Towards a Symbiotic Human-Machine Depth Sensor: Exploring 3D Gaze for Object Reconstruction

Teresa Hirzle[1]**, Jan Gugenheimer**[1]**, Florian Geiselhart**[1]**, Andreas Bulling**[2]**, Enrico Rukzio**[1]
[1]Institute of Media Informatics, Ulm University, Germany
[2]Institute for Visualisation and Interactive Systems, University of Stuttgart, Germany
[1]firstname.lastname@uni-ulm.de, [2]andreas.bulling@vis.uni-stuttgart.de

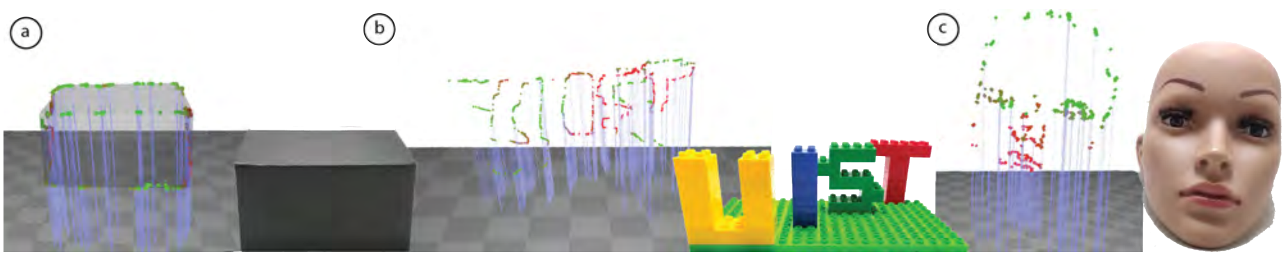

Figure 1. 3D Gaze points from one author looking at three different objects rendered in Cinema4D: (a) black box, (b) letters build in Lego bricks and (c) head of a mannequin

ABSTRACT
Eye tracking is expected to become an integral part of future augmented reality (AR) head-mounted displays (HMDs) given that it can easily be integrated into existing hardware and provides a versatile interaction modality. To augment objects in the real world, AR HMDs require a three-dimensional understanding of the scene, which is currently solved using depth cameras. In this work we aim to explore how 3D gaze data can be used to enhance scene understanding for AR HMDs by envisioning a symbiotic human-machine depth camera, fusing depth data with 3D gaze information. We present a first proof of concept, exploring to what extend we are able to recognise what a user is looking at by plotting 3D gaze data. To measure 3D gaze, we implemented a vergence-based algorithm and built an eye tracking setup consisting of a Pupil Labs headset and an OptiTrack motion capture system, allowing us to measure 3D gaze inside a 50x50x50 cm volume. We show first 3D gaze plots of "gazed-at" objects and describe our vision of a symbiotic human-machine depth camera that combines a depth camera and human 3D gaze information.

CCS Concepts
•**Human-centered computing** → **Human computer interaction (HCI);** *Mixed / augmented reality;* Interaction paradigms;

UIST'18 Adjunct October 14–17, 2018, Berlin, Germany

ACM ISBN 978-1-4503-5949-8/18/10. . . $15.00
DOI: https://doi.org/10.1145/3266037.3266119

Author Keywords
3D gaze; eye-based interaction; human-machine symbiosis

INTRODUCTION
Upcoming augmented reality (AR) head-mounted displays (HMDs) are expected to have integrated eye tracking to be able to understand human activity [1] and provide an additional interaction modality (e.g. pointing [6], explicit interaction [5], indirect selection [11]). In contrast to prior device types that focused on 2D displays, AR HMDs aim to augment the physical environment around the user and therefore also need 3D spatial information. This is currently obtained using depth cameras and SLAM algorithms (e.g. Microsoft HoloLens), which provide good results but still struggle with some specific scenarios (e.g. mirrors/windows, dynamic objects, near objects). To overcome these issues, we aim to leverage eye tracking by using 3D gaze estimation as an additional "human depth sensor". This provides a more sparse point cloud (see Fig. 1), but takes advantage of the flexibility of the human eye (e.g. focusing on close distances).

Several approaches have been developed to measure 3D gaze (e.g. in the real world [13, 3] or in AR [10]). One of the main limitations these algorithms have is that gaze depth information only provides accurate measurements of a user's 3D gaze point up to approximately 0.5 to 1 meter distance (e.g. [7], [9]). In contrast to seeing this as a limitation of 3D gaze, in this work we aim to leverage this effect by treating the human eye as a close range depth sensor.

A similar approach to gain information about a physical object through gaze depth was presented by Leelaswassuk et al. [8]. However, the main focus of their work was to use 3D point-of-regard information in combination with a depth map

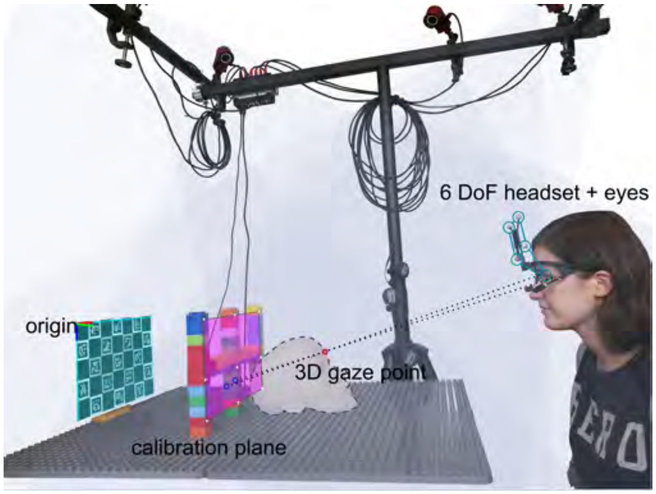

Figure 2. The technical setup to measure 3D gaze inside the tracking volume of approximately 50x50x50cm.

of the environment to allow users hands-free object modelling. Further Vidal et al. [12] demonstrated that gaze can successfully be used to enhance a technical system by recognising if a person looks at augmented or real content.

To explore the feasibility of our approach, we implemented a vergence-based algorithm and built an eye tracking setup consisting of a head-mounted Pupil Labs eye tracker [4] and an OptiTrack motion capture system, allowing us to measure 3D gaze inside an approx. 50x50x50 cm volume. Additionally, we present first 3D gaze visualisations from three different kinds of objects (see Fig. 1). We further discuss our results and future plans to leverage human depth perception by integrating our approach into a symbiotic human-machine depth camera, consisting of an actual depth sensor for far distances enriched with human 3D gaze data for close distances.

3D GAZE SYSTEM

To be able to measure and asses the quality of 3D gaze data we built an initial setup consisting of a head-mounted Pupil Labs eye tracker and a motion capture system (OptiTrack). This allows us to track position and orientation of the headset and the user's eyes inside a 50x50x50 cm volume. To get a more precise measurement of the world camera from the eye tracker, we use an additional tracked ChArUco board (see Fig. 2). All these coordinate systems are fused together inside of a Unity application enabling to visualise the current 3D gaze position in real time.

Our 3D gaze algorithm is based on Wang et al.'s [13] implementation that uses a gaze point triangulation approach. Our algorithm works as follows: We first calibrate the eye tracker with a 9-point calibration on a plane in the real world (calibration plane in Fig. 2). We then project the estimated 2D gaze points (given by the eye tracker) for each eye onto that calibration plane to obtain both corresponding 2D gaze points in the real-world in 3D coordinates. We then use gaze point triangulation to calculate the 3D gaze point of the user, by casting two rays through the user's eyes and corresponding 2D gaze points in 3D coordinates.

RESULTS OF FIRST GAZE-SCANS

To create the gaze-scans in figure 1 we recorded gaze data of one author, who scanned the objects with her eyes, i.e. consciously looking at the objects' outlines and main features at a distance of approx. 30cm. For a proof of concept, we tested our approach with three objects, differing in geometric complexity and contained depth cues. For the first test we used a simple three dimensional geometric form (black box) to get an impression of the feasibility of our approach (Fig. 1 (a)). Based on these results we extended the scan using simple geometric objects placed at different depth levels (Fig. 1 (b)). In a third step we tested the algorithm with an organic object that includes several depth cues itself (Fig. 1 (c)).

SYMBIOTIC HUMAN-MACHINE DEPTH SENSOR

Our final goal with this work is not to generate point clouds that are better (or even at the level) of current or future depth cameras. Instead, we strive to explore how much we can learn about a physical object a user is looking at by observing gaze depth and how we can use this information to enhance already existing depth cameras that are expected to also be part of most future HMDs. We learned from our initial experiments that using 3D gaze information we get a rather sparse "gaze point cloud" compared to current depth cameras. This is mainly because our visual system perceives the environment not only in the centre of the visual field (where we measure the gaze point), but also in the visual periphery, which we do not measure with gaze point estimation [2]. However, the human eye has certain advantages that current depth cameras still struggle with (e.g. focusing on close distances).

We envision a symbiotic scenario extending current technology with "human sensing" data, where a depth camera is able to create a rough understanding of a static environment and gaze depth is merged into the model (e.g. difficult surfaces, close distances). Since our eyes perform an unique eye movement when following moving objects (smooth pursuits [2]), eye data could additionally be used to identify dynamic objects in the scene. We argue that this fusion between human abilities and physical sensors can potentially leverage each individual advantages to overcome current technical difficulties and is also a start to explore future human-machine symbiotic sensors.

CONCLUSION

In this work we presented the vision of a symbiotic human-machine depth sensor that combines advantages of an actual depth sensor with those of human 3D gaze data. We presented a first proof of concept implementation to measure 3D gaze in a 50x50x50 cm volume and explored to what extend we are able to recognise simple objects a user is looking at. In the future we are planing to fuse this information with depth cameras and quantify performance improvements of the environmental depth map.

REFERENCES

1. A. Bulling, J. A. Ward, H. Gellersen, and G. Troster. 2011. Eye Movement Analysis for Activity Recognition Using Electrooculography. *IEEE Transactions on Pattern*

Analysis and Machine Intelligence 33, 4 (April 2011), 741–753. DOI:http://dx.doi.org/10.1109/TPAMI.2010.86

2. Andrew T Duchowski. 2007. Eye tracking methodology. *Theory and practice* 328 (2007).
3. Carlos Elmadjian, Pushkar Shukla, Antonio Diaz Tula, and Carlos H Morimoto. 2018. 3D gaze estimation in the scene volume with a head-mounted eye tracker. In *Proceedings of the Workshop on Communication by Gaze Interaction*. ACM, 3.
4. Moritz Kassner, William Patera, and Andreas Bulling. 2014. Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction. In *Adj. Proc. of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*. 1151–1160. DOI: http://dx.doi.org/10.1145/2638728.2641695
5. Dominik Kirst and Andreas Bulling. 2016. On the Verge: Voluntary Convergences for Accurate and Precise Timing of Gaze Input. In *Ext. Abstr. of the 34th ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*. 1519–1525. DOI: http://dx.doi.org/10.1145/2851581.2892307
6. Mikko Kytö, Barrett Ens, Thammathip Piumsomboon, Gun A Lee, and Mark Billinghurst. 2018. Pinpointing: Precise Head-and Eye-Based Target Selection for Augmented Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 81.
7. Ji Woo Lee, Chul Woo Cho, Kwang Yong Shin, Eui Chul Lee, and Kang Ryoung Park. 2012. 3D gaze tracking method using Purkinje images on eye optical model and pupil. *Optics and Lasers in Engineering* 50, 5 (2012), 736–751.
8. Teesid Leelasawassuk and Walterio W. Mayol-Cuevas. 2013. 3D from Looking: Using Wearable Gaze Tracking for Hands-free and Feedback-free Object Modelling. In *Proceedings of the 2013 International Symposium on Wearable Computers (ISWC '13)*. ACM, New York, NY, USA, 105–112. DOI: http://dx.doi.org/10.1145/2493988.2494327
9. Esteban Gutierrez Mlot, Hamed Bahmani, Siegfried Wahl, and Enkelejda Kasneci. 2016. 3D Gaze Estimation using Eye Vergence.. In *HEALTHINF*. 125–131.
10. Thies Pfeiffer and Patrick Renner. 2014. EyeSee3D: a low-cost approach for analyzing mobile 3D eye tracking data using computer vision and augmented reality technology. In *Proceedings of the Symposium on Eye Tracking Research and Applications*. ACM, 369–376.
11. Ken Pfeuffer, Benedikt Mayer, Diako Mardanbegi, and Hans Gellersen. 2017. Gaze+ pinch interaction in virtual reality. In *Proceedings of the 5th Symposium on Spatial User Interaction*. ACM, 99–108.
12. Mélodie Vidal, David H Nguyen, and Kent Lyons. 2014. Looking at or through?: using eye tracking to infer attention location for wearable transparent displays. In *Proceedings of the 2014 ACM International Symposium on Wearable Computers*. 87–90.
13. Rui I Wang, Brandon Pelfrey, Andrew T Duchowski, and Donald H House. 2014. Online 3D gaze localization on stereoscopic displays. *ACM Transactions on Applied Perception (TAP)* 11, 1 (2014), 3.