or head distance highly relevant for the task of recommendation as predictors of affective and attentional states.

5.5.2 Introduction

To enable recommender systems to suggest suitable items for a particular user, the recommender system should know user preferences and goals, or should be able to infer it from the user feedback. Traditionally, a user in digital environment is model-based on his/her explicit or implicit feedback [1]. Explicit feedback on user intents, interests, skills and knowledge is hard to acquire, people often are not willing to answer questions and in many scenarios it is not possible to get it either. Even though explicit feedback once given is oftentimes qualified as reliable, in many real scenarios its reliability may be low, especially in cases when users' input is somehow forced. In such situations the users do not provide accurate responses either due inability to do so or because they do not pay attention to or even they may want to provide a false feedback. So, implicit user feedback is heavily used to complement or just replace the explicit one in many situations.

Current recommenders placed in a digital environment use inputs for user modeling based mainly on an infrastructure used for implementation of the recommender system:

- *web-based applications* in various domains (e.g., e-shops, e-books for education, various web services such as flight booking, museums, healthcare systems) – traditional inputs are page visits, mouse movements, clicks, keyboard typing;
- *applications on smart phones and tablets* for similar domains as above mentioned web-based systems – traditional inputs are screens visited, gestures, taps, typing;
- *applications on smart glasses* for various scenarios of everyday life – traditional inputs are images of environment.

These inputs represent evidence of the user's intent, but present just a little help in understanding "why" the user has acted in particular way and what is his/her opinion. Gaze data can not only strengthen evidence for particular intent deduced, but also disprove various assumptions on user behavior, e.g., an interest based on his/her activity (e.g., clicks). This can markedly improve recommendations as we get more reliable implicit ratings of the items for particular user currently computed mainly based on learning to rank algorithms. Considering collaborative applications gaze data bring even more accurate inputs.

5.5.3 Envisioned Challenges

We list scenarios and challenges for utilizing eye tracking data as an input for recommendation. They need various levels of gaze as an input from low level signals through features effective for machine learning to inferred knowledge on the user short term or long term characteristics.

Challenges related to the content of eye tracking data:
- detection of user states useful for recommendation – how eye tracking data can help in recognition of confusion, attention, fatigue?
- user skill assessment – how eye tracking data can help in recognizing familiarity with the application, expertise?
- features for machine learning – which features based on eye tracking data can be used for machine learning tasks such as values estimation, classification, clustering, comparing items, finding similar items.

Challenges related to the recommendation algorithms:
- explanation, i.e., presenting reasoning on recommendation to the user using eye tracking data, making recommenders scrutable – should gaze be explicitly presented to the user? can eye tracking data help in increasing trust of users?

Challenges related to processing of eye tracking data following that gaze produces enormous amounts of:
- massive data processing,
- data storage and filtering,
- real time data sharing.

First steps towards ubiquitous gaze are using eye trackers in collaborative or group scenarios. Such scenarios require special infrastructure [2]. Research in this domain has started primarily in educational domain as personalization including recommendations in intelligent tutoring systems is active research area for many years and multiple eye tracker setups are almost exclusively present in educational environments.

References

1 Jawaheer, G. Weller, P. and Kostkova, P. Modeling User Preferences in Recommender Systems: A Classification Framework for Explicit and Implicit User Feedback. ACM Trans. Interact. Intell. Syst, 4(2):8:1–8:26, 2014.
2 Bielikova, M. Konopka, M. Simko, J. Moro, R. Tvarozek, J. Hlavac, P. Kuric, E. Eye-tracking en masse: Group user studies, lab infrastructure, and practices. Journal of Eye Movement Research, 11(3):6:1–6:15, 2018.

5.6 Mixed-initiative Sensemaking Enabled by Ubiquitous Gaze Sensing and Interaction

Leslie Blaha (Pacific Northwest National Lab. – Richland, US)

Main reference Baber, C., Cook, K., Attfield, S., Blaha L. M., Endert, A., and Franklin, L. "A conceptual model for mixed-initiative sensemaking", 2018 CHI Sensemaking Workshop. 2018.

5.6.1 Abstract

We desire computational teammates that can recommend relevant or interesting data sources to support situation awareness, understanding, or decision making. We are currently able to create and transmit more data than people can process and make sense of. Much of this data goes unanalyzed when we lack the human resources to examine it all. Because human analysts have limited capacity for processing data, they must leverage the computational efficiency and data storage capabilities of machines. Machines have the potential to process the large sets intractable for humans to find and recommend relevant information for people, but they need guidance from humans to do so. If teamed, these complementary strengths support useful and timely inferences on large volumes of data, especially in dynamic decision environments where time for sensemaking may be limited. I propose a novel interactive machine learning paradigm directly leveraging gaze information to train machine learning to support sensemaking.

5.6.2 Gaze-enabled Emergency Response Scenario

Video and picture footage is streaming into the emergency management operations center as a wildfire threatens an urban area. This footage both contains valuable information of emergency planning and is greater in volume and velocity than a person can reasonably attend to. One emergency response traffic coordinator is tasked with monitoring the evacuation process. She must keep traffic flowing, route evacuees safely away from the fire, and deploy first responders efficiently to address problems early. She approaches the interactive data display equipped with ubiquitous gaze sensing and interaction technology. She focuses her gaze on the map icon, blinks twice, and opens up the current traffic feeds overlaid on a map of the urban area. Glancing at the video feed from one of the front-line teams, she visually pulls the video to the map with a slow saccade. The computer recognizes that she is checking the fire forecast against the traffic flow and adds an overlay of green-yellow-red coloring to indicate traffic delays. Computational models offer cones of uncertainty for the fire movement. A weather forecast is suggested, and she rejects the suggestion with a quick nod. Instead, she focuses on the location of Fire Team 3. The fixation brings up a live video feed from the chief's helmet camera. The computer asks if she needs a communication channel. Nodding while fixating on the video feed, the computer video calls the chief for a verbal report of conditions. Registering keywords in the report, the computer extracts relevant video clips from the other Fire Team chiefs across the responder locations. The compiled report of fire status is automatically sent to the responder logistics coordinator to consider additional truck deployments. Simultaneously, the traffic coordinator is provided analytic results indicating a need for a road to be blocked and traffic re-routed. She saccades and fixates on the police icon at the top of the screen, initiating a call to the local police chief. Fixating back on the map location with the forecast analytics, she transmits the information to the police department to initiate traffic re-rerouting. With a triple blink, she closes the maps and logs the activity in the operations center research.

5.6.3 Challenges Toward Mixed-Initiative Sensemaking

Ubiquitous gaze sensing and interaction offers novel approaches to enabling mixed-initiative sensemaking on large volumes of streaming data. In dynamic decision environments like the emergency response operations described above, a single person or even a small team does not have the capacity to process all possible data sources to determine which contain important information to assess the situation. However, we can aid humans with computational tools for mixed-initiative sensemaking [1]. Mixed-initiative sensemaking relies on a combination of human and machine intelligence collaborating to complete complex exploratory analysis, reasoning, problem solving, and decision making tasks. Ubiquitous gaze sensing and interaction captures task-related gaze behavior, providing a rich source of information about the data relevant to each individual. We desire intelligent machine analytics using ubiquitous gaze sensing as a key input for providing recommendations about additional data sources or other analytics to help the operators with their tasks. There are a number of computational challenges in the analytics process to achieving mixed-initiative sensemaking. Note that I am assuming the technical capabilities to collect, transmit, and store ubiquitous gaze and interaction data are addressed separately, and herein emphasize the data modeling and interpretation challenges.

Mixed-initiative sensemaking relies on common ground between the human and computational teammates to effectively align the information needs, goals, and interpretations [1]. Establishing common ground is an ongoing process wherein both the human and machine interpret each other's behavior and resolve conflicts. This requires (near) real-time gaze modeling, which constitutes one of the big data challenges for eye movement analytics [2].

Machines need a way to communicate with people in a manner consistent with the sensemaking process and tasks. Visualization of information is an effective way to present the outputs of large scale analytics, and mixed-initiative systems have the potential to suggest relevant data through the visual analytic tools. Additional, machines need a way to ensure information is presented in a size and structure which people will be able to effectively visually process, which can be informed by analytics of gaze tracking.

To be adaptive to the analyst, the machine needs a way to understand the goals and tasks of the user, to track switches of tasks, and to predict which data is relevant and informative to the current tasks. This is challenging in complex sensemaking where a person may be switching frequently between subtasks like information foraging and information synthesis. Machines need a way to infer the task a person is doing from the gaze and interaction behaviors and predict the tasks to which a person is likely to switch to make those transmissions smooth. Doing so ensures information can be ready for the user when it is needed without long analysis or query delays.

Machines need a way to extract task-meaningful gaze behaviors that should be leveraged to inform the machine analytics. Importantly, there are two sets of analytics informing the mixed-initiative sensemaking process: (1) gaze behavior analytics supporting the modeling of the user, the task, and the efficacy of the recommendation and information presentation processes, and (2) analytics from the data interactions informing the data analysis and further mixed-initiative recommendation processes. These may leverage the same gaze inputs but require different calculations and computations. The machine will need flexible algorithms to determine which behaviors are needed to inform the gaze analytics and user modeling and which are needed to inform the data stream analytics.

5.6.4 Gaze-based Interactive Machine Learning

I propose a new interactive machine learning paradigm to support large-scale data analysis leveraging ubiquitous gaze sensing and interaction for mixed-initiative sensemaking. Gaze information provides a rich context for indicating what information is of interest to the analyst, which can be leveraged as a labels or input to both machine learning and computational cognitive models, in addition to the other situation-related data sources. I propose that a mixed-initiative system that combines computational cognitive models of the operator with machine learning creates a computational teammate that can bootstrap small amounts of imagery or information viewed by the operator to analyze large volumes of situation data to support the sensemaking process.

As an analyst is viewing images, the cognitive model is tracking and analyzing gaze behavior to understand the information that is of interest to the analyst. The model provides an interpretation of how the information supports different aspects of the analyst's tasks, including memory or reasoning. A cognitive model of the analyst can predict future steps or tasks the analyst is likely to take and what information might be needed in the data to support this task. It can interface with machine learning or other artificial intelligence-based analytics to search the larger amounts of available data for additional relevant information the analyst will need to support the sensemaking process. Through visual analytic representations, the system recommend that relevant information to the user at a time and in a manner that will not disrupt the current tasks but is readily available when needed (as predicted by the model and adaptive to real-time gaze modeling). Old information or information that seems irrelevant due to lack of attention can be adaptively removed from the interface.

The gaze data, both raw and interpreted by the cognitive models, becomes another dimension on which the machine can train, like another label set useful for machine learning. A classifier, for example, might use fixation points as an indicator of image features of interest and classify other features into "of interest" and "not of interest" categories. After recording the operator viewing a small amount of imagery, then, the machine can pull additional "of interest" data for the operator from the sources and streams that the operator might not otherwise have the bandwidth to view. Interaction sequences supporting repeated tasks can be predicted with decision trees, so the whole sequence of information is made available with fewer steps. As an interactive learning process, the system can track changes in gaze-based interactions that reflect changes in the tasks or information needs of the operators, to adaptively support sensemaking, maintaining common ground.

Note that interactive learning with cognitive models integrated into the system means that the computational processes can be adaptive to individual operators, because cognitive models are tailored to individual operators. For a team of individuals then, the same types of gaze behavior measured ubiquitously combined with interactive machine learning can facilitate sharing of information between operators. Results of analytics from one process can be recommended to others performing similar tasks or transferred between people when their tasks are interdependent.

Importantly, continuous gaze sensing and interaction with the visual information enables an unobtrusive way to extract information from the viewer to inform the machine processes. As discussed at the Ubiquitous Gaze Sensing and Interaction Seminar, the advances in technology and algorithms to support such continuous gaze sensing is within reach; decreasing costs make it a potential new technology for emergency operations. We take advantage of natural viewing behaviors and advances in machine learning and analytics to support operators in time-critical decision making.

References

1 Baber, C., Cook, K., Attfield, S., Blaha L. M., Endert, A., and Franklin, L. A conceptual model for mixed-initiative sensemaking. *2018 CHI Sensemaking Workshop*, 2018.
2 Blascheck, T., Burch, M., Raschke, M. and Weiskopf, D. Challenges and perspectives in big eye-movement data visual analytics. *Big Data Visual Analytics (BDVA)*, IEEE, 2015.

5.7 Pervasive Eye Tracking and Visual Analytics

Tanja Blascheck (INRIA Saclay, FR)

Main reference Tanja Blascheck, Markus John, Kuno Kurzhals, Steffen Koch, Thomas Ertl: "VA^2: A Visual Analytics Approach for // Evaluating Visual Analytics Applications", IEEE Trans. Vis. Comput. Graph., Vol. 22(1), pp. 61–70, 2016.
URL http://dx.doi.org/10.1109/TVCG.2015.2467871

5.7.1 Introduction and Motivation

In this open problem statement, I present my vision of a pervasive use of eye tracking technology. In the future, I imagine that people will use eye tracking technology always and everywhere. This leads to a great potential in improving our daily lives but also comes with many new challenges we have to face and overcome. My vision is that people use eye tracking technology in combination with other sensors (e.g., EEG, GSR, GPS) and devices (e.g., smartwatches, smartphones, augmented reality glasses, other wearables) for a quantified self. For example, we can enhance the mobility of ourselves by monitoring our behavior to ensure, for example, safe travels, while worn devices provide a mechanism to give feedback, for example, on the travel direction based on landmark. Another possible scenario is the improvement of visual data exploration, because we generate more and more data every day, and more novices as well as (domain) experts, want to visually analyze their data. However, the challenges we are facing in these scenario are manifold. Therefore, in this open problem statement, I sketch two possible scenarios for a pervasive use of eye tracking data, the challenges associated with these scenarios, and some possible directions for future work to achieve such a pervasive use of gaze data.

5.7.2 Scenarios

In the following, I sketch two potential scenarios how pervasive eye tracking data can help people in their daily life. The first example focuses on a pervasive use of eye tracking in the context of mobility, in this case, while riding a bike. The second example focuses on the combination of pervasive eye tracking and visualization, in which data collected from many people is used to enhance the experience of data exploration for an individual.

Scenario 1: Pervasive Use of Eye Tracking for Cyclists

Imagine a cyclist named Mary riding her bike down the road on a busy street in a large city. Mary is on her way to a birthday party of a friend who she has never visited before. She has put on her bike helmet, which is equipped with an Electroencephalography (EEG) and an eye tracker with integrated augmented reality glasses, which measures the gaze as well as her head movements. In addition, she is wearing a smartwatch, which measures her Galvanic Skin Response (GSR), which she can use to show information about the direction of travel, her location as well as surroundings, and which gives her feedback about her current stress level and dangerous situations on her way.

The system analyzes the data collected from the worn sensors in real-time comparing it with the surroundings to give her feedback about which way to go or potentially dangerous situations. For example, the system detects that Mary is checking her watch to see which way to go, highlights an important landmark using augmented reality to guide her in the correct direction and detects that she has not seen the car that is approaching from the left. The system then warns her of the car and sends a message to the autonomously driving car at the same time to communicate this possible impact. The car slows down and Mary, warned by the system, sees that the car is stopping and can safely pass the junction following her route.

Scenario 2: Gaze Guided Visual Analytics

Sarah is an interested citizen and wants to investigate her communities' energy consumption to make an appropriate choice about which energy company to choose when she moves to a new apartment. As every laptop, hers is equipped with an eye tracker which records her eye movements and interactions while she is inspecting the website of her community. The energy data is represented using multiple visualizations and Sarah starts exploring these visualizations. The system automatically analyzes her gaze patterns and compares them to an underlying user model based on gaze data collected from a large number of other citizens who have visited the page.

First, the system uses her gaze data to estimate her expertise with the visualizations. After discovering that Sarah has not used the website before, the visualizations adapt themselves and display help information next to Sarah's gaze to educate her how to use the visualizations. After she transitions to the next level of expertise the help information fades away and the system starts to make suggestions on how to proceed with the exploration. Based on the collected eye movement data of many citizens a data narrative has been created which is used to guide Sarah through her exploration. Depending on the interests of Sarah and where she is looking this narrative automatically branches into multiple story lines guiding her through them. Based on her eye movements the system analyzes which data Sarah has already explored and gives her feedback about this and what information she might have missed.

5.7.3 Envisioned Challenges and Solutions

To come closer to these scenarios, we have to overcome different challenges. For the mobility scenario, the most important challenge that has to be solved are more robust eye tracking systems. First, the infrared light from the sun influences the eye tracking glasses and often leads to a loss of the gaze data [10]. In addition, mapping of gaze data for a pervasive use in real-world and real-time systems requires to map the data to known landscapes, for example, using GPS information in combination with the eye tracking data. This combination of different data sources and sensors (e.g., GPS, EEG, GSR) as well as multiple devices (e.g., eye tracking and augmented reality glasses, smartwatch) requires that the data is automatically recorded and synchronized as well as analyzed in real-time [1, 2]. For analyzing and giving visual feedback to the person, novel methods have to be developed that are unobtrusive yet quickly to grasp. For example, smartwatches can be used to display small-scale visualizations about the currently recorded data [4]. However, recording eye movement data in a public environment brings up the question of privacy issues for the person wearing the eye tracking glasses but also for the people that might be recorded while being on the move. Therefore, we have to find ways to ensure privacy.

The challenges for the second scenario include proper guidance when people are using a visual analytics system for data exploration. For example, if a novice is using a novel system containing visualizations, we have to know a person's intention. This requires that we define appropriate user models based on pre-recorded eye movement data, which we can then use as ground truth if a new person is using a system. Another important aspect is to engage and keep people engaged while using such a system. We have to detect from the eye movement data if a person is bored or overwhelmed to counteract this by showing tutorials or switching to an advanced mode. Especially, in scenarios where people are novices and visually illiterate [5] it is important to help them use a novel system. One possibility is to offer appropriate entry points [3] or use storytelling [6, 8] for a compelling narrative. This requires that we develop and provide appropriate guidance mechanisms [7]. Then, we can make sophisticated recommendations [9] to people using appropriate visual feedback about which part of the data to explore next.

References

1. Blascheck, Tanja and John, Markus and Kurzhals, Kuno and Koch, Steffen and Ertl, Thomas. *VA^2: A Visual Analytics Approach for Evaluating Visual Analytics Applications*. IEEE Transactions on Visualization and Computer Graphics, 22(1): 61–70, 2016.
2. Blascheck, Tanja and John, Markus and Koch, Steffen and Bruder, Leonard and Ertl, Thomas. *Triangulating User Behavior Using Eye Movement, Interaction, and Think Aloud Data*. Proceedings of the Symposium on Eye Tracking Research & Applications, ACM:175–182, 2016.
3. Blascheck, Tanja and MacDonald Vermeulen, Lindsay and Vermeulen, Jo and Perin, Charles and Willett, Wesley and Ertl, Thomas and Carpendale, Sheelagh. *Exploration Strategies for Discovery of Interactivity in Visualizations*. IEEE Transactions on Visualization and Computer Graphics, 2018 (in press).
4. Blascheck, Tanja and Besancon, Lonni and Bezerianos, Anastasia and Lee, Bongshin and Isenberg, Petra. *Glanceable Visualization: Studies of Data Comparison Performance on Smartwatches*. IEEE Transactions on Visualization and Computer Graphics, 2018 (in press).
5. Boy, Jeremy and Rensink, Ronald and Bertini, Enrico and Fekete, Jean-Daniel. *A Principled Way of Assessing Visualization Literacy*. IEEE Transactions on Visualization and Computer Graphics, 20(12):1963–1972, 2014.
6. Boy, Jeremy and Detienne, Francoise and Fekete, Jean-Daniel. *Storytelling in Information Visualizations: Does It Engage Users to Explore Data?*. Proceedings of the Annual ACM Conference on Human Factors in Computing Systems, ACM:1449–1458, 2015.
7. Ceneda, Davide and Gschwandtner, Theresia and May, Thorsten and Miksch, Silvia and Schulz, Hans-Jörg and Streit, Marc and Tominski, Christian. *Characterizing Guidance in Visual Analytics*. IEEE Transactions on Visualization and Computer Graphics, 23(1):111–120, 2017.
8. Segel, Edward and Heer, Jeffrey. *Narrative Visualization: Telling Stories with Data*. IEEE Transactions on Visualization and Computer Graphics, 16(6):1139–1148, 2010.
9. Silva, Nelson and Schreck, Tobias and Veas, Eduardo and Sabol, Vedran and Eggeling, Eva and Fellner, Dieter. *Leveraging Eye-gaze and Time-series Features to Predict User Interests and Build a Recommendation Model for Visual Analysis*. Proceedings of the Symposium on Eye Tracking Research & Applications, ACM:13:1–13:9, 2018.
10. Trefzger, Mathias and Blascheck, Tanja and Raschke, Michael and Hausmann, Sarah and Schlegel, Thomas. *A Visual Comparison of Gaze Behavior from Pedestrians and Cyclists*. Proceedings of the Symposium on Eye Tracking Research & Applications, ACM:104:1–104:5, 2018.

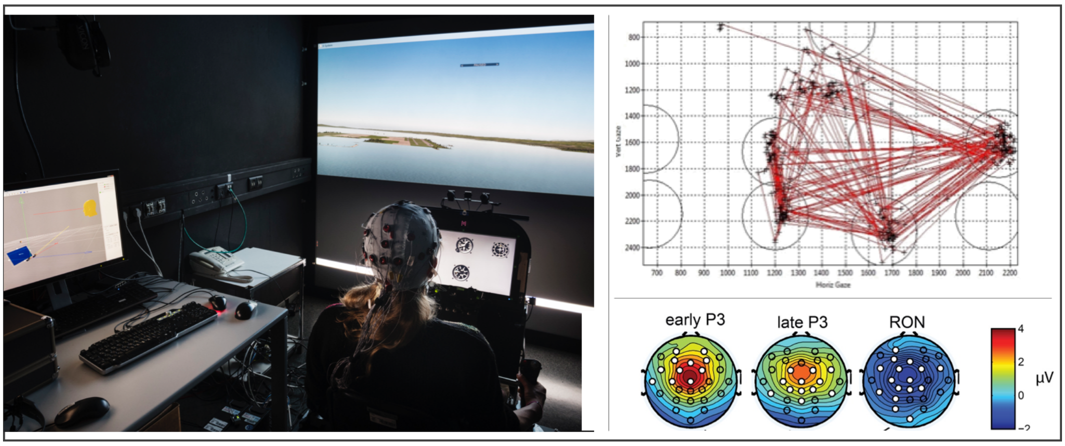

Figure 1 A remote gaze tracking system tracks the eyes of an amateur pilot as she tries to land a fixed-right aircraft (right). Her unpredictable scan path across the flight instruments reveal her levels of anxiety (top-left), and reduced EEG responses to auditory stimuli suggest that she is likely to miss radio messages, i.e., "inattentional deafness" (bottom-left).

5.8 Inferring the Deployment of Limited Attentional Resources

Lewis Chuang (LMU München, DE)

5.8.1 Introduction

Gaze-tracking systems are increasingly prevalent, not only in laboratories but also in our daily environment. The implementation of gaze-tracking systems in the real-world could either be personal, such as head-worn devices, or not, such as those that are implemented in public display systems. In either case, gaze trackers collect data on when the user is looking at what, and how long for. No more, no less. How should we translate this data to draw meaningful inferences of how the user is acquiring and processing visual information? This is necessary in order for us to design computing systems that can be aware of their users needs.

In my lab http://humanmachinesystems.org, we are interested in understanding how limited attentional resources are deployed during user interactions with closed-loop machine systems. We regard attentional resources broadly as any physiological resource that is available to the user, that can allocated to selectively increase the gain of an information channel, often times over other channels; Or as William James [8] have said:

> Attention … is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought, localization, concentration, of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others, and is a condition which has a real opposite in the confused, dazed, scatter brained state which in French is called distraction, and *Zerstreutheit* in German.

For this, we employ the methods of eye tracking and electroencephalography (EEG) (see Figure 1).

5.8.2 Envisioned Challenges

With regards to gaze, this is achieved by moving one's fovea to an area-of-interest (AOI) in the visual scene. This allows spatial information (e.g., edges, contours, object form) to be better resolved and distinguished from the background. This is referred to as overt attention. Many attempts have been made to extract further information from gaze characteristics in order to understand how information might be amplified post-foveation, i.e., covert attention. Unfortunately, our understanding of this is relatively limited. To answer this, it is necessary to first understand the nature of information that serves the purposes of the human observer. This is difficult to answer human observers are unlikely to be consciously operating on image information, in the same terms as would describe the fovea as an image sensor.

It is popular to treat pupil dilation as an index of cognitive load [2]. However, recent findings show that pupil dilation is not only sensitive to ambient lighting, but to the color of the fixated object itself [12]. Clearly, gaze tracking systems are not intend to infer the reflected luminance and ambient luminance of fixated objects. Thus, the depressing message could be that pupil dilation is not a useful measurement at all. However, this dim (and hasty) prognosis is unwarranted [10]. Rather, we need better models of the task and context that gaze is embedded in, in order to discount the variables that influence our estimation of gaze features [11]. A naive solution could be to couple a scene camera with an eye tracker, which would allow us to normalize pupil dilation to ambient illumination. An object recognition algorithm could allow us to further normalize pupil dilation to the likely color of the car model in the scene image.

To summarize, we cannot make meaningful inferences from what an observer is observing ("in-the-wild") without understanding what the observer considers to be meaningful. Furthermore, what the observer perceives to be meaningful could influence our measured gaze characteristic for reasons other than information processing it. Better models that account for context are necessary to allow for meaningful inferences from recorded gaze features.

5.8.3 Envisioned Solutions

We believe that gaze movements are planned actions of a goal-oriented observer that seek out task-relevant information. Thus, it would stand to reason that the predictability of eye movements reflect the ability of the observer to execute this plan. Indeed, we have recently reported that highly anxious *"pilots"* were more likely to generate chaotic and unpredictable eye movements across their flight instruments, especially when they experienced high cognitive load [1]. This is in line with the predictions of attentional control theory [5], which suggests that high anxiety levels and working memory load can compromise executive function. With this in mind, it is possible that ubiquitous gaze tracking systems could be employed to adapt the computing work environment to the user's state.

The EEG response to physical events could also reflect resource availability at the cortical level. In other experiments, we demonstrated that increasing the difficulty of a visuomotor control task diminished the EEG response to irrelevant sounds (i.e., environment sounds), specifically the novelty P3 [13, 14]. The implication is that high cognitive load can prevent users from noticing unexpected but potentially important events, from a calendar notification to emergency warnings. This is termed *inattentional deafness* [4], which could be rectified if the appropriate notifications are presented for the correct context. Another use of EEG could be to validate the design of novel computing systems, in particular those designed to support cognitive work. In another example, we employed EEG to confirm that an *in situ* haptic assembly system that employed augmented reality to reduce visuospatial working memory load, did in fact target the same neural correlates as a standardized test for the same cognitive process [9].

To understand how relevant information is managed and attended to post-fixation, it might be optimal to combine both gaze tracking and EEG methods, e.g., [7, 6], as well as a scene camera to infer the context that this activity is embedded in. More importantly, it is important to ensure that results that are derived under laboratory settings are robust and can generalized to more realistic settings that reflect the variability of real-world settings [3]. As I cast my gaze towards the future, I envision a scenario where a robust understanding of what data means will enable us to delegate to computing systems, the task of making sense of the copious gaze (+ EEG + etc.) data that we are continuously recording of our daily experience.

References

1 Jonathan Allsop, Rob Gray, Heinrich Bülthoff, and Lewis Chuang. Eye movement planning on single-sensor-single-indicator displays is vulnerable to user anxiety and cognitive load. *Journal of Eye Movement Research*, 10(5), 2017.

2 Jackson Beatty. Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, 91(2):276–292, 1982.

3 Lewis L Chuang, Christiane Glatz, and Stas Krupenia. Using eeg to understand why behavior to auditory in-vehicle notifications differs across test environments. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, ACM:123–133, 2017.

4 F. Dehais, M. Causse, François Vachon, N. Regis, E. Menant, and Sébastien Tremblay. Failure to Detect Critical Auditory Alerts in the Cockpit: Evidence for Inattentional Deafness. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 56(4):631–644, 2014.

5 Michael W Eysenck, Nazanin Derakshan, Rita Santos, and Manuel G Calvo. Anxiety and cognitive performance: attentional control theory. *Emotion*, 7(2):336–353, 2007.

6 N. Flad, H. H. Bülthoff, and L. L. **Chuang**. Combined use of eye tracking and EEG to understand visual information processing. In *Proceedings from the International Summer School on Visual Computing (VCSS 2015)*, Fraunhofer Verlag:115–124, 2015.

7 N. Flad, T. Fomina, H. H. Bülthoff, and L. L. **Chuang**. Unsupervised clustering of EOG as a viable substitute for optical eye tracking. In Michael Burch, Lewis Chuang, Brian Fisher, Albrecht Schmidt, and Daniel Weiskopf, editors, *Eye Tracking and Visualization: Foundations, Techniques, and Applications*, Mathematics and Visualization, Springer:151–168, 2016.

8 William James. *The principles of psychology.* Read Books Ltd, 2013.

9 Thomas Kosch, Markus Funk, Albrecht Schmidt, and Lewis L Chuang. Assessing the cognitive demand of assistive systems at assembly workplaces using electroencephalography. *Proceedings of the ACM on Human-Computer Interaction – EICS*, 2018 (accepted).

10 Sebastiaan Mathôt. Pupillometry: Psychology, physiology, and function. *Journal of Cognition*, 1(1), 2018.

11 Vsevolod Peysakhovich, François Vachon, and Frédéric Dehais. The impact of luminance on tonic and phasic pupillary responses to sustained cognitive load. *International Journal of Psychophysiology*, 112:40–45, 2016.

12 Bastian Pfleging, Drea K Fekety, Albrecht Schmidt, and Andrew L Kun. A model relating pupil diameter to mental workload and lighting conditions. In *Proceedings of the Annual ACM Conference on Human Factors in Computing Systems*, ACM:5776–5788, 2016.

13 Menja Scheer, Heinrich H Bülthoff, and Lewis L Chuang. Steering Demands Diminish the Early-P3, Late-P3 and RON Components of the Event-Related Potential of Task-Irrelevant Environmental Sounds. *Frontiers in Human Neuroscience*, 10:73, 2016.
14 Menja Scheer, Heinrich H. Bülthoff, and Lewis L. Chuang. Auditory task irrelevance: A basis for inattentional deafness. *Human Factors*, 1-13, 2018 (in press).

5.9 Short-term Gaze-based User Intent

Andrew Duchowski (Clemson University, US)

Main reference Andrew T. Duchowski: "Gaze-based interaction: A 30 year retrospective", Computers & Graphics, Vol. 73, pp. 59–69, 2018.
URL http://dx.doi.org/10.1016/j.cag.2018.04.002

5.9.1 Introduction

For gaze sensing and interaction to become ubiquitous, use-case scenarios should be developed. Although a myriad tasks may be proposed for everyday use of gaze-based interaction, one compelling assumption made in these scenarios is prediction of the user's intent through analysis of their gaze. For example, when looking at an object, a lamp say, the lamp should infer the user's intent, (turning on the lamp). This style of interaction was envisioned by Vertegaal [8, 9] as Attentive User Interfaces. Since then, various other situations have appeared where gaze-based intent could offer predictive benefit, e.g., decision-support systems [3].

5.9.2 Challenges

In gaze sensing interactive systems, prediction of the user's intent would need to be inferred over variable periods of time, i.e., over the short-, medium-, or long-term. In the medium-term, for example, gaze-based intent could be exploited for divining the next object to be reached for in, for example, a sandwich-making task. Long-term prediction is likely be more complex than in the short-term. The very short-term could be as short as a few milliseconds, during which the location of the saccade end point could be predicted.

5.9.3 Short-Term Prediction of Intent

A fairly straightforward approach to predicting the user's *short-term* visual intent is to estimate what is going to be fixated next by predicting the saccade endpoint mid-flight. The basic premise dates back to [1] who showed that predicting saccade termination was possible by detecting saccadic peak velocity, and then mirroring the saccade velocity profile.

The assumed symmetry of the velocity profile only holds for small amplitude saccades. As saccade amplitudes increase, the velocity profile assumes a Gamma distribution. That is, the velocity profile of small saccades is symmetrical but is skewed for large saccades, and can be modeled by the expression

$$V(t) = \alpha \left(\frac{t}{\beta}\right)^{\gamma-1} e^{-t/\beta}$$

where time $t \geq 0$, and α, $\beta > 0$ are scaling constants for velocity and duration, respectively. Shape parameter $2 < \gamma < 15$ determines the degree of asymmetry [7]. When γ is small, asymmetrical velocity profiles are produced and as γ tends to infinity, the velocity profile assumes assumes a symmetrical (Gaussian) shape.

A more recent approach for saccade endpoint prediction was demonstrated by Arabadzhiyska et al.[2]. They exploit the above observation of saccades obeying ballistic trajectories dependent mainly on saccade amplitude. They then develop and demonstrate an elegant and robust data-driven model that can adequately predict saccade landing position. Model parameters are set by first performing measurements to collect samples of saccades ranging between 5–45 degrees performed by several participants. These samples serve as a kind of look-up table from which saccade characteristics are obtained in real-time use in a foveated display.

5.9.4 Applications: Foveated Displays

Short-term saccade endpoint prediction is particularly well suited to overcoming eye tracking latency associated with foveated rendering. In general, an eye tracker requires at least one frame of eye camera video to compute the gaze point. If there are digital, e.g., , finite-impulse, filters in use to analyze the real-time signal, additional latency is incurred proportional to the filter width.

To be match perceptibility of a full-resolution display, the foveated central inset should appear within 7 ms of fixation onset [5]. Greater delays (e.g., 15 ms following fixation onset), are detectable but have minimal impact on performance of visual tasks when the radius of the foveal inset is large ($\geq 4°$). Due to saccadic suppression, delays as long as 60 ms do not significantly increase blur detection [6]. Note that the latter pertains to the time following saccade termination (60 ms), the former to time following fixation onset (7 ms). Either way, appearance of the foveal inset must be updated before the update is noticed.

Being able to predict, in the short-term, the user's intent to switch gaze to a new location reduces latency of the foveated central inset. Thus estimation of the user's intent through gaze analysis affords computational savings in terms of graphics performance and reduces potential impairment to perception of the scene being rendered.

References

1. James Anliker. *Eye Movements: On-Line Measurement, Analysis, and Control*. Eye Movements and Psychological Processes, Lawrence Erlbaum Associates:185–202, 1976.
2. Arabadzhiyska, Elena and Tursun, Okan Tarhan and Myszkowski, Karol and Seidel, Hans-Peter and Didyk, Piotr. *Saccade Landing Position Prediction for Gaze-Contingent Rendering*. ACM Transactions on Graphics, 36(4):50:1–50:12, 2017.
3. Julia Behrend. *The influence of personality, habits, hierarchy, and role on decision-making in real-life: application to aviation safety*. L'Universite Pierre et Marie Curie, 2018.
4. Han Collewijn, Casper J. Erkelens, and Robert M. Steinman. *Binocular Co-Ordination of Human Horizontal Saccadic Eye Movements*. Journal of Phsyiology, 404:157–182, 1988.
5. Lester C. Loschky and George W. McConkie. *User Performance With Gaze Contingent Multiresolutional Displays*. Proceedings of the Symposium on Eye Tracking Research & Applications, ACM: 97–103, 2000.
6. Lester C. Loschky and Gary S. Wolverton. *How Late Can You Update Gaze-Contingent Multiresolutional Displays Without Detection?*. Transactions on Multimedia Computing, Communications and Applications, 3(4):7:1–7:10, 2007.
7. A J. Van Opstal and J A M Van Gisbergen. *Skewnewss of Saccadic Velocity Profiles: A Unifying Parameter for Normal and Slow Saccades*. Vision Research, 27(5):731–745, 1987.
8. Roel Vertegaal. *Designing Attentive Interfaces*. Proceedings of the Symposium on Eye Tracking Research & Applications, ACM:23–30, 2002.
9. Roel Vertegaal. *Introduction*. Commun. ACM, 46(3):30–33, 2003.

5.10 The Potential of Gaze-based Training in Psychotherapy

Nina Gehrer (Universität Tübingen, DE)

Main reference Nina A. Gehrer, Michael Schönenberg, Andrew T. Duchowski, Krzysztof Krejtz: "Implementing innovative gaze analytic methods in clinical psychology: a study on eye movements in antisocial violent offenders", in Proc. of the 2018 ACM Symposium on Eye Tracking Research & Applications, ETRA 2018, Warsaw, Poland, June 14–17, 2018, pp. 41:1–41:9, ACM, 2018.
URL http://dx.doi.org/10.1145/3204493.3204543

5.10.1 Abstract

Our eye movements guide attentional processes and play a crucial role in our perception and interpretation of the world. Different impairments in these processes have been associated with a variety of psychological disorders. Thus, the use of eye tracking methods in clinical psychology is a promising approach to gain more insight into perceptual and cognitive impairments underlying the etiology of disorders such as autism spectrum disorder (ASD), attention deficit hyperactivity disorder (ADHD), or psychopathy. Furthermore, eye tracking is not only a useful tool for clinical research but also for psychotherapy. The development of gaze-based training as therapeutic intervention or prevention method is a new research area and a promising avenue for innovative treatment strategies.

5.10.2 Introduction

The eyes are the most important interface between our environment and ourselves. The way our gaze scans our surroundings and lingers on certain details determines where we direct our attention, what we perceive and ultimately, how we respond.

In social interactions, detecting and understanding socially important cues is crucial for the functional communication with other individuals and the development of social skills. Thus, we typically show a strong tendency to look at faces and particularly the eyes, which can convey valuable nonverbal information regarding the emotional and cognitive state of the interaction partner [3]. This preference is rooted deeply in our brains and is evident even in infants [4]. Attention-orienting processes to socially salient cues are essential for concepts such as gaze following, joint attention and eye contact, which play a role in the development of higher order social functions such as theory of mind, social bonding and even language acquisition [8]. Accordingly, previous studies have linked impairments in attention orienting to social cues (e.g., the eyes) with various psychiatric conditions, such as autism spectrum disorder (ASD) and psychopathy [9, 2]. Therefore, using eye tracking in clinical research allows us to gain additional insight in psychological processes underlying the perception of social cues and differences that might be associated with dysfunctions or impairments [5].

Furthermore, numerous eye tracking studies have documented deficits in the very basic oculomotor functions in association with psychological disorders [10]. Accordingly, deficiencies in smooth pursuit have been reported in patients with schizophrenia [7]. Further, previous findings in children with ASD or attention deficit hyperactivity disorder (ADHD) have indicated that the associated impairments in inhibition mechanisms might also affect the eye movements [11].

Therefore, investigating eye movements in clinical psychology is a fruitful approach to gain more insight into deficits in oculomotor and attentional processes associated with different psychological disorders.

5.10.3 Potential and Challenges

In clinical psychology and psychotherapy, using eye tracking could help to develop a better understanding of etiology and to learn more about underlying processes of various psychological disorders. Further, eye tracking might be useful as an additional tool in diagnostic procedures if eye movement measures could be shown to provide reliable markers. Finally, gaze-based training is a potentially powerful tool for therapeutic interventions that could address impairments and biases of information processing associated with specific disorders.

First attempts to implement gaze-based training already exist for children with ASD and ADHD. For instance, Chuokoskie et al. [1] developed a robust, low-cost, gaze-contingent game system to train specific oculomotor functions in adolescents with ASD in domestic settings and the preliminary results are very promising. Further, Goodwin et al. [6] are conducting a randomized controlled trial to explore the potential of gaze-based training during infancy as a method to prevent the development of ADHD.

However, the relevance of deficient oculomotor functions and impaired attention orienting in associated psychological disorders remains to be clarified and many important questions are pending: Can these impairments be addressed by specific training? Can improvements in these dysfunctional processes lead to improvements in other symptoms? Do improvements during training transfer to real-life settings? Furthermore, there are technical issues that have to be solved in order to facilitate the development and use of gaze-based therapeutic interventions and to reach their full potential. These challenges start with common issues such as accessibility, robustness, and usability and extend to more complex problems. For instance it would be interesting to develop a mobile eye tracking system that automatically recognizes salient social cues (e.g., eyes or faces) in real-life settings and includes gaze-based recommendations or signals to direct attention to these stimuli. Thus, the development of gaze-based training is a new promising research area albeit the more studies are necessary in order to determine the optimal training targets, methods, their usability, generalizability, and durability.

References

1 Leanne Chukoskie, Marissa Westerfield, and Jeanne Townsend. *A novel approach to training attention and gaze in ASD: A feasibility and efficacy pilot study*. Developmental Neurobiology, 78(5):546–554, 2018.

2 Mark R. Dadds, Yasmeen El Masry, Subodha Wimalaweera, and Adam J. Guastella. *Reduced eye gaze explains "fear blindness" in childhood psychopathic traits*. Journal of the American Academy of Child & Adolescent Psychiatry, 47(4):455–463, 2008.

3 N. J. Emery. *The eyes have it: The neuroethology, function and evolution of social gaze*. Neuroscience & Biobehavioral Reviews, 24(6):581–604, 2000.

4 Teresa Farroni, Gergely Csibra, Francesca Simion, and Mark H. Johnson. *Eye contact detection in humans from birth*. Proceedings of the National Academy of Sciences, 99(14):9602–9605, 2002.

5 Nina A. Gehrer, Michael Schönenberg, Andrew T. Duchowski, and Krzysztof Krejtz. *Implementing innovative gaze analytic methods in clinical psychology*. Proceedings of the Symposium on Eye Tracking Research & Applications, ACM:41:1–41:9, 2018.

6 Amy Goodwin, Simona Salomone, Patrick Bolton, Tony Charman, Emily J. H. Jones, Andrew Pickles, Emily Robinson, Tim Smith, Edmund J. S. Sonuga-Barke, Sam Wass, and Mark H. Johnson. *Attention training for infants at familial risk of ADHD (INTER-STAARS): Study protocol for a randomised controlled trial*. Trials, 17:608, 2016.

7 Sam Hutton and Christopher Kennard. *Oculomotor abnormalities in schizophrenia*. Neurology, 50(3):604–609, 1998.

8 Roxane J. Itier and Magali Batty. *Neural bases of eye and gaze processing: The core of social cognition*. Neuroscience & Biobehavioral Reviews, 33(6):843–863, 2009.
9 Warren Jones, Katelin Carr, and Ami Klin. *Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-old toddlers with autism spectrum disorder*. Archives of General Psychiatry, 65(8):946–954, 2008.
10 Canan Karatekin. *Eye tracking studies of normative and atypical development*. Developmental Review, 27(3):283–348, 2007.
11 Nanda N. J. Rommelse and Stefan Van der Stigchel and Joseph A. Sergeant. *A review on eye movement studies in childhood and adolescent psychiatry*. Brain and Cognition, 68(3):391–414, 2008.

5.11 Eye Movement as Design Material

Hans Gellersen (Lancaster University, GB)

Eye tracking has been long adopted as input device for interaction. The data model of eye trackers makes it convenient to think of them as a pointing device that, much like a mouse, provides a continuous stream of coordinates within a 2D space. Coupled with a display, designers can work with simple abstractions such as points and regions, adopting eye trackers as a black box that hides the intricacies of eye movement. I argue that it is time to open the box – rather than thinking of eye tracking as input device, we should think of eye movement as a material for interaction design.

The conceptual model for gaze interaction appears straightforward: harnessing "what we look at" either as implicit indication of interest a system can respond to, or as an explicit selection of input. However, while a user's mental model might be the same for looking at an object that sits still in the field of view versus one that is in motion, there are fundamental differences in the underlying eye movement processes. Gazing at an object in motion involves smooth pursuit eye movement, a closed loop behavior that is distinct from the saccadic movement otherwise observed. Importantly, this behavior only occurs when there is a moving stimulus for the eyes to follow. This has profound implications for design – whether a user is looking at a moving object can be robustly detected from the correlation of eye movement with the object's motion, as implemented in the *Pursuits* technique [6]. As a consequence, eye tracker and visual environment need only be coupled loosely. There is no need for their coordinate systems to be carefully aligned prior to interaction as the input is based on correspondence of eye motion with motion in the environment. This opens up an entirely new design space for eye gaze: content can be made gaze-aware by presenting it in motion [6]; users can be calibrated implicitly to bootstrap gaze pointing [2]; gaze control can be dynamically associated with ubiquitous devices [5]; and animated widgets can be designed for gaze-only control, even with devices as small as smartwatches [1].

The notion of eye tracking as a pointing device has tended to position eye gaze as alternative to manual action. However, there is a natural interplay and complementarity of eye gaze and manual action – eye gaze precedes action and guides manual input. *Gaze&Touch* demonstrated how the respective strengths of the two modalities can be leveraged for multimodal input, where "gaze selects, and touch manipulates". This can seamlessly extend multitouch and gestural interfaces, by applying manual gestures to objects selected by gaze [3]. Where gaze naturally moves ahead of manual action, manual input can be translated to the gaze location – shifting the frame of reference for manual input, such that the eyes take on the larger and less accurate movement on the interface, while the hand performs smaller and fine-grained input. *Gaze-shifting* also showed that the coincidence of eye gaze and manual input is significant – the same manual input action can take on different meanings, modulated by gaze attention [4]. How to couple gaze and manual action will be particularly intriguing as we move from touch to touchless interactions that at present lack clear conceptual models.

There is much more to gaze that is waiting to be uncovered and developed for interaction design. State of the art eye trackers are single user devices, and there is a vast space to explore multi-user eye gaze, concepts such as mutual gaze and joint attention, and gaze as social signal [7]. As we move from the narrow fields of view that we have in front of desktop interfaces, to interaction with virtual and augmented reality, consideration of gaze will also require a more holistic approach that accounts for head and body movement. Gaze shifts in the real world involve complex interaction of these movements – smaller shifts are performed with the eyes only, whereas larger ones involve head and body movement. How these observations translate to interaction with novel types of display is an open question and fundamental for the design of techniques that couple gaze, head pose, and body pose for natural forms of interface.

References

1. Augusto Esteves, Eduardo Velloso, Andreas Bulling and Hans Gellersen. *Gaze-Shifting: Direct-Indirect Input with Pen and Touch Modulated by Gaze*. Proceedings of UIST'15: User Interface Software & Technology, ACM:457–466, 2015.
2. Ken Pfeuffer, Mélodie Vidal, Jayson Turner, Andreas Bulling and Hans Gellersen. *Gaze-Shifting: Direct-Indirect Input with Pen and Touch Modulated by Gaze*. Proceedings of UIST'13: User Interface Software & Technology, ACM:261–270, 2013.
3. Ken Pfeuffer, Jason Alexander, Ming Ki Chong and Hans Gellersen. *Gaze-touch: combining gaze with multi-touch for interaction on the same surface*. Proceedings of UIST'14: User Interface Software & Technology, ACM:509–518, 2014.
4. Ken Pfeuffer, Jason Alexander, Ming Ki Chong, Yanxia Zhang and Hans Gellersen. *Gaze-Shifting: Direct-Indirect Input with Pen and Touch Modulated by Gaze*. Proceedings of UIST'15: User Interface Software & Technology, ACM:373–383, 2015.
5. Eduardo Velloso, Markus Wirth, Christian Weichel, Augusto Esteves and Hans Gellersen. *AmbiGaze: Direct Control of Ambient Devices by Gaze*. Proceedings of DIS'16: Designing Interactive Systems, ACM:812–817, 2016.
6. Mélodie Vidal, Andreas Bulling and Hans Gellersen. *Pursuits: spontaneous interaction with displays based on smooth pursuit eye movement and moving targets*. Proceedings of UbiComp'13: Pervasive and Ubiquitous Computing, ACM:439–448, 2013.
7. Mélodie Vidal, Rémi Bismuth, Andreas Bulling and Hans Gellersen. *The Royal Corgi: Exploring Social Gaze Interaction for Immersive Gameplay*, Proceedings of the Annual ACM Conference on Human Factors in Computing Systems, ACM:115–124, 2015.

5.12 Why USGI Needs Better Eye Trackers

Kenneth Holmqvist (Universität Regensburg, DE)

Main reference Holmqvist, Kenneth and Andersson, Richard "Eye tracking: A comprehensive guide to methods, paradigms and measures," Lund: Lund Eye-Tracking Research Institute, 2017.

5.12.1 Introduction

None of the UGSI work will be real until we have an eye tracker that can provide good enough data. Loosely based on [1, Chapters 3 and 6], I describe the issues with camera-based VOGs and introduce some of the alternatives.

5.12.2 Envisioned Challenges

Imagine an AR system with a classical video-camera-based system. Our user (consumer?) is trying it out. S/he is doing banking tasks, way finding, checking emails or maybe interacts with an augmented character in a sports application. However because the eye tracker is camera-based, it suffers from a whole range of issues that have been well-studied over the last 20 years. Firstly, the software for rendering the menus and the AR visualizations try to predict the saccadic landing positions mid-flight, which in theory is possible, since saccades are ballistic and have a well-known velocity shape. However, the pupil-corneal reflection technique overestimates the true velocity of the eye [9], and always miscalculates the target by half a degree or more, which makes the rendering seem jumpy. Young people in their teens and up to thirty years of age report that the augmented reality objects slide back and forth, several degrees, for no obvious reason, not knowing that it is because of the variable pupil dilation and the motion of the pupil center that follows with changes in pupil dilation [10]. Furthermore, rendering works better for the brown-eye user than for blue-eyed users who often experience instabilities in the image. Users with contact lenses find the AR objects to be jumping back and forth several degrees at high speed, and refuses to wear it. People who happened to be in the sun when using it report a total loss of functions. A user who was in a room with an old, hot light-bulb report the that all objects are off, and move away from her when she tries to look at them. Everyone report a lot of instability in the imagery, and no-one wants to buy this system [1, Chapter 6].

5.12.3 Envisioned Solutions

Precision, accuracy and robustness at an unprecedented scale is necessary. Precision must be at the absolute minimum to avoid noise in the image. The closer to 0, the better. It is easy to calculate what the accuracy requirements on the eye tracker are if we know the size of the objects that the user interacts with [3]. I think we should not be satisfied until we can distinguish which line of text the user looks at in an email at a normal reading distance. Then we need below 0.05° average accuracy, to be compared with around 0.5° for the best VOGs – assuming inexperienced participants. For a long time, there has been a mistaken belief that the optimal accuracy is limited by the size of the fovea. This is based on the erroneous assumption that we can use any part of the fovea for detailed inspection of very small objects. However, as shown by [2], it is possible to calibrate a DPI system to much finer accuracy, and once having done that, they find movements of the eye smaller than the diameter of the fovea with an accuracy below 0.05°.

Robustness is the tricky parameter to achieve. Current video-based eye trackers work for 90-95% of the student population in Europe, but as soon as they have an eye physiology that makes the eye cleft more narrow, or droopy eye lids, downward going eye lashes, large and variable pupil sizes, contact lenses, blue eyes, mascara or eye-liner, not to speak about glasses with anti-reflective coating, data are not usable even for eye movement studies with lenient data quality requirements.

One important thing is to stop using the pupil as a feature in eye tracking. The pupil feature is the root to many of the issues; it varies in size (causing offsets), the pupil border is more difficult to detect in infrared for blue-eyed people, and the pupil moves differently than the eye ball during saccades, because the inertia in the lens affects the pupil. None of this is OK if we are building an eye tracker for augmented reality, so we cannot use the pupil feature.

Another priority is to stop using cameras. Video cameras that capture frames of the eye have several drawbacks. Firstly, they require a lot of energy, and more energy for higher resolutions. Second, they are pixel-based, which means that the smooth analog movement of the eye is quantized. This creates artifacts in the data when small movements are recorded, so that at gaze directions, the movements are amplified and at other gaze directions, movements are compressed. This effects seems to be worse for lower camera resolutions and fewer corneal reflections. Furthermore, what is the point of recording and transferring a high-resolution image at a high frame rate of many hundred Hz, when we only use two coordinate values from it? It seems utterly wasteful.

The corneal reflection (aka glint) is a better signal. Analogue recordings of eye movements were used from the very start of eye-movement research [4, 5, 6] and data that we have from that time are excellent, but the eye trackers were difficult to use and uncomfortable for participants. Contemporary video-based eye trackers do use the corneal reflection, so in principle we could recreate the good signals. Alas, the quantizing resolution of the eye cameras typically introduce a lot of noise in the signal of the corneal reflection which make small movements (microsaccades) unreliably measured, and smooth pursuit and saccade waveforms noisy.

In 1973, the Dual-Purkinje Imaging (DPI) eye tracker was introduced. Because of its excellent data quality, the DPI has been a major working horse in psychology labs throughout the 1980s and 1990s. It utilizes the corneal reflection plus the reflection the back of the lens, that is, the first and the fourth Purkinje reflections. A system of lenses and mirrors leads the two reflections to each a quadrant detector, which attempts to alter the mirror to keep the reflection at the center. This energy needed to do that is output as two analog voltage signals, one for horizontal and one for vertical. The DPI is difficult to use, by todays standards, but some of the design choices are worth looking into. In particular, the idea of an analog eye tracker with no sampling frequency and no quantization of measurement space are very appealing features for data to be used in augmented reality. Using the 4th Purkinje reflection is not a good idea, as the 4th Purkinje contributes strongly to the erroneous measurements of velocity of saccades, and their post-saccadic oscillations. The 4th Purkinje also disappears behind the pupil border for large off-center gaze directions, which makes the tracking range small for small-pupil user.

Scleral search coils result in excellent data. [7] pronounced coils to be the gold standard in eye tracking. Recently, they have been found to have a data quality on par with the DPI [8]. However, who would want to wear coils on their eye balls during augmented reality activities? And moreover, the participant must be positioned inside an oscillating magnetic field, which is very impractical.

Retinal eye tracking has been used since the 1950s. No other measurement technique can capture such small movements. However, because existing system have been variants of ophtamoloscopes, the eye tracker (camera) efficiently blocks the view of the participant. It has never become a technique that could be sold to researchers.

Today, analog micro-electronics such as MEMS and quadrant detectors can do many of the things the DPI could do. We are likely to very soon see very good eye trackers that are based on such techniques.

References

1. Holmqvist, Kenneth and Andersson, Richard. *Eye tracking: A comprehensive guide to methods, paradigms and measures*. Lund: Lund Eye-Tracking Research Institute, 2017.
2. Poletti, Martina and Rucci, Michele *A compact field guide to the study of microsaccades: challenges and functions*. Vision Research, 118:83–97, 2015.
3. Jacob L Orquin and Kenneth Holmqvist, *Threats to the Validity of Eye-Movement Research in Psychology*, Behavior Research Methods. 50(4):1645–1656, 2017.
4. Dodge, R. and Cline, T. S.*The angular velocity of eye movements*, Psychological Review, 8:145–157, 1901.
5. Buswell, G. T. *How People Look at Pictures*. Chicago: University of Chicago Press, 1935.
6. Yarbus, A. L. *Eye Movements and Vision*. New York, NY: Plenum Press, 1967.
7. Collewijn, H. *Vision Research: A Practical Guide to Laboratory Methods*, Oxford: Oxford University Press, 1988.
8. Hee-kyoung Ko and D. Max Snodderly and Martina Poletti *Eye movements between saccades: Measuring ocular drift and tremor*, Vision Research, 122:93-104, 2016.
9. Ignace Hooge and Kenneth Holmqvist and Marcus Nyström, *Pupil-CR technique is not suitable for studying detailed dynamics of eye movements* Vision Research, 128:6-18, 2016.
10. Drewes, J., *Smaller is better: drift in gaze measurements due to pupil dynamics* PLoS ONE, 9(10):1–6, 2014.

5.13 Who Watches the Watchmen: Eye Tracking in XR

Eakta Jain (University of Florida – Gainesville, US)

Main reference Daniel J. Liebling, Sören Preibusch: "Privacy considerations for a pervasive eye tracking world", in Proc. of the The 2014 ACM Conference on Ubiquitous Computing, UbiComp '14 Adjunct, Seattle, WA, USA – September 13–17, 2014, pp. 1169–1177, ACM, 2014.
URL http://dx.doi.org/10.1145/2638728.2641688

5.13.1 Introduction

The push towards reliable, affordable, and universal eye tracking is being driven by the promise of headset-based mixed reality (XR), which includes augmented reality (AR) and virtual reality (VR), in particular, social VR. Social XR envisions a future where everyone wears headsets for long periods of time. These headsets could be augmented reality or virtual reality, the important thing being that eye trackers will be built into them. At the very beginning, eye trackers will provide the data needed to improve XR systems at the enabling technology level, such as for foveated rendering and alleviating the vergence-accommodation conflict, and for creating the social avatar's eyes. After this, eye tracking data will be used to improve user-friendliness of the system, for example, by combining gaze and gesture to improve gesture recognition, or by creating novel user interaction design by leveraging shared attention. Eye-tracking will also be valued for user identification, for example by iris-scanning, and for improved security, such as continuous authentication via individual specific patterns of eye movements. We will find that these systems can contribute to tracking health and wellness, and become a critical enabler for artificial intelligence-based personalized interventions.

The flip side to these amazing possibilities is that we will allow someone to watch us with an unprecedented intimacy. Eye-tracking data encodes where we are looking, and that is not entirely under our conscious control. Where does a man look when he accompanies his wife shopping? This is clearly private information, said man will argue. Eye-tracking data reveals subtle preferences that can be used for targeted advertisements. In fact, as we will be making our way through virtual and augmented worlds, advertising will shift from clearly demarcated banners to product placement that is integrated into the augmented or virtual experience.[1] Eye-tracking data contains indicators of medical or behavioral conditions [4]. An insurance company might create a social VR app that lets you walk through a hypothetical claims process with a simulated insurance agent, but collects eye movement markers to check for pre-conditions.

In many ways, the scenario above is similar to how we today 'wear' our phones, and allow it to collect multi-modal data on a near-continuous basis. Some of this data is used to improve the quality of the service, such as voice data to improve speech recognition for voice commands, and location data to make search results more relevant to the user. The systems and software architects of these platforms created application programming interfaces (API) that passed this information to third party apps to develop new tools and services to make the user's life better connected, more convenient, safer, and healthier. Though the user is asked to give permission to third party apps to access their data, those developers often request more information than strictly necessary, and users do not fully understand what they are giving up [7, 3, 1]. The interfaces are not necessarily the easiest to navigate and understand [8].

I propose that we think along three vectors before rather than after eye tracking becomes a pervasive sensor in consumer products.

Social VR platform architects: As the platforms for VR and AR develop, the platform creators have full control over the data that is collected and then passed up the software stack. They will determine, for example, the handles that will be provided to third party app developers, the resolution of data they will get, and how well the data are separated. In this domain, the data being recorded will be much richer than previous domains. For example, location, whole body gestures, and fine-grained facial tracking will be needed to create compelling avatars. Eye-tracking data will be collected to enable foveated rendering, consistent focus cues, and to replicate the user's gaze on their social avatar. This last bit is critically different from other domains: even if the user was to turn off foveated rendering, and opt out of gaze-based interaction, they *need* to let the system track their eyes for their avatar in social VR applications. That does not mean they wish to be targeted for advertising based on probabilistic predictions of medical conditions based on eye tracking data for example. From a systems architecture perspective, the open question is: what are the privacy preserving APIs that will control the type and resolution of data that is passed up the software stack to third party apps?

User experience designers: The idea that enabling some form of data gathering for the purpose of task A can also enable task B is not necessarily intuitive to users. Eye tracking

[1] This shift can already be seen in search results where sponsored links are often interspersed with search results.

as a general term encompasses several data: raw gaze location, vergence, pupil diameter, fixations, saccades, blinks, microsaccades, and so on. A lay user may not make the connection that if the social VR app collects eye tracking data to create her real-time avatar as she hangs out with her friends, and has access to the scenes she was looking at (browsing history), then the app can infer what objects she looked at and for how long. The open questions here relate to the visualization and user experience (UX) innovations that are needed to educate the user and give her the controls to customize what she shares.

Eye-tracking technologists: Eye-tracking data is unique because it reveals our interests and preferences as well as other intrinsic characteristics such as affect, age and health. As eye tracking gets built into VR and AR headsets, it should be possible to perform a given task A while guaranteeing that some other task B cannot occur (a privacy guarantee for task B). For example, if high sample rate eye tracking data is smoothed, could we create a believable social avatar without being able to access the markers that might be indicative of degenerative conditions [5]? Several computer science subcommunities have thought about the privacy problem [2, 6]. The eye tracking community needs to think about the tasks that the different eye tracking data can be used for, and the technical frameworks that allow for the separation of different tasks so that one can be turned on and the other turned off. This understanding is critical to enable the responsible use of eye tracking data by platform developers, and the widespread acceptance of this technology by lay users.

References

1 Chia, Pern Hui, Yusuke Yamamoto, and N. Asokan. *Is this app safe?: a large scale study on application permissions and risk signals.* Proceedings of the 21st International Conference on World Wide Web. ACM:311–320, 2012.

2 Dwork, Cynthia, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. *Our data, ourselves: Privacy via distributed noise generation.* In Annual International Conference on the Theory and Applications of Cryptographic Techniques, Springer:486–503, 2006.

3 Felt, A. P., Greenwood, K., and Wagner, D. *The effectiveness of application permissions.* Proceedings of the 2nd USENIX Conference on Web Application Development, 7–7, 2011.

4 Liebling, Daniel J., and Sören Preibusch. *Privacy considerations for a pervasive eye tracking world.* Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, ACM:1169–1177, 2014.

5 Orlosky, Jason, Yuta Itoh, Maud Ranchet, Kiyoshi Kiyokawa, John Morgan, and Hannes Devos. *Emulation of physician tasks in eye-tracked virtual reality for remote diagnosis of neurodegenerative disease.* IEEE Transactions on Visualization and Computer Graphics, 23(4):1302–1311, 2017.

6 Pittaluga, F., Koppal, S. J., and Chakrabarti, A. *Learning privacy preserving encodings through adversarial training.* arXiv preprint arXiv:1802.05214, 2018.

7 Tam, Jennifer, Robert W. Reeder, and Stuart Schechter. *I'm Allowing What? Disclosing the authority applications demand of users as a condition of installation.* Technical report, 2010.

8 Wang, Na, Jens Grossklags, and Heng Xu. *An online experiment of privacy authorization dialogues for social applications.* Proceedings of the Conference on Computer Supported Cooperative Work, ACM:261–272, 2013.

5.14 Gaze-driven Education: Sensing, Understanding, Intervention, and Adaption

Radu Jianu (City – University of London, GB)

Main reference Radu Jianu, Sayeed Safayet Alam: "A Data Model and Task Space for Data of Interest (DOI) Eye-Tracking Analyses", IEEE Trans. Vis. Comput. Graph., Vol. 24(3), pp. 1232–1245, 2018.
URL http://dx.doi.org/10.1109/TVCG.2017.2665498

5.14.1 Abstract

Cheap yet reliable eye trackers now let us collect gaze-data that is unprecedented in scale and diversity. This opens up novel opportunities to advance learning. The new data can help us understand how students learn and how to design more effective learning materials, as well as provide a way to track learners' progress and provide tailored feedback. To take advantage of these opportunities a series of challenges need to be tackled.

5.14.2 Introduction

Reliable eye trackers have become sufficiently affordable that they can now be fitted to regular workstations. This opens up novel opportunities to advance learning. We could collect extensive naturalistic gaze-data from people using computers to learn in schools, universities, and homes. Interpreting such data could: (i) advance our understanding of how students learn; (ii) inform the design of more effective visual learning materials; (iii) inform instructor interventions by capturing what students look at and missed during particular learning sessions; (iv) and support the design of novel learning systems that adapt automatically to enhance the learning experience of students.

5.14.3 Envisioned Challenges

Reaping these benefits hinges on addressing several research challenges. First, we need to collect and store data from many users learning over long periods of time and we need to be able to interpret and make sense of such data effectively. The power of the proposed approach comes from capturing and mining data from hundreds and perhaps thousands of people spending many hours learning using computers in naturalistic settings. In contrast, traditional eye tracking studies dealt with data collected in carefully controlled experiments and at much smaller scales (e.g., a few minutes' worth of gaze-data collected from several tenths of participants). The methods used to collect, store, and interpret gaze-data collected in traditional experiments cannot scale to and cope with the amount and diversity of gaze-data we would collect in naturalistic settings.

Second, we need to determine how to use gazes collected from single and multiple learners to provide personalized recommendations and feedback. Consider instructors who need to assess a student's learning and provide guiding feedback and recommendations in a timely manner and with minimal overhead. They need visual and analytic tools that can summarize students' gazing behavior and capture deviations from normative learning behavior. In other words, we need specialized tools that can support effective diagnostics and interventions in learning.

Finally, we need to find ways in which the next generation of learning systems can use eye tracking to automatically adapt to the needs and progress of individual learners. Specifically, we need to determine how to link individual and multiple learners' gaze-data to useful adaptations of learning systems, and how these adaptations can be shown to learners in ways that are conducive to learning.

5.14.4 Envisioned Solutions

A possible solution to the collection and interpretation of naturalistic eye tracking data lies in the related concepts of semantic areas of interest (AOIs) [1] and data of interest (DOIs) [4]. If learners watch digital content, we can easily match their momentary gaze-points to specific content-items shown on the screen automatically and in real-time [4]. Examples of content-items include concrete definitions, examples, or illustrative images present in the learning material. To support analysis, such content-items can be annotated with descriptive attributes to capture, for example, the type of learning idiom (e.g., 'definition', 'example', 'exercise'), the type of visual representation (e.g., 'text', 'illustration','animation'), or which learning concept they refer to (e.g., 'variable', 'loops', 'expressions').

In this way, an individual student's learning behavior can be captured as a collection of richly-annotated content-items viewed over time. This would facilitate novel data-centric interpretations and analyses. Education researchers and instructors could explore their students' attention data at a level of abstraction that relates to the semantics of the learning materials. For example, a researcher could easily ask if there's a correlation between learning performance, as indicated perhaps by a quiz, and the type of content students look at (e.g., "Do effective learners look at definitions or examples more?", "Do they focus on text or images?"). Alternatively, an instructor could check whether a student skipped too quickly over a particular learning concept or whether they systematically don't pay attention to definitions. The instructor could then provide targeted recommendations to that student.

To build adaptive learning systems that track learners' attention and progress to provide automatic feedback and adaption, we can draw inspiration from existing research into recommendation systems. A useful distinction is that between content-based filtering and collaborative filtering. A system could look at a single learner's performance and adapt based on their individual learning profile (e.g., visual learner vs. verbal learner) and learning progress. Or, a system could mine gaze-data from many learners to infer prototypical learner profiles and effective learning patterns, then match individual learners to such prototypes and guide them along personalized learning paths.

It's important that adaptive responses appear coherent and unobtrusive to learners. It's useful to distinguish between explicit and implicit feedback. Explicit feedback could take the form of clearly distinguishable messages that recommend a course of action to the learner (e.g., revisit a particular definition). Implicit feedback could take the form of subtle, unobtrusive changes in the learning interface. Examples include repeating and possibly rephrasing concepts when the system detects a learner glanced over them, or changing the saliency or positioning of particular learning content based on a learner's reading habits. The design space of implicit feedback is broad and worth investigating.

Incipient research efforts in the directions outlined above already exist. Jianu and Blascheck explored gaze analyses that build on the use of semantic AOIs and DOIs [1, 4]. Conati et al.'s work on adaptive learning systems and gaze-adaptive interfaces provides a valuable stepping stone [2, 3, 5]. However, such efforts are still relatively isolated and more work is needed to fulfill the vision outlined above.

References

1 Blascheck, T., Burch, M., Raschke, M. & Weiskopf, D. Challenges and perspectives in big eye-movement data visual analytics. *Big Data Visual Analytics (BDVA)*, IEEE, 2015.

2 D. Bondareva, C. Conati, R. Feyzi-Behnagh, J. M. Harley, R. Azevedo, and F. Bouchet. *Inferring learning from gaze data during interaction with an environment to support self-regulated learning*. International Conference on Artificial Intelligence in Education, Springer:229–238, 2013.

3 C. Conati, C. Merten, S. Amershi and K. Muldner. *Using eye tracking data for high-level user modeling in adaptive interfaces*. National Conference on Artificial Intelligence, 22(2):1614–1617, 2007.

4 R. Jianu and S. S. Alam. *A Data Model and Task Space for Data of Interest (DOI) Eye-Tracking Analyses*. IEEE Transactions on Visualization and Computer Graphics, 24(3):1232–1245, 2018.

5 S. Kardan and C. Conati. *Comparing and combining eye gaze and interface actions for determining user learning with an interactive simulation*. In International Conference on User Modeling, Adaptation, and Personalization, Springer:215–227, 2013.

5.15 From Lab to the Real World: Eye Tracking Grows Up

Enkelejda Kasneci (Universität Tübingen, DE) and Michael Raschke (Blickshift GmbH – Stuttgart, DE)

Main reference Tanja Blascheck, Michael Burch, Michael Raschke, Daniel Weiskopf: "Challenges and Perspectives in Big Eye-Movement Data Visual Analytics", in Proc. of the Big Data Visual Analytics, BDVA 2015, Hobart, Australia, September 22–25, 2015, pp. 17–24, IEEE, 2015.
URL http://dx.doi.org/10.1109/BDVA.2015.7314288

Visual search behavior plays a key role in our ability to complete everyday activities. Especially during the last decade, eye tracking technology has been increasingly employed in numerous research studies across several application domains to analyze the eye movements of users. Therefore, the eye tracking community has been intensively working towards methods to move the analysis of eye movements out of the laboratory to encompass activities in the real world. Probably one of most interesting application domains in the last decade has been the driving scenarios, for which numerous studies have investigated eye movements of drivers to identify deficits in visual search patterns or types of hazardous situations that may cause accidents. In addition, in the autonomous driving context, eye tracking has been considered as a non-invasive way for driver observation.

Related to application, few months ago in May 2018, Elon Musk tweeted that eye tracking is an ineffective technology for driving assistance[2] systems. In contrast to this statement, there are three main arguments for the effectiveness of eye tracking: (1) While the costs for the eye tracking hardware are continuously decreasing, (2) eye tracking has developed to a pervasive technology. Moreover, (3) we expect a break through regarding the application of eye tracking related to an emerging combination of this technology with perceptual models, advanced machine learning methods and big data technologies.

First, the costs. In 2009, the cost for lab-based eye trackers have been in the range of more than ten thousand Dollars. Since then, the prices have shown a steep drop and

[2] https://twitter.com/elonmusk/status/996102919811350528

will further drop in the near future. Moreover, we believe that within the next ten years, eye tracking will become a widely used standard sensor. Furthermore, in our opinion, eye tracking will have the first break through into the mass market in the automotive sector, followed by VR/AR applications, MedTech, and education.

Especially, the automotive industry forces currently the development of a robust eye detection. Current state of the art eye tracking hardware is not robust with regard to noise and especially with regard to changing illumination conditions [4, 5]. Since current corneal reflection methods are based on infrared LEDs and sensors and sunlight also emits infra-red parts, the quality of eye detection decreases outside laboratory conditions. However, a growing number of new methods to cope with these these challenges have been presented. Many of them are using machine learning algorithms to overcome the light condition problem [6, 7]. In light of these recent developments, we believe that eye tracking technology will mature within the next few years and be applicable to such scenarios.

In contrast to the previously mentioned tweet by Musk, eye tracking has shown to be a promising technology in the context of autonomous driving. The next step towards the fully automated driving is the level of conditional automation, where the autonomous system controls the vehicle for a limited time interval. Some recent work, where eye tracking has been employed to observe the driver, has indicated the effectiveness of this technology to ensure the take-over readiness of the driver in critical situations [2, 3]. More specifically, [3] have proposed the first driver assistance system able to classify the take-over readiness of a driver in conditionally automated driving scenarios. This system works preemptively and at high accuracy, where the driver is warned in advance if a low take-over readiness is to expect.

The third prediction of this position paper is that through the application of machine learning methods there be will powerful analysis methods for eye tracking data in the future [1]. These analysis methods will lead to a quantum leap in the development of perception models for a pervasive understanding of the user's visual perception. Recorded eye movements are the input data for calculating probabilities about which visual objects users have recognized in their environment and which objects they have not seen [9, 11]. As soon as such reliable perception models become available and applicable to online scenarios, current challenges in the realm of human-machine interaction will be solved. In addition, in VR and AR, foveated rendering will not only help to decrease power consumption through reduced computational resources, but will enable a realistic rendering of natural scenes and improve user experience. Just by considering these two specific scenarios (driving and VR/AR), we believe that in the near future the eye movements of millions of users will be recorded and analyzed continuously.

The upcoming combination of new hardware solutions, algorithms from artificial intelligence and big data technologies will change the way how humans interact with computers in a fundamental way. User interfaces will be personalized and machines will adapt the human way how to perceive an environment and will learn to empathically interact with their users. A next step will be that machines will use these learnings from human vision itself to optimize their visual perception and artificial thinking processes.

On the way towards this new human-machine interaction paradigm, fundamental questions have to be answered on how we want to use this new technology. During the Dagstuhl seminar Ubiquitous Gaze Sensing and Interaction 2018 we started a discussion about the process of how to define guidelines for the development of eye tracking towards a broadly accepted technology by the society.

This position paper aims to motivate a continuation of this discussion and even to intense it. We believe that especially discussions about ethical implications and issues of data privacy will be crucial for the further positive development of eye tracking technology and its acceptance by the society. Since eye tracking will become a pervasive technology, possibly affecting millions of people, its misuse has to be avoided.

The main challenge for such perspective discussions is that to date we lack a clear picture of the technology in the future. However, many scientific prototypes reaching from new eye tracking sensors to methods from artificial intelligence to interpret eye movements have indicated what might be possible with more advanced technology. Furthermore, similar user tracking technological solutions from other fields show that big data analytics in the field of personal data can lead to misuse. In light of available solutions from related areas, a first step in the context of eye tracking will be to compare the data structure and data processing in the other fields with the eye tracking technology.

We believe that it is the time to raise awareness in our community on ethical implications of eye tracking technology and to organize a framework for discussion and working groups that might propose guidelines hand in hand with current technological developments. To start such a process, we propose to undergo the following steps:

1. First, we connect and bring together all relevant players in the community to discuss ethical implications and data privacy issues. Scientists from domains like biometric measurement and genetic research will be asked about challenges in their fields, available solutions, and guidelines. During this first step, the main goal will be to raise awareness regarding possible implications of eye tracking technology in the future society.
2. Second, we bring together a small group of key players to conduct further steps.
3. Going from this first connection of people from science and tech, the organizer group will invite other disciplines, such as philosophy, social sciences, and bio sciences to the discourse.
4. A milestone could be a dedicated event, where the social, ethical and data privacy implications will be discussed. A result of this first milestone will be a first draft of tasks, which have to be done towards the development of eye tracking applications that have a positive impact on the society.
5. The next step is to create a scientific program to study the implications of eye tracking on the society in more detail. During this creation, a manifest will be written and frequently updated to provide international guidelines of using eye tracking for the society. This manifest shall be signed by all relevant players in science and in tech community to underline its importance.

Our hope is that this research program will support the further development of eye tracking and acceptance of this promising technology in the future. We should actively influence its development and discuss this new technology with the society.

We would like to thank the Blickshift team, especially Michael Stoll, for a very detailed discussion of this topic.

References

1 Blascheck, T., Burch, M., Raschke, M. & Weiskopf, D. Challenges and perspectives in big eye-movement data visual analytics. *Big Data Visual Analytics (BDVA)*, IEEE, 2015.

2 C. Braunagel, D. Geisler, W. Rosenstiel, E. Kasneci. *Online recognition of driver-activity based on visual scanpath classification*. IEEE Intelligent Transportation Systems Magazine, 9(4):23–36, 2017.

3 C. Braunagel, W. Rosenstiel, E. Kasneci. *Ready for Take-Over? A New Driver Assistance System for an Automated Classification of Driver Take-Over Readines*. IEEE Intelligent Transportation Systems Magazine, 9(4):10–22, 2017.

4 W. Fuhl, M. Tonsen, A. Bulling, E. Kasneci. *Pupil detection for head-mounted eye tracking in the wild: an evaluation of the state of the art*. Machine Vision and Applications, 27(8):1275–1288, 2016.
5 W. Fuhl, D. Geisler, T. Santini, W. Rosenstiel, E. Kasneci. *Evaluation of state-of-the-art pupil detection algorithms on remote eye images*. Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct, ACM:1716–1725, 2016.
6 W. Fuhl, T. Santini, G. Kasneci, E. Kasneci. *PupilNet: Convolutional Neural Networks for robust pupil detection*. arXiv preprint arXiv:1601.04902, 2016.
7 T. Santini, W. Fuhl, E. Kasneci. *PuRe: Robust pupil detection for real-time pervasive eye tracking*. Computer Vision and Image Understanding, 120(5):40–50, 2018.
8 T. Santini, W. Fuhl, E. Kasneci. *Calibme: Fast and unsupervised eye tracker calibration for gaze-based pervasive human-computer interaction*. Proceedings of the Annual ACM Conference on Human Factors in Computing Systems, ACM:2594–2605, 2017.
9 E. Kasneci, T. Kübler, K. Broelemann, G. Kasneci. *Aggregating physiological and eye tracking signals to predict perception in the absence of ground truth*. Computers in Human Behavior, 68:450–455, 2017.
10 T. Kübler, C. Rothe, U. Schiefer, W. Rosenstiel, E. Kasneci. *SubsMatch 2.0: Scanpath comparison and classification based on subsequence frequencies*. Behavior Research Methods, 49(3):1048–1064, 2016.
11 E. Kasneci, G. Kasneci, T. Kübler, W. Rosenstiel. *Online Recognition of Fixations, Saccades, and Smooth Pursuits for Automated Analysis of Traffic Hazard Perception*. Artificial Neural Networks, Springer:411–434, 2015.

5.16 Challenges in Gaze-based Intention Recognition

Peter Kiefer (ETH Zürich, CH)

5.16.1 Introduction

Recent technological developments have made ubiquitous gaze sensing a goal realistically reachable in the near future. This enables new generations of systems that adapt to the user's gaze in order to provide assistance. It is expected that intelligent assistance can be provided by recognizing cognitive states of a user [1]. While previous work has considered the gaze-based recognition of cognitive states such as interest [2], boredom [3], or cognitive load [4], this extended abstract discusses the recognition of a cognitive state that is on a particularly high cognitive level: intention.

5.16.2 Challenges

The following challenges for gaze-based intention recognition can be identified:

Establishing a common terminology and framework: the terminology related to cognition is not used consistently throughout the Human Computer Interaction (e.g., [5]) and eye tracking literature (e.g., [6]). This includes terms, such as, cognitive state, intention, plan, activity, action, cognitive load, goal and task. There is a need to review previous work in these and related fields to establish a common ground.

Selection of gaze features for building the models: different features of gaze have been suggested for gaze-based activity recognition (e.g., [7]). These need to be considered, combined, and possibly extended for inferring higher-level cognitive states.

Bringing together short-term and long-term models: the models for short-term prediction (i.e., in the range of seconds or milliseconds, e.g., [8]) established in the eye tracking and vision research communities need to be combined with models for longer term intention recognition and prediction (i.e., in the range of several minutes) well-known in Artificial Intelligence and Cognitive Science (e.g., [9, 10]).

Accounting for hierarchical, parallel and interleaved intentions: intentions can be seen as hierarchical concepts (i.e., an intention can be implemented by several sub-intentions) that occur in parallel (i.e., a subject may have several intentions at the same time) or interleaved (i.e., an intention can be 'paused' and superseded by some other intention for a while, but picked up again later) (e.g., refer to [11]).

Bottom-up vs. top-down: there is a need to re-visit classic discussions in the literature regarding the benefits, drawbacks and potential combinations of data-driven and model-driven approaches.

Computational methods and platforms for gaining efficiency: gaze data come at high frequency, and the acceptable time lag between the occurrence of an intention and the according assistance is small. This calls for efficient algorithms and computing platforms.

Context-awareness: adding context to the inference model will benefit the recognition accuracy. In particular, knowing the current situation of the user (such as, being at work or in a restaurant) will help in disambiguating which kinds of intentions are possible in that situation (e.g., [12]).

Relation to affective states: the relation between affective states (possibly also inferred from eye movements) and intentions requires further investigation.

Research practices and infrastructure: one challenge for the community consists in creating and sharing gaze datasets for different domains, annotated with intentions, which can be used for benchmarking purposes.

References

1. Andreas Bulling and Thorsten O. Zander. Cognition-aware computing. *IEEE Pervasive Computing*, 13(3):80–83, 2014.
2. Pernilla Qvarfordt and Shumin Zhai. Conversing with the user based on eye-gaze patterns. In *Proceedings of the Annual ACM Conference on Human Factors in Computing Systems*, ACM:221–230, 2005.
3. Peter Kiefer, Ioannis Giannopoulos, Dominik Kremer, Christoph Schlieder, and Martin Raubal. Starting to get bored: An outdoor eye tracking study of tourists exploring a city panorama. In *Proceedings of the Symposium on Eye Tracking Research & Applications*, ACM:315–318, 2014.
4. Andrew T. Duchowski, Krzysztof Krejtz, Izabela Krejtz, Cezary Biele, Anna Niedzielska, Peter Kiefer, Martin Raubal, and Ioannis Giannopoulos. The index of pupillary activity: Measuring cognitive load vis-à-vis task difficulty with pupil oscillation. In *Proceedings of the Annual ACM Conference on Human Factors in Computing Systems*, ACM:282:1–282:13, 2018. ACM.
5. Kristian Lukander, Miika Toivanen, and Kai Puolamäki. Inferring intent and action from gaze in naturalistic behavior: A review. *International Journal of Mobile Human Computer Interaction*, 9(4):41–57, 2017.
6. Roman Bednarik, Hana Vrzakova, and Michal Hradis. What do you want to do next: a novel approach for intent prediction in gaze-based interaction. In *Proceedings of the Symposium on Eye Tracking Research & Applications*, ACM:83–90, 2012.

7 Andreas Bulling, Jamie A Ward, Hans Gellersen, and Gerhard Troster. Eye movement analysis for activity recognition using electrooculography. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(4):741–753, 2011.
8 Elena Arabadzhiyska, Okan Tarhan Tursun, Karol Myszkowski, Hans-Peter Seidel, and Piotr Didyk. Saccade landing position prediction for gaze-contingent rendering. *ACM Transactions on Graphics*, 36(4):50:1–50:12, 2017.
9 Gita Sukthankar, Christopher Geib, Hung Hai Bui, David Pynadath, and Robert P. Goldman. *Plan, activity, and intent recognition: Theory and practice.* Newnes, 2014.
10 John R. Anderson, Michael Matessa, and Christian Lebiere. ACT-R: A theory of higher level cognition and its relation to visual attention. *Human-Computer Interaction*, 12(4):439–462, 1997.
11 Peter Kiefer. *Mobile Intention Recognition.* PhD Thesis, Otto-Friedrich-Universität Bamberg, Germany, Springer, 2011.
12 Peter Kiefer and Christoph Schlieder. Exploring context-sensitivity in spatial intention recognition. In *Workshop on Behavior Monitoring and Interpretation, 30th German Conference on Artificial Intelligence (KI-2007)*, BMI:102–116, 2007.

5.17 Gaze Language: A New Channel of Communication in Augmented Reality

Krzysztof Krejtz (SWPS University of Social Sciences and Humanities, PL)

5.17.1 Introduction

This paper focuses on communication challenges during **Play & Learn** scenario in Augmented Reality. In this scenario a group of agents collaborate together by means of Augmented Reality technology to achieve a common goal. Group collaboration is defined as "a coordinated, synchronous activity that is the result of a continued attempt to construct and maintain a shared conception of a problem" [7]. There are four elements of the collaboration: situational context, interpersonal interaction, mutual problem understanding, and collaboration effects [1]. The successful collaboration effects requires mutual understanding of a problem, and sharing feelings, attitudes, social norms between group members. Shared Reality Theory [4] claims that mutual understanding and feeling enhance personal connection and involvements with the group [2]. The concept of Shared Reality resulted in an enormous body of literature on collaborative work, learn and play (see e.g., [6]).

The common problem understanding and knowledge construction within the group is achieved through interpersonal dialog, which is often internalized, according the the theory of collaborative learning. The interpersonal dialog requires unrestricted communication with minimized noise and permeable channel(s) of communication [8], see Figure 2.

We postulate that the communication channels may be enriched by the constant monitoring and visualization gaze of the each group agent, see Figure 2. The challenges concerning implementation of the gaze communication are deeply rooted in the psychological and cultural context (see [3]), concerning gaze signals meaning, their influence on interpersonal relationships and self-control.

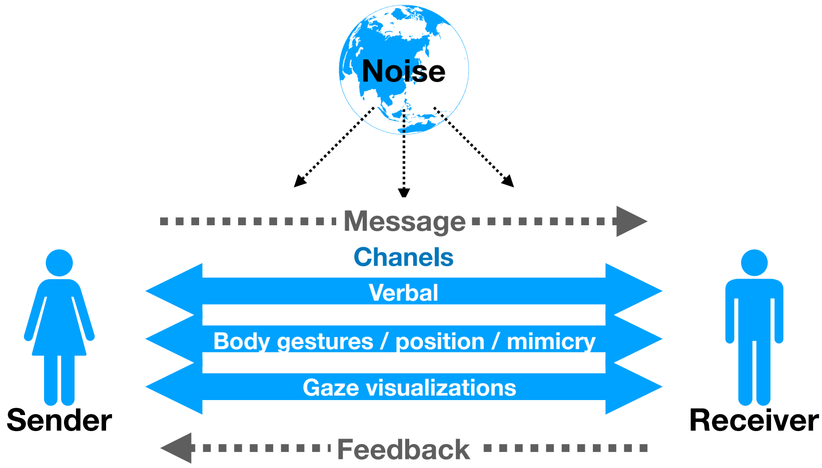

Figure 2 The Shannon-Weaver communication model enriched with Gaze Communication channel.

5.17.2 Gaze Channel Challenges

Gaze social signaling is used by individuals in everyday interpersonal communication. Similarly to gestures, body position and facial expressions or mimicry the gaze role is supportive to the main (usually verbal) channels of communication. It supports verbal channel of communication by intentional and unintentional signaling of the sender emotional or mental state and intent (see [5]). For example, quick glance at the watch during a conversation may communicate the intent of finishing the conversation, or looking away from the interlocutor face may communicate the intent of changing the topic of the conversation.

Gaze social signaling is also used for supporting meta-communication (establishing and communicating the relationship between interlocutors) and is strongly dependent on cultural context. For instance, looking into the interlocutor's eyes may reflect a dominant position, challenging of the partner, or ensuring a good report between them. It has to be stressed that the gaze signaling, similarly to the body gesture communication, has no clear meaning without situational and cultural context.

Ubiquitous gaze monitoring and its online visualization may foster the role of gaze social signaling, changing it into a potentially important communication channel. The gaze visualizations may foster recognition of mental and emotional states between group members. And help them to establish shared reality within the group. As a result faster and more accurate solutions (in he context of a problem solving groups), deeper learning (in context of learning groups) or higher satisfaction from the game (in the context of a group play) may be expected.

Social presence awareness and self-control could also foster the focus on a task and help in establishing useful group norms (see [2]). Seeing that most students focus their visual attention on learning material may help need of learning (a norm of working hard). Noticing that most group members focus their attention on task, in problem solving groups, may trigger the desire for the solution finding.

However, applied to different social situation online display of the gaze may cause new problems. First of all, the continuous visualization of other people gaze may trigger strong social presence awareness and in turn induce higher self-control, fear of being evaluated by others. This may be especially important for socially anxious participants who may undermine their performance because of their fear of social evaluation. Second, the ability of understanding the explicit gaze signals may be limited. The meaning of different gaze signals in group communication will need to be established in the broader process of social negotiations and learning.

5.17.3 Envisioned Solutions

The solutions for the sketched challenges require mainly explicit and implicit training of new social skills of communication with the use of own gaze visualization and reading of the other group members social signaling by visualized gaze. The training for teachers or experts of the use of new social signaling can be prepared and implemented as additional program in the educators curriculum. The training for broader audience can be implemented in a series of gaming apps where the basic elements of own gaze control and others gaze signal understanding may be embed and required for achieving a game goals.

The challenges need also technical solutions which, for example, will allow for momentary disengagement from the group work (switching off the following of others gaze or displaying the own gaze). That would be specially important in the first adaptation phase of the online gaze monitoring and visualization technology to the classrooms, task solving groups or multiuser games situations.

References

1 Pierre Dillenbourg. What do you mean by collaborative learning? In Pierre Dillenbourg, editor, *Collaborative-learning: Cognitive and Computational Approaches*, Elsevier:1–19, 1999.
2 Gerald Echterhoff, E. Tory Higgins, and John M. Levine. Shared reality: Experiencing commonality with others' inner states about the world. *Perspectives on Psychological Science*, 4(5):496–521, 2009.
3 William B. Gudykunst. *Bridging differences: Effective intergroup communication.* Sage Publications Inc, 2004.
4 Curtis D. Hardin and E. Tory Higgins. Shared reality: How social verification makes the subjective objective. In R. M. Sorrentino and E. T. Higgins, editors, *Handbook of Motivation and Cognition, Vol. 3: The Interpersonal Context.*, Guilford Press:28–84, 1996.
5 J. Kunecke, A. Hildebrandt, G. Recio, W. Sommer, and O. Wilhelm. Facial emg responses to emotional expressions are related to emotion perception ability. *PLoS ONE*, 9:1–11, 2014.
6 J. M. Levine, L. L. Thompson, and D. M. Messick. *Shared Cognition in Organizations: The Management of Knowledge.* Organization and Management Series. Taylor & Francis, 2013.
7 Jeremy Roschelle and Stephanie D. Teasley. The construction of shared knowledge in collaborative problem solving. In Claire O'Malley, editor, *Computer Supported Collaborative Learning*, Springer:69–97, 1995.
8 C. E. Shannon, W. Weaver, R. E. Blahut, and Joseph T. Tykociner. *The Mathematical Theory of Communication.* Number Teil 11 in Illini books. University of Illinois at Urbana-Champaign Library and University of Illinois (Urbana) and Warren Weaver, 1963.

5.18 Communicating Visualization with Gaze-guided Storytelling

Kuno Kurzhals

Main reference Chao Tong, Richard C. Roberts, Rita Borgo, Sean Walton, Robert S. Laramee, Kodzo Wegba, Aidong Lu, Yun Wang, Huamin Qu, Qiong Luo, Xiaojuan Ma: "Storytelling and Visualization: An Extended Survey", Information, Vol. 9(3), p. 65, 2018.
URL http://dx.doi.org/10.3390/info9030065

5.18.1 Abstract

Storytelling for visualization is important for multiple reasons: (1) communicating information to people, (2) providing guidance to understand complex data coherences better, and (3) motivating people to engage with the data. From simple infographics to complex visual analytics systems, visualization research in recent years indicates a growing interest of this topic. To this point, storytelling in visualization is realized by static summaries and animations that can be influenced by interaction. We discuss the possibilities of applying eye tracking data as an alternative interaction modality. Such a gaze-guided approach has the advantage that it can individually adapt to the users attention with or without explicit interaction.

5.18.2 Introduction

The dissemination of results plays an important role in all research fields. For the communication with different target audiences, the means of presentation also vary from descriptive statistics, summary reports, and visualization to support findings. With the application of visualization, it is often easier to convey facts and circumstances to a broader audience than with just statistical results that require a certain degree of expertise from the audience. This idea of visualizing numbers and concepts to tell data stories is present and commonly known from infographics [4, 7].

Over the last years, the importance of data storytelling was also emphasized for interactive visualization and visual analytics in scientific [9, 5] and information visualization [1, 6]. A recent overview of existing techniques is provided in the survey by Tong et al. [8]. According to Kosara and Mackinlay, "Presentation–specifically, its use of elements from storytelling–is the next logical step in visualization research and should be a focus of at least equal importance with exploration and analysis" [3]. To achieve this, the authors list, among other aspects, interaction, annotation, and highlighting as important future research directions. With interaction, self-running presentations can be extended to individual experiences of data exploration. With eye tracking technology, it is possible to approximate the users' current attention focus and react to this information. Hence, our focus is on the question: "How can gaze data be incorporated to enhance storytelling for interactive data visualization?" We discuss challenges and scenarios related to this question and how they could be addressed in the future.

5.18.3 Envisioned Challenges

When looking at gaze as an input parameter for interaction, it has to be differentiated between eye tracking for explicit (e.g., as a mouse replacement) and implicit (e.g., an attentive display) input [2].

For explicit input, many scenarios replace the mouse by a gaze cursor as a freehand alternative. Consequently, all related issues, in particular the Midas touch problem have to be addressed when used for interaction with a visualization. One important question here is, how should the visualization react to the current gaze input? On typical scenario could be a public display without touch interface. Here, a narrative presentation could introduce the user to the data and the related circumstances. Then, the user is free to select individual components for further exploration. A direct conversion of established desktop interfaces is not always possible, due to the mentioned issues. Especially navigation through the visualization might be cumbersome without appropriate adjustments to the gaze input.

A more promising direction is the implicit use of gaze to interact with storytelling. Here, the system can make subtle changes to the presentation without the user noticing it. We identified three scenarios that seem promising for further investigation:

Gaze-guidance: The visualization can emphasize specific elements to guide the user's attention during presentation. In contrast to a static design, the system can actually identify if the visual cue was sufficient or has to be intensified (e.g., by flickering highlights).

Attendance-based adjustment: Measuring the gaze distribution and other related metrics, the system can react with respect to the user's attendance. If the user needs time to explore the presented visualization, or if the gaze data indicates low attention, the presentation can be adjusted accordingly, for example by slowing it down, or including more of the aforementioned guidance.

Branching stories: Both examples before assume a fixed storyline. For some cases, it might be beneficial to provide branching paths that adjust to the user, for example to explain an issue in detail. If the system is able to derive an assessment of the subjective understanding based on the gaze data, it can derive from the main story and provide additional explanations to help with the communication of facts. Vice versa, short cuts in the storyline can be taken if the system recognizes that the user is interested in one specific aspect.

We think that explicit and implicit input play an important role for gaze-guided storytelling. While explicit input will be necessary in exploration-focused scenarios, implicit input for guidance and subtle adjustments can significantly enhance storytelling based on animated presentations.

References

1. Nahum Gershon and Ward Page. What storytelling can do for information visualization. *Communications of the ACM*, 44(8):31–37, 2001.
2. Robert JK Jacob. Eye movement-based human-computer interaction techniques: Toward non-command interfaces. *Advances in Human-Computer Interaction*, 4:151–190, 1993.
3. Robert Kosara and Jock Mackinlay. Storytelling: The next step for visualization. *IEEE Computer*, 46(5):44–50, 2013.
4. Jason Lankow, Josh Ritchie, and Ross Crooks. *Infographics: The power of visual storytelling*. John Wiley & Sons, 2012.
5. Kwan-Liu Ma, Isaac Liao, Jennifer Frazier, Helwig Hauser, and Helen-Nicole Kostis. Scientific storytelling using visualization. *IEEE Computer Graphics and Applications*, 32(1):12–19, 2012.
6. Edward Segel and Jeffrey Heer. Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1139–1148, 2010.
7. Mark Smiciklas. *The power of Infographics: Using pictures to communicate and connect with your audiences*. Que Publishing, 2012.
8. Chao Tong, Richard Roberts, Rita Borgo, Sean Walton, Robert S. Laramee, Kodzo Wegba, Aidong Lu, Yun Wang, Huamin Qu, Qiong Luo, et al. Storytelling and visualization: An extended survey. *Information*, 9(3):65:1–65:42, 2018.

9 Michael Wohlfart and Helwig Hauser. Story telling for presentation in volume visualization. In *Proceedings of the 9th Joint Eurographics/IEEE VGTC Conference on Visualization*, IEEE:91–98, 2007.

5.19 Gaze as a Service for Ubiquitous Gaze Sensing and Augmented Reality

David P. Luebke (NVIDIA – Charlottesville, US)

5.19.1 Abstract

Augmented reality offers tremendous promise, but must be coupled with a minimal interface that avoids overwhelming the user with information or requiring cumbersome input. Gaze sensing will be a key component of such interfaces. When designing such an interface, we should think of gaze as a *service* to which an ecosystem of apps and devices can subscribe. Because of the uniquely personal and sensitive nature of gaze data, the community should consider the possible granularities at which that data could be provided or withheld, and the privacy implications for such systems using gaze.

5.19.2 Introduction

Imagine a crowded event, such as a large party or a reception at a conference, full of activity–bands playing, conversations everywhere, people known and unknown. You meet the eyes of a person across the room; they smile and start walking toward you. You unobtrusively tap a finger ring with your thumb and textual information appears directly above the person's head. Perhaps you don't know the person well: this floating label is a virtual name card, reminding you of their name and affiliation. Perhaps you see this person often: the label is a calendar reminder of your lunch meeting tomorrow, or the last couple of texts you exchanged a few hours ago. Perhaps you are actively working on a project with the person; a brief summary of their recent commits to your shared codebase appears. By the time you reach each other and begin talking, your shared context–for they too have this informational superpower–has been established.

Augmented reality (AR) offers the promise of superimposing information on your view of the world, with much industrial and academic research targeting a form factor ultimately as fashionable (or covert) as a pair of glasses. Meanwhile, major leaps are enabling artificial intelligence (AI) to analyze, recognize, and understand your environment. AI is, or will soon be, capable of recognizing the people in the room, the words in their conversations, the social groupings and postures, and the song the band is playing. However, still missing is the user interface to make all that superimposed information useful. Our AR-equipped partygoer does not want their view cluttered with virtual name tags hovering over every person or transcribed speech bubbles from every conversation, any more than a first responder in a crisis situation–a firefighter, fire chief, or medic–wants labels on every bystander or every distant siren or fellow responder.

A useful ARAI system should understand the specific task of the specific user, inferring and presenting only the information needed for that user and that task—while also understanding the high-level goals of the user well enough to flag important or anomalous information that may require changing tasks. Such a system should also predict likely actions for the task and moment, and require absolutely minimal input to enact them-a design principle known as "Do-What-I-Mean" or DWIM. How can even the most advanced AI system predict the user's attention and intent sufficiently? The crucial missing element is gaze: gaze sensing, augmented with various context both internal (such as EEG, ECG, GSR, pupillometry, pulse, etc.) and external (first-person camera or video feed, location services, user history, etc.). Such augmented gaze data–sometimes called "gaze+X"–will prove a crucial element of future augmented reality interfaces.

5.19.3 Challenges

This vision presents many challenges. The gaze tracking itself must be robust, working under almost all conditions (daylight, indoor, nighttime, driving through dappled light with flashes of bright sunlight and shade, etc.) for almost all users (myopic, presbyopic, nystygmatic, amblyopic, etc.). Consumer scenarios will require all-day battery life and a vanishingly unobtrusive form factor; professional scenarios (such as our first responder) may need special hardening such as thermal protection for firefighters. But beyond these hardware challenges lie important system and platform challenges such as handling of privacy and the design of gaze sensing as a service.

5.19.4 Envisioned Solutions

Minimality will be a a key design principle for ARAI systems: require minimal explicit input from the user, and provide minimal output tailored to the user, situation, and task at hand. Modeling user attention and intent from gaze+X will let us minimize the input from the user. Of course gaze sensing, even the nebulous "gaze+X", is not mind-reading. In the scenario above, the use of a simple hands-free affordance–the tap of a finger ring with the thumb of the same hand–plus the user's gaze point, plus enough information about the object of the user's gaze (the other person, and perhaps the fact that their eyes have just met)–gives the additional context needed for a sufficiently advanced and personalized AI system to guess the user's intent and what options to present. Other scenarios might use speech ("Who is that?", "What model is that yellow car over there?", "Where does that door lead?") or more complex tactile affordances. The point is that gaze provides context vital for reducing the input and cognitive effort required to query or instruct the system.

I believe we should think of gaze as a *service* to which apps can subscribe. Such a service would have many different levels; some examples ranging from least private and personal to most sensitive:

- Basic common gestures (probably provided by the operating system and common across all apps) for direct manipulation of UI elements, selection from menus, etc.
- Objects gazed at, again at different levels:
  - Immediately (at the moment the ring is tapped in the above scenario)
  - In recent history (last few seconds, few minutes, today, etc.)
- People gazed at
  - With a special call-out for the action of meeting somebody's eyes, signaling interaction
- Statistics on gaze data, such as one might use to measure health, biometric identification, drowsiness, arousal, cognitive load, etc.
- Raw gaze tracks along with the accompanying first-person video feed

Such a service would exist in an ecosystem of apps, both personal and networked. This network implies a framework for handshaking and consensual sharing of gaze data by different participants in the same area; for example a teacher or trainer could use gaze from students or trainees to better evaluate their understanding. Finally, the community must articulate the privacy concerns, accounting for all the various granularities of gaze data referenced above, and propose solutions to protect privacy and educate users about the risks and benefits of sharing gaze data.

5.20 Basic Explicit Gaze-based Interaction Techniques in VR/MR

Diako Mardanbegi (Lancaster University, GB)

Main reference Ken Pfeuffer, Benedikt Mayer, Diako Mardanbegi, Hans Gellersen: "Gaze + pinch interaction in virtual reality", in Proc. of the 5th Symposium on Spatial User Interaction, SUI 2017, Brighton, United Kingdom, October 16–17, 2017, pp. 99–108, ACM, 2017.
URL https://doi.org/10.1145/3131277.3132180

5.20.1 Introduction

When the first iPhone was launched, it defined a new vocabulary for interaction with computers using a simple set of gestures. Although those interaction methods have been proposed before iPhone, the gesture-driven touchscreen brought gestures to the mainstream. We envision that when the Virtual Reality (VR) and Mixed Reality (MR) devices become more ubiquitous in the near future, they define a new set of interaction techniques that will soon become the mainstream. The promising synergy between eye tracking and near-eye displays that we are observing today suggests that our eyes and in particular gaze would provide a key input for those interaction techniques and play an important role in interaction in VR/MR [2]. However, there is no standard set of gaze interaction techniques that could support basic interaction tasks such as selecting, dragging, zooming, undo, etc.In this report, we address some of the challenges that need to be more thoroughly addressed before XR devices with eye tracking functionality become ubiquitous.

We are relying so much on our smart phones as our main personal computing device that are always with us and provide fast access to information making our communication possible. In terms of form factor we see that displays have become the main component of these personal computers that not only provide visual content but also used as the main interaction channel where users can directly input their commands via manual input. We envision that in near future, smart phones are replaced with some sort of mixed-reality head-mounted devices that are capable of displaying visual content directly in front of the user's eye. Thinking about this new form factor, our eyes seem to be playing an important role in communicating with the device not only as an input channel for visual information but also as an output channel that provide an abundance of information about the subject and the environment (e.g., context, visual attention, cognitive load, biometric, fatigue, health, etc.).

We envision a scenario in which the person wakes up in the morning and puts on his/her wearable computer and uses that for all day. The device provides relevant information such as weather, notifications, news headlines, calendar and schedule depending on where the user is looking at or what the user's mood is. During breakfast, or other daily activities, attention analysis using eye movements could facilitate automatic recognition of the intended activity, detection of potentially missing steps, and providing supportive information. The user is at the train station, he looks at the information shown on the train schedule display which is far. The device would then assist the user by enlarging the information and making the text readable in the field of view. In a driving or cycling task, the device provides navigation assistance by showing the map or by attention guiding that takes into account the attention information from the other drivers on the road and whether for example the user is not paying attention to the surroundings. In the shopping mall, the user can get offers and suggestions based on the information about gaze and eye movements. We could also think of many possible applications where the device facilitates interaction with others in a party or a social event. The simplest examples would be that the device provides information about other people as a memory assistant. The eye movements and gaze data recorded during the day, could be used for automatic summarization and journaling at the end of the day.

While in many of these examples gaze is used implicitly, we envision that for such a continuous use of a AR/MR technology, it's crucial to have a set of few explicit gaze-supported and hands-free interaction techniques to help performing actions such as pointing and selecting digital information or even objects in real world.

There are two unwanted things that we want to prevent from happening in this exciting moment: a) first is when the gaze-based interaction techniques proposed by the first VR/MR devices are not designed appropriately and the users start getting used to a set of nonintuitive and unnatural eye-based interaction techniques which will be hard to correct later, b) and the second condition is when gaze become more like a service that various third party apps are allowed to subscribe to that and utilize that to perform generic tasks such as selection. This may have a hugely negative impact on the overall user experience because different apps may utilize gaze differently to perform similar tasks. Similar thing happened when the Microsoft Kinect provided gesture recognition for games and the users had to often use the body movements and gestures differently to perform the same kind of task across different games affecting the overall user experience of the technology. The other main challenge is that because of the inaccuracy and unreliability of even state-of-the-art eye trackers defining explicit commands that don't work all the time could result in user frustration affecting the user experience.

The above mentioned challenges are mainly associated with the explicit use of gaze. The problem with addressing these challenges is that the VR/MR technologies are still in a premature state and many of the 3D interaction tasks are not fully defined yet. This suggests that the early VR/MR devices with integrated eye tracking should perhaps focus more on the implicit ways of using gaze (e.g., [5]) and avoid defining interaction techniques that require the users to deliberately use their eyes to control UI elements. In the meanwhile, I believe the community should identify a set of explicit commands that can be commonly used across VR/MR devices and even third-party apps for basic tasks such as selection, scrolling, zooming, etc.I also think that because of the inaccuracy issue, the first explicit gaze interaction applications should not be reliant on gaze data. There are already gaze interaction techniques that can be implemented without the need for precise gaze estimation (e.g., [3, 1, 4]) and I believe such techniques could be good candidates for explicit use. Another suggestion would be that any explicit interaction that relies on gaze should potentially come with an alternative method where users can easily switch between modalities when gaze interaction fails.

References

1 Jalaliniya, S., and Mardanbegi, D. Eyegrip: Detecting targets in a series of uni-directional moving objects using optokinetic nystagmus eye movements. Proceedings of the Annual ACM Conference on Human Factors in Computing Systems, ACM: 5801-–5811, 2016.
2 Ling, R., Mardanbeigi, D., and Hansen, D. W. Synergies between head- mounted displays and headmounted eye tracking: The trajectory of development and its social consequences. Living Inside Social Mobile Information, 131-–156, 2014.
3 Vidal, M., Bulling, A., and Gellersen, H. Pursuits: spontaneous interaction with displays based on smooth pursuit eye movement and moving targets. Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing, ACM:439—448, 2013.
4 Zhang, Y., Bulling, A., and Gellersen, H. Sideways: A gaze interface for spontaneous interaction with situated displays. Proceedings of the Annual ACM Conference on Human Factors in Computing Systems, ACM:851—860, 2013.
5 Ken Pfeuffer, Benedikt Mayer, Diako Mardanbegi, and Hans Gellersen. Gaze + pinch interaction in virtual reality. Proceedings of the 5th Symposium on Spatial User Interaction. ACM:99–108, 2017.

5.21 Don’t Make Me Click: Immersive Information Spaces at a Glance

Thies Pfeiffer (Universität Bielefeld, DE)

Main reference Jella Pfeiffer, Thies Pfeiffer, Anke Greif-Winzrieth, Martin Meißner, Patrick Renner, Christof Weinhardt: “Adapting Human-Computer-Interaction of Attentive Smart Glasses to the Trade-Off Conflict in Purchase Decisions: An Experiment in a Virtual Supermarket”, in Proc. of the Augmented Cognition. Neurocognition and Machine Learning – 11th International Conference, AC 2017, Held as Part of HCI International 2017, Vancouver, BC, Canada, July 9–14, 2017, Proceedings, Part I, Lecture Notes in Computer Science, Vol. 10284, pp. 219–235, Springer, 2017.
URL http://dx.doi.org/10.1007/978-3-319-58628-1_18

Mixed reality (MR) technologies allow us to create experiences mixing digital and physical content. As current MR has a strong focus on the visual domain, it seems natural to consider eye tracking as one modality that will allow us to swiftly interact with both visualizations and objects in the physical environment alike. General availability of eye tracking is supported, as it emerged to be a key technology for enabling perceived high resolution rendering for virtual reality (VR) and augmented reality (AR) headsets (foveated/gaze-contingent rendering). With low latency eye tracking technology available in future MR systems, all the required technologies for gaze-based interaction will be readily available. The following text outlines a scenario depicting multiple uses of gaze-based interaction in the context of immersive information spaces.

5.21.1 Introduction: Knowledge Work in a Mixed Reality Future

When talking about digital objects, the majority of it is information (texts, pictures, videos, 3D objects) that has been digitized or digitally created. Most of this information is either linked to physical entities or has established incarnations in physical form (e.g., books, pictures, products). However, as of today, accessing and in particular manipulating these digital objects require knowledge and tools that are in most cases completely different to those that work in our physical reality. For accessing the information, we will have to bring them up to a dedicated surface on a smartphone, tablet or computer screen and then we will have a very abstract, rather generic way of interacting with these objects with a very small set of degrees of freedom, often not appropriate to the format of the information.

In a first step, mixed reality devices will get those displays out of our world. We will no longer see dead black screens in offices or black holes on our walls in the living room. Using wearable mixed reality devices, information can be presented anywhere. With current technology prototypes, such as the Microsoft HoloLens, this vision can already be realized to some extend.

Imagine that, while doing your daily routine in the bathroom, e.g., tooth brushing, you could spend the lazy 2-3 minutes with browsing the recent headlines from your preferred news feed. By monitoring your eye movement patterns, the MR system could detect moments of mindless gaze (or anticipatory gazes at a proxy location where the headlines typically are blended in). This would trigger the presentation of the headlines, e.g., as an overlay on top of your mirror, and by monitoring your attention (what you saw/what you mean), you can get abstracts or full texts in one continuous experience without any explicit interactions (what you get).

In another situation, you are reading a text book on statistics. While going through the texts, you encounter a reference to a statistic procedure that you have no experience with. The MR system detects that you slowed down you reading process, interprets that as uncertainty and offers a brief summary about the procedure hovering beyond the text book. You follow this suggestion, read the summary and the system subsequently will provide additional information (examples, figures, etc.) as a trial, which, when followed with your gaze, will unfold a branching network of available information. Such a concept can easily be extended to libraries [1]. Similar ways to present additional information to existing physical entities can be imagined, e.g., in the area of shopping [5].

But not only receiving information, also information giving can be handled by such a system. When interacting with the personal household robot, areas that require vacuuming or dusting could be communicated to the robot just by gazing at the relevant areas. The robot, in turn, may communicate its schedule to allow the humans to adjust it to match the personal plans (e.g., not to be disturbed while reading the book on statistics) [2, 3].

5.21.2 Envisioned Challenges

As described above, the technologies that are required to realize this vision are around the corner. Major problems are provisioning of power, form factor and the realization of a robust tracking of eye movements that cover 99.9 percent of the population (to not exclude non-trackable persons). Major challenges are more on the human-computer interaction part: robust and generic models for basic aspects of cognitive processing have to be developed (e.g., detecting information search, information processing, reading, mindless gaze, task switches, etc.) that will form the basic atomic "user events" that can be used to trigger more complex interactions. A key interaction metaphor for such unfolding information interfaces would have to support a generic undo/backtracking command.

The MR system also is required to detect the environment in so far as to be able to blend in the digital information in an appropriate fashion (e.g., so that text is readable for the user and at an accessible position). Some of this is already rudimentarily available in systems such as the Microsoft HoloLens, however, not at a quality level that would be required for a smooth integration.

Completely missing is an extensive experience with the design of interactive objects that are physically not interactable. As no one was able to present text in mid air (except for some experiments, such as the HoloPro) or on available surfaces at a larger scale, there are not many design guidelines addressing particular problems that go along with such designs. There is, however, knowledge in the are of the design for head-up-displays, augmented reality or advertisement boards that may be tapped in.

5.21.3 Envisioned Solutions

The analysis of eye movements will play a major role and be a key enabling technology for a smooth interaction with the digital and physical objects [4]. However, one should not expect a gaze-only interface, but a multi-modal interface that integrates gaze with other modalities, such as speech, gestures, and some controller-based system for high-precision inputs (e.g., using textiles). Monitoring head and gaze orientation together with the scanning of the environment and a digital display technology will already a establish a robust human-in-the-loop interaction system.

The proposed solution will thus be that of a gaze-enabled smart glasses system [6, 7] connected to a cloud system with geo-, object- and action-referenced digital information. It will come with a personalized user model (up to basic cognitive and perceptual level).

References

1 Thies Pfeiffer, Friedrich Summann, Jens Hellriegel, Sebastian Wolf, & Christian Pietsch. *Virtuelle Realität zur Bereitstellung integrierter Suchumgebungen*. o-bib. Das offene Bibliotheksjournal, 4(4):94–107, 2017.

2 Sebastian Meyer zu Borgsen, Patrick Renner, Florian Lier, Thies Pfeiffer, & Sven Wachsmuth. *Improving Human-Robot Handover Research by Mixed Reality Techniques*. VAM-HRI 2018. The Inaugural International Workshop on Virtual, Augmented and Mixed Reality for Human-Robot Interaction. 2018.

3 Patrick Renner, Florian Lier, Felix Friese, Thies Pfeiffer, & Sven Wachsmuth. *WYSIWICD: What You See is What I Can Do*. HRI Companion: ACM/IEEE International Conference on Human-Robot Interaction Companion, ACM:382–382, 2018.

4 Jonas Blattgerste, Patrick Renner, & Thies Pfeiffer. *Advantages of Eye-Gaze over Head-Gaze-Based Selection in Virtual and Augmented Reality under Varying Field of Views*. COGAIN : Workshop on Communication by Gaze Interaction, ACM:1:1–1:9, 2018.

5 Jella Pfeiffer, Thies Pfeiffer, Anke Greif-Winzrieth, Martin Meißner, Patrick Renner & Christoph Weinhardt. *Adapting Human-Computer-Interaction of Attentive Smart Glasses to the Trade-Off Conflict in Purchase Decisions: An Experiment in a Virtual Supermarket*. In D. D. Schmorrow and C. M. Fidopiastis (Eds.), Augmented Cognition. Neurocognition and Machine Learning, Springer:219–235, 2017.

6 Thies Pfeiffer, Steven K. Feiner, & Walterio W. Mayol-Cuevas. *Eyewear Computing for Skill Augmentation and Task Guidance*. In A. Bulling, O. Cakmakci, K. Kunze, and J. M. Regh (Eds.), Eyewear Computing–Augmenting the Human with Head-Mounted Wearable Assistants. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 6(1):199–203, 2016.

7 Thies Pfeiffer. *Smart Eyewear for Cognitive Interaction Technology*. In A. Bulling, O. Cakmakci, K. Kunze, and J. M. Regh (Eds.), Eyewear Computing – Augmenting the Human with Head-Mounted Wearable Assistants. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 6(1):177–177, 2016.

5.22 Envisioning Gaze-informed Interaction

Pernilla Qvarfordt (FX Palo Alto Laboratory, US)

Joint work of Pernilla Qvarfordt, Diako Mardenbegi, Shumin Zhai, David Beymer, Jacob T. Biehl, Gene Golovchinsky, Tony Dunnigan
Main reference Qvarfordt, P.: "Gaze-informed multimodal interaction," In The Handbook of Multimodal-Multisensor Interfaces, ACM:365–402, 2017.
URL https://doi.org/10.1145/3015783.3015794

5.22.1 Abstract

Eye tracking technology have developed into smaller and cheaper devices. As a result, usage of eye tracking technology is moving out from the lab and into real world applications. As usage changes, designers of gaze-based interactive system needs to consider how to make the interaction reliable and efficient. Some of these options and factors are outlined here with examples of how they have been designed into some gaze-informed interactive systems.

5.22.2 Introduction

Already from the earliest eye tracking research, we know that our gaze is determined by motor control, perceptual and cognitive factors [1]. Our thinking and intentions are reflected in how we look at things in a particular context. In face-to-face communication, we frequently make use of other's gaze to inform our comprehension of a situation. Although gaze tracking have the potential of being more accurate and provide more powerful interpretation of a users' intent and interest than a human is capable of, many challenges remain to be solved.

To design gaze-informed interaction requires a good understanding of how gaze and other modalities are synchronized and coordinated. Today, this understanding is partial and specific application domains need to be explored and documented. We need to build models of specific interaction scenarios as well as general models of users' preferences and interest. Moving towards truly gaze-informed interaction requires a joint research effort of many researchers with different background and expertise. However, to fully achieve gaze-informed interaction that is natural and smooth, care needs to take how the interaction itself is designed. Here, I outline a few principles that can be used for achieving a robust, smooth and interesting interaction with a system that uses gaze input as one input method.

5.22.3 Short on Gaze Interaction and Gaze-Informed Interaction

When envisioning gaze-interaction, the first thing often imagined is gaze-based pointing. The argument for gaze-interaction is that when our hands are occupied or cannot be used, we can instead the gaze as a pointing device. This explicit or active use of gaze in interaction, is without a doubt important in particular for users who have limited abilities to use a mouse, and is what I call gaze interaction. The gaze signal is used to actively influence the outcome of the system, for instance to point at object and items. The argument against gaze pointing is that it is not natural since we use our eyes to look at our surroundings, not to point with. A mayor problem with gaze interaction is how to distinguish these two cases: looking vs. selecting. Methods range from dwell time [2] to adding additional modalities, be a button press [3, 4], a foot pedal press [5], or even performing a tooth click [6].

Gaze-informed interaction on the other hand, views the information obtained from the gaze signal as one source for understanding the users' interaction. The way we look at the same scene can differ depending on the task we were given or how engaged we are in an activity. Since our eyes are partly driven by our cognitive processes, we can potentially use differences displayed in eye gaze when engaged in a task or an activity and infer from this signal what those activities and tasks are to provide more appropriate context information to the user. If we can entangle the specific gaze patterns for a particular task, it may be possible to build gaze-informed interactions where it appears as the system can read the users' mind in that it can based on data from eye tracking infer user's intention, preferences or workload. This kind of application would use gaze as a completely implicit input method. However, there are many challenges for reaching that vision. One of them being that we need a really good understanding of the task and user behavior when performing the task. Since this form of interaction is implicit, rather than calling it gaze interaction, I call it gaze-informed interaction since the information contained in the gaze informs the interaction with the system.

5.22.4 Principles for Reliable Gaze-based Interaction

One challenge with gaze interaction and gaze-informed interaction is the eye tracking technology. Eye tracking have improved considerably, but when moving out from the lab to the real world performance issues are amplified since the situation is changed. Lighting is very changing, the users make larger and more frequent movements; both these factors affects the reliability, accuracy and precision of the eye trackers. Although, we believe that eye tracking will become more reliable in the future, the principles outlined here can serve to provide an extra layer of reliability when design gaze-based interaction.

Complementing Modalities

Different modalities have different strength and weaknesses. If combined well, the resulting system can become more reliable, efficient or fun to use than each modality used alone. One such example could be MAGIC pointing [11]. In MAGIC pointing, hand and eye works together to make a pointing selection. The long movements with the mouse is performed using eye tracking, while the short precise movements are performed with the mouse. Each modality performs the action that it does with best performance, be it speed or precision.

Mouse pointing does not contain much noise, but when two input methods with fair amount of uncertainty or noise are combined, the result can be increased certainty of user's intended action. For instance, [9] used gaze information to correct errors in speech recognition. Since both speech recognition and gaze data are error prone, Zhang et al. created an N-best list from each modality and used the item highest on both lists as the final result. Slaney at al. [8] used a similar approach for verbal web browsing tasks. Text from the web-page was extracted from regions on the web page where the user looked while speaking. The large vocabulary speech recognizer's N-best list was re-scored based on the attended text. Speech recognition improved by using another noisy signal, eye tracking.

Fall Backs and Redundancies

The second principle is to design fall backs and redundancies for when the gaze signal fails. In gaze pointing, no fall back is provided, since it is the primary input method. If the eye tracking fails, the user cannot make any selections. How a fall back should look like depends on the system. In some cases, the fall back is simply to accept that the gaze signal is absent or is noisy. In [13], we designed a note taking system for wearable displays that used gaze to point at areas of interest within an image captured by the wearable system. The user simply looked at the area of interest, signaled the system to start recording a voice memo.

Where the user looked became the anchor of the annotation. If no gaze could be collected, or the gaze were not stable at the moment the recording started, the system would attach the annotation to the complete image. This would result in a less precise annotation, but this result would be better than not being able to make an annotation.

When designing fall backs and redundancies, it is important to consider the cost of eye tracking failures from a user perspective. One example of this is from a system we designed to support visual inspection [14, 15]. It used eye tracking to suggest regions not inspected that match the characteristics of already viewed regions. In visual inspection, finding all treat targets is the most important factor for success. If the system had failed to record an area as viewed, the cost of inspecting it again is low in comparison to missing treat target. In a different system where, for instance, speed is more important, the cost calculation is likely different and the fall back and redundancies built in would also have been designed differently.

Understanding Context

Within a particular context, the interpretation of the gaze becomes more powerful. A look is not just a look when viewed in context, it can provide an extensive resource for interpretation the user's intention and task progress in an interactive system.

The context can be one of many things, but it is either centered around the user or around the stimuli. Within a user's gaze, multiple signals, such as the pupil size, movement speed and direction, may be extracted and used. Pupil size can be used to detect user's workload [12]. Analyzing gaze patterns, such as consecutive fixations, can provide information of higher level cognitive processes when the user is trying to connect the dots. The user also performs other actions, such as gestures, speech, mouse movements, etc.These actions analyzed together with the gaze can provide a framework for detecting task or task progress. Turning to the stimuli, it contains information as well, e.g., objects, text analysis, etc., that can provide clues to what the user attends to at a specific moment and helps to infer users' task or intention.

How user-center context indicator and stimuli-centered context indicator can work together to provide an adaptive and efficient interaction, is illustrated in a tourist information system we developed [10]. By first looking at how a remote tourist consultant provided information to a tourist using a maps as a visual aid, we identified particular gaze patterns that were telling of users' intention and interest [16]. For example, the tourist consultant often used the tourist's gaze pattern over a map to infer when a particular topic was saturated and it was time to switch to another topic. The tourist gaze patterns often served as indicators of what new topic was of interest. Specific gaze patterns was also found, for instance, the tourist often looked back and forth between two objects identified in the map before asking about distances between them. Based on these finding, we could implement a system that only used gaze to carry out the same conversation as the tourist consultant [10]. Gaze directions, duration and objects, such as bus routes, hotels and attractions, were all used to infer user's interest and specific information need as it changed over time. This example shows the power of the context, using a speech conversation and the visual information in a map, the context could be extracted and modeled so that when only using gaze, the same task could be completed.

5.22.5 Ethical Considerations

Gaze-interaction have clear ethical considerations. Since eye trackers collects information of when and where a person is looking is collected and analyzed, an interactive system would make use of sensitive personal data. An anecdote that I encountered when I started to work with eye tracking, was that of a famous HCI researcher testing eye trackers for usability testing and revealed that he spent considerably time looking at a beautiful woman when seemingly reading a web page. Eye tracking can reveal intimate truths about about a user that he or she may not consciously be aware about. However, using eye tracking as user input device can enrich the user interaction and increase the system performance. Balancing user privacy and system performance is important for making interactive system using eye tracking not only compelling from performance point of view, but also acceptable from a user privacy point of view.

Ethical and privacy needs to be address on all levels in interactive system design. From a privacy perspective, the storing of gaze data is highly sensitive. An interactive gaze-informed system likely do not need have gaze-data stored to perform its function, however, the system might perform better, be more accurate and make better interpretation of the gaze, if user profiles or past interactions are stored and learned from. In interactive system, the gaze data goes through a number of analytical steps. For each step, the resulting data may be either more or less sensitive. By evaluating ethical and privacy effects on each step, the system designer may be able to find a balancing point that allows that system retain powerful analysis while preserving the user's privacy and integrity.

5.22.6 Conclusions

Gaze-based interactive system can provide a highly adaptive and unique user experience. However, reaching this goal is not without challenges. Although eye tracking technology is getting ready for challenges outside the lab, the user's behavior outside the lab can make eye tracking challenging. Technology can improve, but to build reliable gaze interactive system, designers need to think about handling occurrences of incomplete and noisy data. Building in fall backs, redundancies and utilizing context indicator, it is possible to design engaging gaze interactive systems.

References

1 A. L. Yarbus. *Eye Movements and Vision*. Plenum Press, 1967.
2 R. J. K. Jacob. What You Look at is What You Get: Eye Movement-based Interaction Techniques, *Proceedings of the Annual ACM Conference on Human Factors in Computing Systems*, ACM:11–18, 1990.
3 M. Yamato, K. Inoue, A. Monden, A. and K. Torii and K.-I. Matsumoto. Button Selection for General GUIs Using Eye and Hand Together, *Proc. of the Working Conf. on Advanced Visual Interfaces*, ACM:270–273, 2000.
4 M. Kumar, A. Paepcke, and T. Winograd. EyePoint: Practical Pointing and Selection Using Gaze and Keyboard, *Button Selection for General GUIs Using Eye and Hand Together*, ACM:421–430, 2007.
5 K. Klamka, A. Siegel, S. Vogt, F. Göbel, S. Stellmach,and R. Dachselt, R. Look & Pedal: Hands-free Navigation in Zoomable Information Spaces Through Gaze-supported Foot Input, *Proceedings of the Annual ACM Conference on Multimodal Interaction*. ACM:123–130, 2015.

6 X. Zhao, E. D. Guestrin, D. Sayenko, T. Simpson, M. Gauthier, and Popovic. Typing with Eye-gaze and Tooth-clicks, *Proceedings of the Symposium on Eye Tracking Research & Applications*, ACM:341–344, 2012.
7 M. Kaur, M. Tremaine, N. Huang, J. Wilder, Z. Gacovski, F. Flippo, and C. S. Mantravadi. Where is "It"? Event Synchronization in Gaze-Speech Input Systems, *Proceedings of the International Conference on Multimodal Interfaces*, ACM:151–158, 2003.
8 M. Slaney, R. Rajan, A. Stolcke, A. and P. Parthasarathy. Gaze-enhanced speech recognition, *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, IEEE:3236–3240, 2014.
9 Q. Zhang, A. Imamiya, K. Go, and X. Mao. Overriding Errors in a Speech and Gaze Multimodal Architecture, *Proceedings of the International Conference on Intelligent User Interfaces*, ACM:346–348, 2004.
10 P. Qvarfordt, and S. Zhai. Conversing with the User Based on Eye-gaze Patterns, *Button Selection for General GUIs Using Eye and Hand Together*, ACM:221–230, 2005.
11 S. Zhai, C. Morimoto,and S. Ihde. Manual and Gaze Input Cascaded MAGIC Pointing, *Button Selection for General GUIs Using Eye and Hand Together*, ACM:246–253, 1999.
12 M. Pomplun, and S. Sunkara. Pupil dilation as an indicator of cognitive workload in human-computer interaction, *Proceedings of the International Conference on HCI*, 542–546, 2003.
13 D. Mardenbegi, and P. Qvarfordt. Creating Gaze Annotations in Head Mounted Displays, *Proceedings of the International Symposium on Wearable Computers*, ACM:161–162,2015.
14 P. Qvarfordt, J. T. Biehl, G. Golovchinsky, T. Dunningan. Understanding the Benefits of Gaze Enhanced Visual Search, *Proceedings of the Symposium on Eye Tracking Research & Applications*, ACM:283–290, 2010.
15 P. Qvarfordt. Gaze-informed multimodal interaction,*The Handbook of Multimodal-Multisensor Interfaces*, ACM:365–402, 2017.
16 P. Qvarfordt, D. Beymer, and S. Zhai. RealTourist – a study of augmenting human-human and human-computer dialogue with eye-gaze overlay, *Proceedings of INTERACT*, Springer:767–780, 2005.

5.23 Detecting Mindless Gaze

Martin Raubal (ETH Zürich, CH)

Main reference Erik D. Reichle, Andrew E. Reineberg, Jonathan W. Schooler: "Eye Movements During Mindless Reading," Psychological Science, Vol. 21(9), pp. 1300–1310, 2010.
URL https://doi.org/10.1177/0956797610378686

5.23.1 Abstract

Eye-tracking data contains mostly fixations, eye movements that stabilize over a stationary object of interest for a certain temporal duration [1]. Thresholds for determining fixations are arbitrary (about 100ms) and we assume that during a fixation people perceive an object meaningfully, which allows us to infer their cognitive processes [3]. This is an assumption though and the question is how often do people fixate objects 'mindlessly' (looking through an object or daydreaming), i.e., they fixate only the 'syntactic Area Of Interest' but do not relate to its semantics. It is important to detect such mindless gazes because otherwise we would incorrectly infer meaning and cognitive processes.

5.23.2 Introduction

Eye-tracking data contains mostly fixations, eye movements that stabilize over a stationary object of interest for a certain temporal duration [1]. Thresholds for determining fixations are arbitrary (about 100 ms) and we assume that during a fixation people perceive an object meaningfully, which allows us to infer their cognitive processes [3]. This is an assumption though and the question is how often do people fixate objects 'mindlessly' (looking through an object or daydreaming), i.e., they fixate only the 'syntactic Area Of Interest' but do not relate to its semantics. It is important to detect such mindless gazes because otherwise we would incorrectly infer meaning and cognitive processes.

5.23.3 Challenges and Research Questions

The first challenge is to come up with a clear definition of mindless gaze. Looking through an object and therefore not perceiving the stimulus is not the same as perceiving the stimulus but semantically misinterpreting it. Can both be defined as mindless gaze? Once a clear definition is reached, several research questions could be tackled by designing an experiment for detecting mindless gaze:

- What constitutes mindless gaze and which methods are best suited for its detection? This question connects to research on eye movements during mindless reading [4].
- Is there a correlation between mindless gaze and galvanic skin response (GSR) (or other bodily measures)?
- When testing which objects people have perceived, how can one distinguish between short-term memory capacity and mindlessness?
- People may fixate objects during a time-critical task but 'miss them semantically'. Can this be a result of mindless gaze?
- What about tasks where identifying chunks is important (such as when playing chess)? A specific fixation per se is meaningless but successful if the chunk (in chess a meaningful configuration of pieces) is perceived and correctly identified as such.

Several domains and tasks are suitable for an experiment to detect mindless gaze. The experiment must be designed in such a way that allows for testing the participants' semantic interpretation of fixated objects. Objects that were fixated for a certain amount of time but which participants cannot remember afterwards or attach meaning to, may be classified as belonging to 'syntactic AOIs' rather than 'semantic AOIs'. This allows for distinguishing between meaningful and meaningless AOIs in the sense that the former are being utilized for solving the task at hand. One could, for example, imagine the scenario of an emergency center, where people must solve a cartographic map task [2] under time pressure. The type of task is important, therefore we expect different results depending on whether people must solve a concrete problem versus only explore an area.

It will be interesting to see whether the lack of connections of fixations to 'interpreted objects' is sufficient to identify mindless gaze or whether such detection requires data triangulation, e.g., GSR synchronized with the fixations. One can envision several potential application areas for mindless gaze identification, such as learning and education to detect whether pupils are studying or daydreaming.

References

1 Duchowski, A. *Eye Tracking Methodology: Theory and Practice*. Springer, 2017.
2 Göbel, F., P. Kiefer, I. Giannopoulos, A. Duchowski and M. Raubal *Improving map reading with gaze-adaptive legends*. Proceedings of the Symposium on Eye Tracking Research & Applications, ACM:1–9, 2018.

3 Just, M. and P. Carpenter *A theory of reading: From eye fixations to comprehension*. Psychological Review, 87(4):329–354, 1980.
4 Reichle, E., A. Reineberg and W. Schooler *Eye Movements During Mindless Reading*. Psychological Science, 21(9):1300–1310, 2010.

5.24 Challenges and Opportunities of Gaze Sensing in Pervasive Visual Analytics

Daniel Weiskopf (Universität Stuttgart, DE)

Main reference Kuno Kurzhals, Daniel Weiskopf: "Eye Tracking for Personal Visual Analytics", IEEE Computer Graphics and Applications, Vol. 35(4), pp. 64–72, 2015.
URL http://dx.doi.org/10.1109/MCG.2015.47

5.24.1 Abstract

Visual analytics systems were traditionally designed for a professional desktop environment. However, there are recent trends to bring visual analytics to other environments, including smartphones, large display walls, or head-mounted displays. With this trend, I expect that visual analytics will become pervasive. I will discuss challenges and opportunities that come with combining pervasive visual analytics and pervasive gaze sensing, in particular, related to gaze sensing technology, gaze-based interaction, evaluation, and privacy.

5.24.2 Introduction

Visual analytics has been established as a new direction within visualization that focuses on interactive visual interfaces that facilitate the analysis of complex data [5]. Its strength is the combination of visualization, human-computer interaction, and often some kind of integrated and partially automated data analysis (e.g., using machine learning, data mining, or statistical methods). Originally, most visual analytics systems were designed for a professional workplace, typically in a desktop environment. However, there are recent trends to bring visual analytics to other environments. One example is immersive visual analytics, which puts visual analytics into immersive environments, e.g., with head-mounted displays.

Another scenario is visual analytics on smartphones to support the access to pervasive simulation data "out in the field": This could be a civil engineer or an architect running and visualizing a simulation of a newly planned extension of a building within the already existing old building, for example, to assess its impact on the surround and discuss that with stakeholders like the users of the building. Another example includes visual support for first responders that need information access on mobile, robust, and lightweight devices.

Another scenario is personal visual analytics [4], i.e., visual analytics on mobile phones targeting data that is collected in a personal setting, which typically includes quantified-self applications.

All these scenarios heavily rely on non-desktop visual analytics, which comes with many challenges that are common to pervasive applications in general. In the following, I will pick a few challenges and research directions that are particularly relevant for pervasive gaze sensing and visual analytics.

5.24.3 Challenges and Research Directions

Gaze sensing technology

A reliable technological basis is a fundamental requirement that is common to virtually any pervasive gaze sensing application. This is particularly true for pervasive visual analytics because this application will often run for long time spans and in critical (work) environments, i.e., the eye tracking technology should be as unobtrusive and reliable as possible. Also calibration and re-calibration should be simple for the user, ideally, it should be implicit. While there has been much progress in this direction of research and technology, the basis is not yet completely there for pervasive visual analytics. However, with the current speed of development, it is foreseeable that this situation will change in the near to mid-term future, especially since demands come not only from pervasive visual analytics, but from almost all applications of pervasive gaze sensing.

Gaze-based interaction

Gaze-based interaction plays a critical role in many examples of pervasive gaze sensing. This is true for pervasive visual analytics as well, for example, for the general problem of explicit interaction by gaze, but also for indirect approaches like foveated rendering.

However, there are some specific challenges, too. For example, immersive visual analytics is still facing the problem of how to interact with the display of spatial and nonspatial data, including the selection of objects in semi-transparent renderings (such as in volume visualization of scalar fields) or abstract displays of networks or high-dimensional data. Another example is the recognition of user intent or activity, which could support user interaction indirectly; here, recognition mechanisms will have to be adapted to visual analytics, which is different from many other applications of pervasive gaze sensing. The third example is human-robot interaction: Pervasive visual analytics will play an important role in scenarios where human-robot interaction is critical, such as data display for an engineer who works in an industry 4.0 factory or at a building construction site with heavy-load robots. Here, the interaction with the visual display should be tightly linked to the interaction with the robotic system. For many of the aforementioned professional applications, we have to consider how interaction can be scaled across different types of devices. For example, the engineer mentioned above may partially collaborate with her or his colleagues in front of a display wall in a meeting room, and partially out in the factory with a head-mounted display or just a smartphone. Finally, the scenario of personal visual analytics comes with further interaction challenges because we have to support it in a casual setting.

Evaluation

In general, evaluation is difficult for visual analytics because it has to take into account the various aspects of user involvement: how users perceive, understand, and work with the visual representation. There is even a specific workshop series that addresses novel evaluation methodologies for visualization research: The BELIV Workshops ("Evaluation and Beyond – Methodological Approaches for Visualization", https://beliv-workshop.github.io).

Fortunately, pervasive gaze sensing opens up new opportunities for evaluating visual analytics. While eye tracking has been used as a tool for visual analytics research in general [2], pervasive analytics and gaze sensing will come with additional challenges and opportunities: Can gaze sensing serve as a reliable means of quantifying the effectiveness and efficiency of visual analytics? How can we analyze and understand complex and massive gaze data collected during in-the-wild or longitudinal experiments? The latter question leads to the problem of data analysis. Here, visual analytics could play a complementary role–as a means of visually analyzing gaze data [1]. However, the unconstrained settings of pervasive gaze sensing lead to hard data analysis issues that will require us to include the analysis of the visual context surrounding the user [3].

Privacy

Pervasive gaze sensing comes with privacy issues in general because extensive data is collected from the user, but also her or his environment, i.e., others who might be recorded with pervasive camera systems. For pervasive visual analytics, this has also sociological and legal issues because it is often used in a professional setting at the workplace. These issues are complemented by the ones that touch the private sphere in the context of personal visual analytics [4].

References

1 T. Blascheck, K. Kurzhals, M. Raschke, M. Burch, D. Weiskopf, and T. Ertl. Visualization of eye tracking data: A taxonomy and survey. *Computer Graphics Forum*, 36(8):260–284, 2017.
2 K. Kurzhals, B. D. Fisher, M. Burch, and D. Weiskopf. Eye tracking evaluation of visual analytics. *Information Visualization*, 15(4):340–358, 2016.
3 K. Kurzhals, M. Hlawatsch, C. Seeger, and D. Weiskopf. Visual analytics for mobile eye tracking. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):301–310, 2017.
4 K. Kurzhals and D. Weiskopf. Eye tracking for personal visual analytics. *IEEE Computer Graphics and Applications*, 35(4):64–72, 2015.
5 J. J. Thomas and K. A. Cook, eds. *Illuminating the Path: The Research and Development Agenda for Visual Analytics.* National Visualization and Analytics Center, 2005.

6 Conclusion and Outlook

Ubiquitous Gaze Sensing and Interaction turned out to be an engaging Dagstuhl Seminar bringing together researchers and industry with multiple perspectives and backgrounds. The intensity of discussions and the willingness to continue the discussions and create new research partnerships were evident. Several topics attracted much of the participants' attention. In particular, popular discussions were on Data Privacy and Gaze + X. These two topics were discussed in more than one break-out session with different configurations of researchers and sparked collaborations on papers and idea exchange over traditional academic fields. From this perspective, we achieved what we set out to do when planning this seminar.

The final discussion of the seminar was how to continue the discussions sparked at this seminar and how to include more researchers than those present at these discussions. A number of ideas were put forth, these included organizing a special issue on the topic of Gaze + X, conference workshops to invite new researchers from different fields to continue the discussion, and finishing the papers that were getting started during the workshop.

Participants

- Amy Alberts
Tableau Software – Seattle, US
- M. Stella Atkins
Simon Fraser University – Burnaby, CA
- Roman Bednarik
University of Eastern Finland – Joensuu, FI
- Hans-Joachim Bieg
Robert Bosch GmbH – Stuttgart, DE
- Maria Bielikova
STU – Bratislava, SK
- Leslie Blaha
Pacific Northwest National Lab. – Richland, US
- Tanja Blascheck
INRIA Saclay – Orsay, FR
- Andreas Bulling
MPI für Informatik – Saarbrücken, DE
- Lewis Chuang
LMU München, DE
- Andrew Duchowski
Clemson University, US
- Sara Irina Fabrikant
Universität Zürich, CH
- Nina Gehrer
Universität Tübingen, DE
- Hans Gellersen
Lancaster University, GB
- Kenneth Holmqvist
Universität Regensburg, DE
- Eakta Jain
University of Florida – Gainesville, US
- Radu Jianu
City – University of London, GB
- Enkelejda Kasneci
Universität Tübingen, DE
- Peter Kiefer
ETH Zürich, CH
- Krzysztof Krejtz
SWPS University – Warsaw, PL
- Kuno Kurzhals
Universität Stuttgart, DE
- David P. Luebke
NVIDIA – Charlottesville, US
- Radoslaw Mantiuk
West Pomeranian Univ. of Technology – Szczecin, PL
- Diako Mardanbegi
Lancaster University, GB
- Thies Pfeiffer
Universität Bielefeld, DE
- Pernilla Qvarfordt
FX Palo Alto Laboratory, US
- Michael Raschke
Blickshift GmbH – Stuttgart, DE
- Martin Raubal
ETH Zürich, CH
- Laura Trutoiu
Magic Leap – Seattle, US
- Daniel Weiskopf
Universität Stuttgart, DE