Gaze-Aware Mixed-Reality: Addressing Privacy Issues with Eye Tracking

Working Paper

Author(s):
Göbel, Fabian; Kurzhals, Kuno; Raubal, Martin; Schinazi, Victor R.

Publication date:
2020

Permanent link:
https://doi.org/10.3929/ethz-b-000409514

Gaze-Aware Mixed-Reality: Addressing Privacy Issues with Eye Tracking

Fabian Göbel
Institute of Cartography and Geoinformation
ETH Zurich
Zurich, Switzerland
goebelf@ethz.ch

Kuno Kurzhals
Institute of Cartography and Geoinformation
ETH Zurich
Zurich, Switzerland
kunok@ethz.ch

Martin Raubal
Institute of Cartography and Geoinformation
ETH Zurich
Zurich, Switzerland
mraubal@ethz.ch

Victor R. Schinazi
Institute of Cartography and Geoinformation
ETH Zurich
Zurich, Switzerland
scvictor@ethz.ch

CHI'20 Extended Abstracts,, April 25–30, 2020, Honolulu, HI, USA
Proceedings of the 1st Workshop on Exploring Potentially Abusive Ethical, Social and Political Implications of Mixed Reality Research in HCI

Abstract
Current Mixed Reality (MR) systems rely on a variety of sensors (e.g., cameras, eye tracking, GPS) to create immersive experiences. Data collected by these sensors are necessary to generate detailed models of a user and the environment that allow for different interactions with the virtual and the real world. Generally, these data contain sensitive information about the user, objects, and other people that make up the interaction. This is particularly the case for MR systems with eye tracking, because these devices are capable of inferring the identity and cognitive processes related to attention and arousal of a user. The goal of this position paper is to raise awareness on privacy issues that result from aggregating user data from multiple sensors in MR. Specifically, we focus on the challenges that arise from collecting eye tracking data and outline different ways gaze data may contribute to alleviate some of the privacy concerns from aggregating sensor data.

Author Keywords
Mixed Reality; Eye Tracking; Gaze-based Interaction; Privacy.

CCS Concepts
•**Human-centered computing** $\rightarrow$ *User interface design;* **HCI theory, concepts and models; Interaction techniques;**

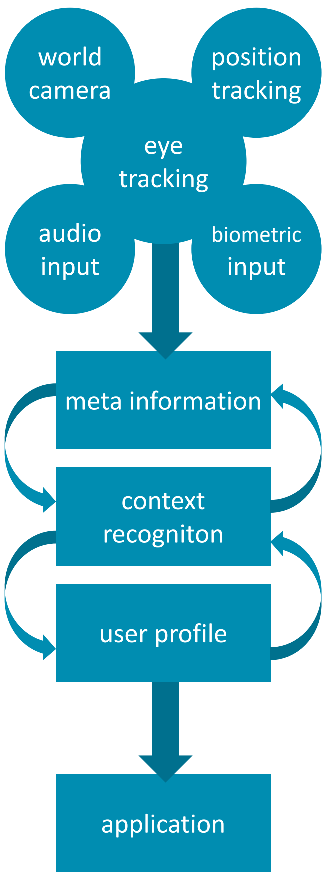

Figure 1: Typical data processing pipeline for a MR device. It is expected, that all listed categories of sensory input will be available in future MR devices.

Introduction

Mixed Reality (MR) devices are becoming an integral part of our daily lives as new interfaces continue to ubiquitously couple users to a variety of sensors that are used to maximize the quality of the MR experience. For example, the latest release of Microsoft's MR headset (Hololens 2[1]) includes eye tracking capabilities that allow applications to measure the user's point of regard (POR) for gaze supported selection and manipulation of virtual content [25]. Eye tracking also allows researchers to measure changes in arousal relative to visual stimuli. Indeed, researchers have successfully utilized gaze features such as pupil dilation, fixation and saccades, or blink rate to predict differences in stress levels [2, 9, 14, 17, 22]. While these technologies will certainly enhance our interactions with the physical and virtual world, they also raise a series of ethical concerns regarding data handling and privacy.

Notably, the collection of eye tracking data in combination with other sensors in MR devices poses additional privacy concerns. Here, the collected data may be combined to produce fine grained information about the user and manipulated to control their behavior. Critically, these privacy issues apply both to the users (individuals wearing the devices) but also to the general public that is being recorded and tracked, often without their consent. In this paper, we present a scenario illustrating the potential associated with multi-sensor MR and highlight how gaze may be used to alleviate some of the privacy and ethical concerns that arise from the aggregation of sensitive data.

Issues with Multi-Sensor MR

Figure 1 illustrates a typical data processing pipeline for sensory input that is necessary to provide the MR experience. We identified five main categories of sensory input that are either already in use or feasible to include in the future: *world camera*, *biometric input*, *audio*, *position tracking* and *eye tracking*. Depending on the scenario, the individual sensors (or a combination of them) provide the data input that is needed to build a detailed profile about the user and the environment. Here, semantic information is derived from the data and enriched by meta information from other sources outside the system (e.g., social media, news, common knowledge) in order to create a comprehensive model of the situation. For example, when fixating on a product in the supermarket, the list of ingredients is automatically downloaded from the internet and can be presented to the user via the MR application. Below we discuss how these sensors are used and outline their potential ethical issues.

[1] https://docs.microsoft.com/en-us/hololens/hololens2-hardware

World Camera: This category comprises all optical devices that capture the user's surroundings as well as the people and the objects contained within it. In order to provide an immersive user experience, these cameras need to constantly scan the environment [21]. This type of passive and "always on" recording raises privacy issues including the identity and actions of the person wearing the MR device and others that are present but are unaware of being recorded.

Biometric Measures: Wearable devices (e.g., Fitbit, Apple Watch) are often used to measure biometric signals such as the heart rate in an effort to help individuals to monitor and improve their health [27]. Beyond providing information about the physical and mental state of a person, these data can also be used for authentication purposes and to infer various types of activities [20]. In MR, data from biometric sensors can be used to provide personalized content by adjusting the MR experience according to body signals and to counteract stress situations via biofeedback. The same data can also be maliciously used to profile and expose individuals and their medical conditions.

Gaze-aware MR scenario

An elderly person is walking down a street with the latest MR device. Suddenly, the biometric sensors of the device report a loss of blood pressure. The eye tracker recognizes a change in the gaze behavior stemming from confusion and visual search. Further analysis of the biometric data (in conjunction with the user profile) allows the system to assess the situation and propose different options (e.g., show locations of a pharmacy, hospital, or doctor) that can be used to help the person. Using a gaze gesture, this person can now initiate a conversation with relatives or the hospital to communicate the emergency. As the location of the person is constantly being tracked, the MR map application can also provide directions to the closest points of interest and the chosen destination. Arriving at the hospital, the MR glasses of the doctor immediately disclose the medical record after eye contact and face recognition with the user confirms the interaction.

Box 1: Example scenario for the application of gaze-aware MR.

Audio Input: Voice assistance in MR is used to offer a natural and hands-free mode for interaction with the system. While voice is used as an explicit input to trigger the action, a microphone needs to be passively listening to identify the activation key word. This circumstance poses privacy issues if the device is unintentionally activated revealing sensitive details about conversations between users. These types of data breaches are becoming more common as evidenced by a recent newspaper report on large amounts of audio data that were being listened to by Amazon workers for manual annotation[2].

Position Tracking: These are the data inputs used to localize the MR users and position them on a map of the environment or a room through GPS and SLAM [7, 16]. While local information is needed to accurately place virtual objects, spatial information beyond the current field of view allows the system to provide solutions such as navigation instructions or weather forecasts. Different privacy issues arise with this type of data [6, 11]. For example, location tracking allows for the exposure of intimate details such as frequently visited places or people's daily habits.

Eye Tracking: Recording the users' eye movements and mapping them to areas or objects of interest within the real or virtual world can be applied for interaction purposes and for behavior analysis. Furthermore, gaze behavior (e.g., fixations, scan paths) can reveal information about ongoing cognitive processes [5, 10] and a user's cognitive load [12] associated with a task. Eye tracking also provides detailed data on changes in pupil diameter and blink rate. Similar to heart rate, dilation and contraction of the pupils are controlled by the sympathetic and parasympathetic nervous systems, respectively. Indeed, researchers have found that variability in pupil diameter may be used as an index for the evaluation of mental states that is comparable to heart rate [1, 23]. The relationship between blinking rates and stress is still not clear. Here, researchers have reported higher [17] and lower [14] frequencies of blinks during stressful situations. Taken together, eye tracking data can be used to infer individual characteristics including age, gender, race, sexual preference, body mass index, hormonal cycle, and health (see [15]). As eye movements occur intentionally and unintentionally, it is difficult for a user to control what to share with the MR system.

[2] https://www.theguardian.com/technology/2019/apr/11/amazon-staff-listen-to-customers-alexa-recordings-report-says

Gaze-aware MR

In the following section, we present a possible mixed reality scenario (see Box 1) to motivate a discussion of the implications of a multi-sensor MR system and how different sensors can be combined with eye tracking in order to remedy some of the aforementioned privacy concerns.

Eye Tracking + World Camera: A common challenge with the world camera recording is that objects that are not needed for the current task are also recorded as part of the interaction. In our scenario, the world camera might possibly record sensitive data such as the people or objects in the hospital. Since gaze is used to target an object before the interaction takes place [13, 24], information of the current POR can be used to filter the world camera video feed. This allows the system to process only task critical information within a small area around the current fixation.

Eye Tracking + Biometric Measures: The collection of biometric data should be limited since it can reveal sensitive details about the user [19] (e.g., medical conditions). Here, eye tracking can help to identify situations in which the collection of biometric data is beneficial and supported by the user. In our scenario, the recording of biometric data could be stopped whenever the user is in close interaction with items from a previously defined blocklist, for example, containing recognized faces or objects.

Eye Tracking + Voice Input: Eye tracking can also be used to avoid privacy issues with voice assistance in MR that requires a microphone to be listening at all times. We propose to use gaze gestures as a trigger to turn the microphone to *listening mode* and thus ensure user's awareness [4]. Besides being distinct from natural gaze behavior and avoiding unintentional triggering, gaze gestures (e.g. blinking) can be detected based on raw gaze data without the need for an active world camera. In our scenario, the microphone is only activated after the user performs the specific gaze-gesture to initiate the call.

Eye Tracking + Position: The collection of eye tracking data by itself can also be sensitive. In the presented scenario, such data collection should be deactivated once the user enters the treatment room in the hospital. Here, geofencing can be used for defining restrictive areas where gaze data will not be collected [18]. While this is often applied in outdoor scenarios, marker-based approaches can also be used for geo-fencing in indoor environments [3].

Conclusion

In this position paper, we presented the benefits and privacy risks of the most common sensors used in MR. We highlighted some of the ethical and privacy challenges associated with eye tracking in MR and discussed the manner in which gaze and other sensor data can be used to alleviate some of these issues. Specifically, we showed that the combination of the world camera with gaze can avoid recording objects irrelevant to the task. Similarly, when the microphone is combined with gaze as an explicit trigger, constant listening for a keyword is not necessary to activate the voice assistant. We also demonstrated how gaze can help restrict the collection of sensitive biometric data and the way that location can be used to block gaze recording.

While we focused on a single MR user, the raised issues become increasingly challenging in multi-user scenarios [8]. Aggregating gaze data from multiple individuals allows for the creation of detailed models of visual attention [26] that can be used to detect non-conforming behavior and manipulate intent. For this reason, it is critical to develop a set of guidelines that can be used to control the collection of data and protect the privacy rights of users and the general public. Bar-Zeev, one of the inventors of the first HoloLens, outlines a series of policies with respect to eye tracking[3] that can also be applied in the case of MR. These include, (1) the notion that MR sensor data is highly sensitive and should be treated with the same privacy protocols as health data, (2) the strict regulation of data streaming, (3) the need for transparency with regard to the profiling that is taking place, and (4) the assurance that the collected and processed data are used for the benefit and not the exploitation of the user.

As MR technology becomes more integrated with our daily lives, it is our duty as researchers and developers to consider the ethical consequences of collecting and processing data from multiple sensors for MR applications. The combination of gaze with different sensory data provides an integrative and promising way to address some of these privacy concerns by making use of sensors to influence data collection before processing takes place.

[3]https://www.vice.com/en_us/article/bj9ygv/the-eyes-are-the-prize-eye-tracking-technology-is-advertisings-holy-grail

REFERENCES

[1] Ane Alberdi, Asier Aztiria, and Adrian Basarab. 2016. Towards an Automatic Early Stress Recognition System for Office Environments Based on Multimodal Measurements: A Review. *Journal of Biomedical Informatics* 59 (2016), 49–75.

[2] Armando Barreto, Jing Zhai, Naphtali Rishe, and Ying Gao. 2007. Measurement of Pupil Diameter Variations As a Physiological Indicator of the Affective State in a Computer User. *Biomedical Sciences Instrumentation* 43 (2007), 146–151.

[3] Dominik Bucher, David Rudi, and René Buffat. 2018. Captcha Your Location Proof - a Novel Method for Passive Location Proofs in Adversarial Environments. In *Progress in Location Based Services*, Peter Kiefer, Haosheng Huang, Nico Van de Weghe, and Martin Raubal (Eds.). 269–291.

[4] Heiko Drewes and Albrecht Schmidt. 2007. Interacting with the Computer Using Gaze Gestures. In *Human-Computer Interaction*, Abascal J. Barbosa S.D.J. Baranauskas C., Palanque P. (Ed.). Vol. 4663. Springer, Berlin, Heidelberg, 475–488.

[5] Andrew T. Duchowski, Krzysztof Krejtz, Izabela Krejtz, Cezary Biele, Anna Niedzielska, Peter Kiefer, Martin Raubal, and Ioannis Giannopoulos. 2018. The Index of Pupillary Activity: Measuring Cognitive Load Vis-à-vis Task Difficulty with Pupil Oscillation. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*.

[6] Matt Duckham and Lars Kulik. 2006. Location Privacy and Location-aware Computing. *Dynamic and Mobile GIS* 3 (2006), 63–80.

[7] Paolo Fogliaroni, Bartosz Mazurkiewicz, Markus Kattenbeck, and Ioannis Giannopoulos. 2019. Geographic-aware Augmented Reality for Vgi. *Advances in Cartography and GIScience of the ICA* 2 (2019), 1–9.

[8] Fabian Göbel, Tiffany C. K. Kwok, and David Rudi. 2019. Look There! Be Social and Share. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems: Workshop on Challenges Using Head-Mounted Displays in Shared and Social Spaces*. 1–6.

[9] Nadja Herten, Tobias Otto, and Oliver T. Wolf. 2017. The Role of Eye Fixation in Memory Enhancement under Stress - an Eye Tracking Study. *Neurobiology of Learning and Memory* 140 (2017), 134–144.

[10] Marcel A. Just and Patricia A. Carpenter. 1976. Eye Fixations and Cognitive Processes. *Cognitive Psychology* 8, 4 (1976), 441–480.

[11] Carsten Keßler and Grant McKenzie. 2018. A Geoprivacy Manifesto. *Transaction in GIS* 22, 1 (2018), 3–19.

[12] Peter Kiefer, Ioannis Giannopoulos, Andrew Duchowski, and Martin Raubal. 2016. Measuring Cognitive Load for Map Tasks Through Pupil Diameter. In *Proc. of the Int. Conf. on Geographic Information Science*. 323–337.

[13] Michael Land, Neil Mennie, and Jennifer Rusted. 1999. The Roles of Vision and Eye Movements in the Control of Activities of Daily Living. *Perception* 28, 11 (1999), 1311–1328.

[14] Wenhui Liao, Weihong Zhang, Zhiwei Zhu, and Qiang Ji. 2005. A Real-time Human Stress Monitoring System Using Dynamic Bayesian Network. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition - Workshops*. 1–8.

[15] Daniel J. Liebling and Sören Preibusch. 2014. Privacy Considerations for a Pervasive Eye Tracking World. In *Proc. of the ACM International Joint Conf. on Pervasive and Ubiquitous Computing: Adjunct Publication*. 1169–1177.

[16] R. Mautz and S. Tilch. 2011. Survey of Optical Indoor Positioning Systems. In *Proc. of the Int. Conf. on Indoor Positioning and Indoor Navigation*. 1–7.

[17] Mohd Norzali Haji Mohd, Masayuki Kashima, Kiminori Sato, and Mutsumi Watanabe. 2014. Facial Visual-infrared Stereo Vision Fusion Measurement As an Alternative for Physiological Measurement. *Journal of Biomedical Image Processing* 1, 1 (2014), 34–44.

[18] Kaylin T. Nguyen, Jeffrey E. Olgin, Mark J. Pletcher, Madelena Ng, Leanne Kaye, Sai Moturu, Rachel A. Gladstone, Chaitanya Malladi, Amy H. Fann, Carol Maguire, Laura Bettencourt, Matthew A. Christensen, and Gregory M. Marcus. 2017. Smartphone-based Geofencing to Ascertain Hospitalizations. *Circulation. Cardiovascular Quality and Outcomes.* 10, 3 (2017).

[19] Greig Paul and James Irvine. 2014. Privacy Implications of Wearable Health Devices. In *Proc. of the Int. Conf. on Security of Information and Networks*. 117–121.

[20] Ken Pfeuffer, Matthias J. Geiger, Sarah Prange, Lukas Mecke, Daniel Buschek, and Florian Alt. 2019. Behavioural Biometrics in VR: Identifying People from Body Motion and Relations in Virtual Reality. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*. 1–12.

[21] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. 2016. You Only Look Once: Unified, Real-time Object Detection. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 779–788.

[22] Peng Ren, Armando Barreto, Jian Huang, Ying Gao, Francisco R. Ortega, and Malek Adjouadi. 2014. Off-line and On-line Stress Detection through Processing of the Pupil Diameter Signal. *Annals of Biomedical Engineering* 42, 1 (2014), 162–176.

[23] Kiyomi Sakamoto, Shoichi Aoyama, Shigeo Asahara, Haruki Mizushina, and Hirohiko Kaneko. 2009. Relationship between Emotional State and Pupil Diameter Variability under Various Types of Workload Stress. In *Ergonomics and Health Aspects of Work with Computers*, Karsh Ben-Tzion (Ed.). Vol. 5624. Springer, Berlin, Heidelberg, 177–185.

[24] Linda E. Sibert and Robert J. K. Jacob. 2000. Evaluation of Eye Gaze Interaction. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*. 281–288.

[25] Sophie Stellmach and Raimund Dachselt. 2012. Look & Touch: Gaze-supported Target Acquisition. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*. 2981–2990.

[26] Yusuke Sugano, Xucong Zhang, and Andreas Bulling. 2016. Aggregaze: Collective Estimation of Audience Attention on Public Displays. In *Proc. of the Annual Symposium on User Interface Software and Technology*. 821–831.

[27] Raphael P. Weibel, Jascha Grübel, Hantao Zhao, Tyler Thrash, Dario Meloni, Christoph Hölscher, and Victor R. Schinazi. 2018. Virtual Reality Experiments with Physiological Measures. *Journal of Visualized Experiments* 138 (2018), 1–8.