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

Teresa Hirzle¹, Jan Gugenheimer¹, Florian Geiselhart¹, Andreas Bulling², Enrico Rukzio¹
¹Institute of Media Informatics, Ulm University, Germany
²Institute for Visualisation and Interactive Systems, University of Stuttgart, Germany
¹firstname.lastname@uni-ulm.de, ²andreas.bulling@vis.uni-stuttgart.de

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 extent 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.

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...
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 assess 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 extent we are able to recognise simple objects a user is looking at. In the future we are planning to fuse this information with depth cameras and quantify performance improvements of the environmental depth map.

**REFERENCES**


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

Prototype
To be able to measure and assess the quality of 3D gaze data we built a prototype consisting of a head-mounted eye tracker (Pupil Labs) and a motion capture system (OptiTrack). Our 3D gaze algorithm is based on a gaze point triangulation approach.

Gaze-Scans
To create the gaze-scans in 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.

A Symbiotic Human-Machine Depth Sensor
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). 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
We presented a first proof of concept implementation to measure 3D gaze in a 50x50x50 cm volume and explored to what extent we are able to recognise simple objects a user is looking at. In the future we are planning to fuse this information with depth cameras and quantify performance improvements of the environmental depth map.

Teresa Hirzle¹, Jan Gugenheimer¹, Florian Geiselhart¹, Andreas Bulling², Enrico Rukzio¹
¹Institute of Media Informatics, Ulm University
²Institute for Visualization and Interactive Systems, University of Stuttgart

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 extent we are able to recognise what a user is looking at by plotting 3D gaze data.