What Are We Missing? Adding Eye-Tracking to the HoloLens to Improve Gaze Estimation Accuracy

Hidde van der Meulen
University College Dublin
Dublin, Ireland
hidde.vandermeulen@ucdconnect.ie

Andrew L. Kun
University of New Hampshire
Durham, NH, United States
andrew.kun@unh.edu

Orit Shaer
Wellesley College
Wellesley, MA, United States
oshaer@wellesley.edu

ISS '17, October 17–20, 2017, Brighton, United Kingdom

ACM ISBN 978-1-4503-4691-7/17/10.
https://doi.org/10.1145/3132272.3132278

Abstract
The Microsoft HoloLens keeps track of its location and rotation relative to the environment but lacks the ability to capture eye gaze data. We assess a novel method to extend the HoloLens with a head mounted eye-tracker. Using a combination of eye gaze data and head rotation we compared gaze behavior between real and virtual objects. Results indicate that eye-tracking plays an important role in accurately determining a user's gaze for real objects in contrast to virtual objects.

Author Keywords
Eye-tracking; Augmented Reality

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

Introduction
The Microsoft HoloLens is a wearable mixed reality display, which allows users to interact using gesture and voice commands with digital content, which is overlaid onto the real world. The device adjusts projections of virtual 3d objects based on its position in the environment and user's head rotations. This allows users to place a 3d objects in space and walk around it

as if it is a real object. Interaction with virtual objects is possible using a virtual cursor, a small circle that is always centered in front of the user. Users can aim the cursor by changing their physical position and head rotations. Custom apps can log and export the HoloLens's location and rotation to estimate where participants look while wearing the device. Since the HoloLens does not provide eye-tracking, the available data is limited to a user's head position. This means that there is an uncertainty and possible inaccuracy in the data if users move their eyes away from the center. People move their eyes while exploring a scene with an average saccade size is between 4 and 5 degrees during scene perception [8].

Modern inexpensive eye-trackers can be applied in varying settings such as large touchscreen tables, driver simulators and more recently in combination with AR devices [4,6]. We combined a Pupil Lab eye-tracker with a HoloLens device to study differences in gaze prediction between real and virtual objects (Figure 1).

Figure 1. The Microsoft HoloLens with an attached Pupil Labs high speed binocular eye-tracker.

Related work

Other efforts have also explored the use eye tracking to better understand the relationship between user gaze when using AR devices, and various outcomes, such as performance on a task. For example, in the automotive domain, Medenica et al. [5] simulated AR in a driving simulator, and found that an AR-based navigation system allows drivers to keep their eyes on the road more than two other systems that used an LCD display. Kim and Dey [2] also found that a (simulated) AR-based navigation system reduced the time drivers looked away from the road, compared to a head-down display. Bolton et al. [1] also used a head-up display to implement AR navigation information and used eye tracking to assess the time drivers spent looking at different areas of interest, as well as to identify common gaze sequences. In contrast to the work we present here, these previous efforts used simulated or real head-up displays to implement augmented reality. However, a head-up implementation of AR requires deploying the technology in the user's infrastructure. In contrast, our focus is on head-worn AR glasses, which can display AR information without deploying technology in the user's infrastructure. Note that we will need head-worn eye trackers to assess users' visual behaviors when they wear AR glasses.

Renner and Pfeiffer [9] also explored the use of head-worn eye tracking in an application that is intended to employ AR glasses. However, their experiment simulated AR glasses using a virtual reality (VR) headset. And while their approach allows for careful control of the experiment (since they can control all aspects of the experiment in VR), there is also a need for understanding how users will interact with real AR glasses – this is what our work is aimed at.

Experiment

We conducted an experiment with a 2x2x2 mixed design (gaze location calculation x target type x path). In calculating the gaze location, we compared head rotations to a combination of head rotations and eye-tracking. We had two types of trials: either with 4 virtual or 4 paper (real) targets. Gaze location calculation and target type were within-subjects variables. Participants were instructed to walk one of two predetermined paths in front of the curtain while looking at the 4 targets in a predefined order, see Figure 2 – the path was a between-subjects variable.

We collected data from 20 participants (all undergraduate students, 12 female) and dismissed data of 8 participants based on low eye-tracker confidence (< 80%). We report on n=12 (7 female). The low confidence was caused by a suboptimal angle of the pupil cameras under the HoloLens.

Participants completed a practice trial and two (counterbalanced) experimental trials with virtual and real targets. Participants heard the sequence of walking and target instructions with a new instruction every 4 seconds. Participants heard either a direction in which to take a small step or a target number (1-4) to look at. The total sequence took 2 minutes per trial and gave participants an equal amount of time to look at each target.

Setup

We used a modified Pupil Labs eye-tracker attached to the HoloLens (Figure 1). To combine head rotations and eye-tracker angles, we need to calibrate the eye-tracker and synchronize recording. A custom HoloLens application that listens to socket commands and a networked computer allow us to simultaneously send commands to the HoloLens and eye-tracker software. A calibration command will let the HoloLens show markers at predetermined locations within its display, participants are instructed to look at the markers. The locations and timings of the calibration markers are sent to the eye-tracker software which then correlates the recorded pupil positions. The correlations allow us to estimate gaze points in real time using recorded pupil positions.

Using the same protocol, we can initiate simultaneous data recording on the HoloLens and eye-tracker. When this recording is started, the instruction sequence for gaze targets and stepping sequence starts playing to the participants.

Data processing and results

After collecting all data, we interpolated eye-tracker data for data points with a confidence lower than 20% (i.e. blinks). Then we calculated the intersections of gaze on the wall for each point in time using the physical location (which changes as participants make steps as instructed by the sequence) and rotations (which changes as participants look at different targets during the sequence).

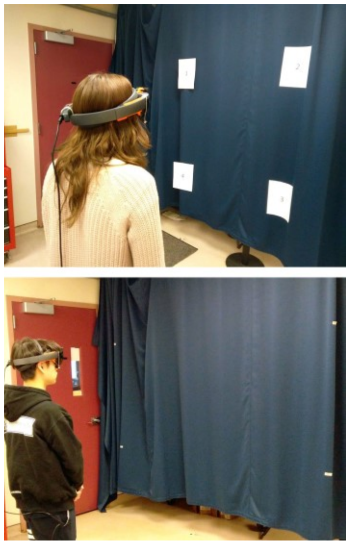

Figure 2. Participants wearing the HoloLens and eye-tracker looking at the numbered and identically positioned real (top) and virtual (bottom) targets. Note that the virtual targets are only visible for the participant using the HoloLens.

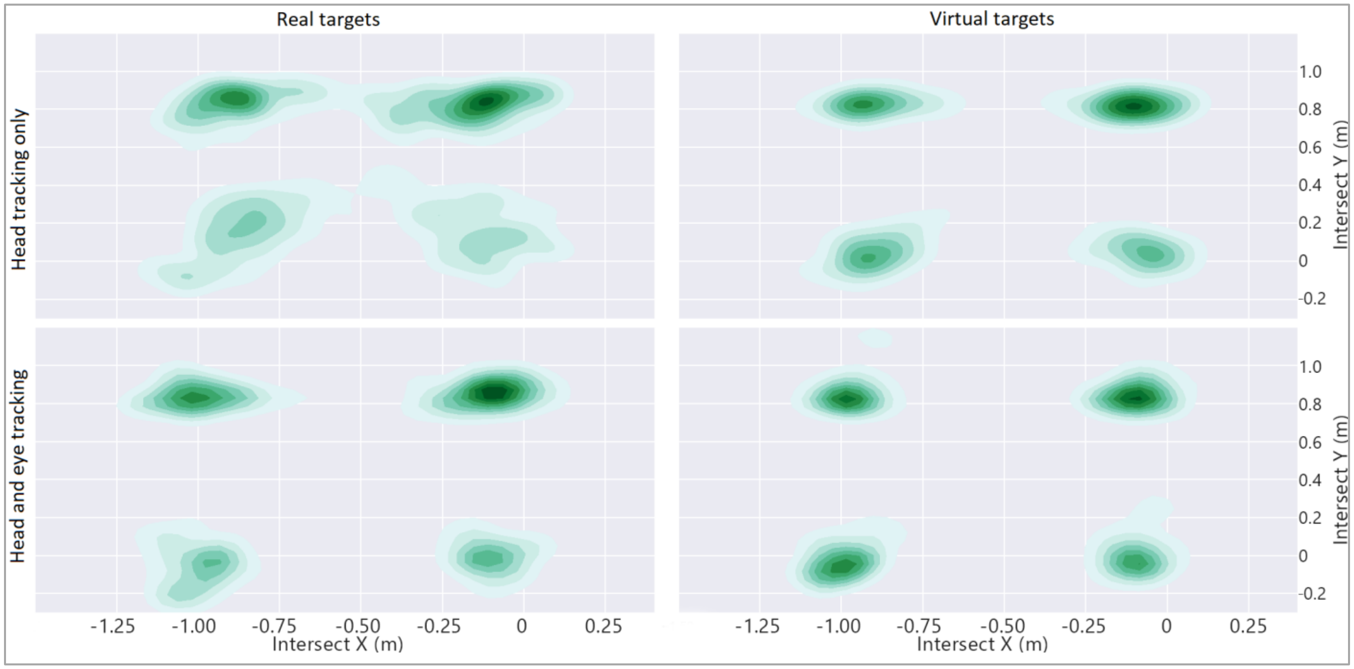

Figure 3. Heatmaps showing the cumulative estimated gaze locations on the wall (n=12). The X-axis shows the calculated gaze estimates of all participants in the horizontal space and the Y-axis the estimates in the vertical space

The intersections were estimated using just head-rotations or a combination of head and eye-rotations for all trials with virtual targets and trials with real targets. Figure 3 shows the combined gaze estimates of all participants looking at the 4 targets. The resulting heatmap seems to be strongly influenced by the addition of eye tracking data with more concentrated estimates around the target locations as a result. The eye-tracker data seems to have less effect on the estimation for virtual targets.

Discussion and future work

We believe that the difference of gaze behavior between real and virtual targets is caused by the narrow field of view that the HoloLens offers. This result is consistent with previous work [4]. To see a virtual target within the HoloLens display, participants need to aim their head towards the target while for a real target a combination of eye and head-movements will work well. Thus, for virtual targets, gaze location can be estimated using only head rotation while for real targets the addition of gaze angle is important.

In future work, we will develop a tool to work with real-time gaze data using multiple HoloLens devices and eye-trackers. We will use the tool in our work to enhance museum experiences [7], and to transform automated vehicles into places of work and play [3].

Our work complements the work on AR that used simulated and real head-up displays [1,2,5], because it allows us to start to explore visual behaviors that we can expect to see with AR glasses which have a limited field of view.

Limitations

All participants were students with a technical background and they had no prior experience in using the HoloLens. Both factors could have influenced the results, experiments with a broader group and longer training could strengthen the findings. We also had to discard data of 8 participants because for them the eye tracker was positioned too low to get a good-enough image of their eyes, resulting in a relatively low amount of data. Future experiments could investigate ways to improve the confidence of the eye-tracking.

Another limitation is that we tested the technology in a controlled setting with little freedom of movement for participants and two simple target types. We will need to explore the combination of head tracking and eye-tracking in more natural use cases of the Microsoft HoloLens such as advanced 3d holograms and a greater freedom of movement.

Acknowledgments

This work was supported in part by the Broadband Center of Excellence at the University of New Hampshire.

References

[1] Bolton, A., Burnett, G. and Large D.R. An investigation of augmented reality presentations of landmark-based navigation using a head-up display. *Proc. AutomotiveUI '15* (2015), 56-63.

[2] Kim, S. and Dey, A.K. Simulated augmented reality windshield display as a cognitive mapping aid for elder driver navigation. *Proc. CHI '09* (2009), 133-142.

[3] Kun, A.L., Boll, S. and Schmidt, A. Shifting Gears: User Interfaces in the Age of Autonomous Driving. *IEEE Pervasive Computing 15*, 1 (2016), 32-38.

[4] Kun, A.L., van der Meulen, H. and Janssen, C.P. Calling While Driving: An Initial Experiment with HoloLens. *Driving Assessment (2017).*

[5] Medenica, Z, Kun, A.L., Paek, T. and Palinko O. Augmented reality vs. street views: a driving simulator study comparing two emerging navigation aids. *Proc. MobileHCI '11* (2011), 265-274.

[6] van der Meulen, H., Varsanyi, P., Westendorf, L., Kun, A. and Shaer, O. Towards Understanding Collaboration around Interactive Surfaces. *Proc. User Interface Software and Technology* (2016), 219-220.

[7] Pollalis, C., Fahnbulleh, W., Tynes, J. and Shaer, O. HoloMuse: Enhancing Engagement with Archaeological Artifacts through Gesture-Based Interaction with Holograms. *Proc. Tangible, Embedded, and Embodied Interaction* (2017), 565-570.

[8] Rayner, K. Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology 62*, 8 (2009), 1457-1506.

[9] Renner, P. and Pfeiffer, T. Attention guiding techniques using peripheral vision and eye tracking for feedback in augmented-reality-based assistance systems. *3D User Interfaces* (2017), 186-194