Benefits of Locating Overt Visual Attention in Space Using Binocular Eye Tracking for Mixed Reality Applications

Thies Pfeiffer and Nikita Mattar

{tpfeiffe, nmattar}@techfak.uni-bielefeld.de
A.I. Group, Faculty of Technology, Bielefeld University, Germany

Abstract: The "Where?" is quite important for Mixed Reality applications: Where is the user looking at? Where should augmentations be displayed? The location of the overt visual attention of the user can be used both to disambiguate referent objects and to inform an intelligent view management of the user interface. While the vertical and horizontal orientation of attention is quite commonly used, e.g. derived from the orientation of the head, only knowledge about the distance allows for an intrinsic measurement of the location of the attention. This contribution reviews our latest results on detecting the location of attention in 3D space using binocular eye tracking.

1 Motivation and Related Work

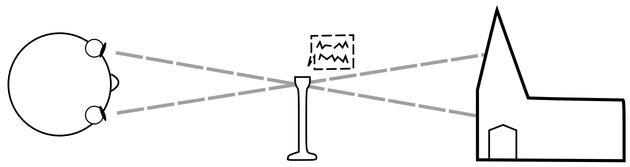

Figure 1: Binocular eye tracking helps disambiguating between partly occluding fixation targets.

Where is the user looking at? An exact knowledge about the current location of the user's overt visual attention is essential for most Mixed Reality (MR) applications. The system has to disambiguate between attention targeted at relevant objects and those that are either irrelevant or, moreover, the system is not even aware of. This is especially a problem in dynamic environments such as tourist scenarios with a high frequency of occlusions, e.g. by other people. From the perspective of the MR application, this is a foreground (known objects) vs. background (unknown objects) problem. Knowledge about the depth of the overt visual attention can help to disambiguate between partly occluded objects and thus improves the robustness of the system (see Figure 1).

Information about the overt visual attention also helps to focus on certain points of interest and thus reduces the amount of information to display. A coarse approximation of the direction of attention can be derived from head movements (e.g. see [BHF02]).

Where to display augmentations? Intelligent view management, i.e. the spatial organisation of the MR user interface, is crucial especially in high demanding scenarios like vehicle control. This is getting increased attention with Heads-Up Displays (HUD) finding their way from special purpose solutions for aviation into consumer automotive settings. However, the high frequent data delivered by advanced sensors has to be channeled, e.g. by intelligent view management, to be turned into helpful information. Thus, an important question is where to present these information without distracting the operator and yet to catch his attention. Information should, e.g., be displayed directly within the scene rather than be presented on a fixed plane [MF96]. This is only one approach to reduce typical problems related to HUD displays such as cluttering, cognitive tunneling and time needed for the attention switch between the HUD and the external environment.

The location of the user's current overt visual attention could be used to decide where to place user interface elements: either inside or outside the user's focus, depending on their relevance. It can also be used to filter object-linked information with a distance-based level of detail depending on where the user is looking at. One approach to determine the current overt visual attention is based on binocular eye tracking, which provides depth information in addition to the horizontal and vertical orientation derived from monocular eye tracking.

2 Locating Overt Visual Attention with Binocular Eye Tracking

Recently we evaluated binocular eye tracking systems and tested algorithms to estimate the depth of a fixation [PLW09] in virtual and real scenarios in desktop-based settings. We also demonstrated the application of 3D fixation estimation in an object selection task. In our settings with objects located in the personal perimeter ($< 1m$), we achieved a mean accuracy in depth estimation of $-1.9cm$ with a precision (standard deviation) of $9.7cm$.

In a second study, we moved from a stationary desktop-based setting to a fully interactive immersive setting in a CAVE-like environment. This setting combines eye tracking with optical marker based body tracking. For this we developed an automatic single-user calibration procedure with 3D calibration targets [Pfe08]. In the same publication we evaluated precision and accuracy of the system in a monocular fixation task. We measured a horizontal accuracy of $1.18\,^{\circ}$ (sd $1.51\,^{\circ}$) and a vertical accuracy of $2.52\,^{\circ}$ (sd $2.24\,^{\circ}$). The system has been used for gaze-based selection of 3D buildings on a city map, for a demonstration see our video hosted on YouTube [PLW07].

To measure the application-level latency, we developed a visual ping task: Initially the user is fixating a given target. Then a new target is presented and the time is taken until saccades towards this target exceed $2.5\,^{\circ}$. In this task our setup achieved a mean latency of $307.9\,ms$ (sd $99.9\,ms$). This includes both system and human processes. The system's latency alone was estimated to be about $70\,ms$.

3 Discussion

The resolution we achieved in the estimation of the fixation depth allows for a stable disambiguation of foreground vs. background in common scenarios. It is thus also suited to provide essential information for an intelligent view management.

In the praxis, however, there are still some challenges. First, the eye tracking system we used requires a stationary processing unit and is thus not suited for mobile operation. However, currently much progress can be observed in miniaturizing eye tracking devices and mobile solutions have already been presented. In addition, detection accuracy has to be constantly monitored to detect and compensate drifts. This could require a re-adjustment of the headgear and repeated calibration procedures.

The distance from the user, the depth, is an important factor, not least because of the physiology of the human eye, which is faster in adapting smooth pursuits within the same distance than switching between distances. This knowledge can be used to improve the foreground vs. background disambiguation needed in Mixed Reality applications as well as the view management for the user interface. Visualizations adapting to the user's attention include manipulations of the transparency, adaptive blurring, level of detail and scene-linked placements. Moreover, knowledge about the location of overt visual attention in space could also be used for gaze-based interaction as well.

This work has been partially funded by the German Research Foundation within the Collaborative Research Center 673 *Alignment in Communication*.

References

[BHF02] B. Bell, T. Höllerer, and S. Feiner. An annotated situation-awareness aid for augmented reality. In *Proceedings of the 15th annual ACM symposium on User interface software and technology*, pages 213–216. ACM New York, NY, USA, 2002.

[MF96] R.S. McCann and D.C. Foyle. Scene-linked symbology to improve situation awareness. In *AGARD CONFERENCE PROCEEDINGS AGARD CP*, pages 16–16. AGARD, 1996.

[Pfe08] Thies Pfeiffer. Towards Gaze Interaction in Immersive Virtual Reality: Evaluation of a Monocular Eye Tracking Set-Up. In Marco Schumann and Torsten Kuhlen, editors, *Virtuelle und Erweiterte Realität - Fünfter Workshop der GI-Fachgruppe VR/AR*, pages 81–92, Aachen, 2008. Shaker Verlag GmbH.

[PLW07] Thies Pfeiffer, Marc E. Latoschik, and Ipke Wachsmuth. Gaze Interaction in Immersive Virtual Reality - 3D Eye Tracking in Virtual Worlds. YouTube video, http://www.youtube.com/watch?v=Zx0xAmJtTyc, 2007.

[PLW09] Thies Pfeiffer, Marc Erich Latoschik, and Ipke Wachsmuth. Evaluation of Binocular Eye Trackers and Algorithms for 3D Gaze Interaction in Virtual Reality Environments. *Journal of Virtual Reality and Broadcasting*, 5.2008(16), jan 2009.