Tracking Active Observers in 3D Visuo-Cognitive Tasks

Markus D. Solbach

John K. Tsotsos

solbach@eecs.yorku.ca

tsotsos@eecs.yorku.ca

Department of Electrical Engineering and Computer Science, York University

Toronto, Ontario, Canada

ABSTRACT

Most past and present research in computer vision involves passively observed data. Humans, however, are active observers in real life; they explore, search, select what and how to look. In this work, we present a psychophysical experimental setup for active, visual observation in a 3D world dubbed PESAO. The goal was to design PESAO for various active perception tasks with human subjects (active observers) capable of tracking the head and gaze.

CCS CONCEPTS

• General and reference → Experimentation; • Hardware → Digital signal processing.

KEYWORDS

eye tracking, motion tracking, head tracking, active observer, active vision

ACM Reference Format:

Markus D. Solbach and John K. Tsotsos. 2021. Tracking Active Observers in 3D Visuo-Cognitive Tasks. In *2021 Symposium on Eye Tracking Research and Applications (ETRA '21 Adjunct)*, May 25–27, 2021, Virtual Event, Germany. ACM, New York, NY, USA, 3 pages. <https://doi.org/10.1145/3450341.3458496>

1 INTRODUCTION

Most past and present research in computer vision involves passively observed data. Humans, however, are active observers in their everyday lives; they explore, search, select what and how to look [Bajcsy 1985, 1988; Bajcsy et al. 2017]. Nonetheless, how exactly active observation occurs in humans so that it can inform the design of active computer vision systems is an open problem [Aloimonos et al. 1988; Dickinson et al. 1997; Krotkov and Brown 1992]. In this work, we present a psychophysical experimental setup for active, visual observation in a 3D world dubbed PESAO. The goal was to design PESAO for various active perception tasks with human subjects (active observers) capable of tracking the head and gaze. While many psychophysical studies explore human performances, there seem to be none that involve active observers with 3D real stimuli. PESAO allows us to bring many studies to the three-dimensional world and to involve active observers. To our knowledge PESAO is the first of its kind experimental setup, combining precise head

Figure 1: Left: A Subject performing an experiment with PESAO. Right: A visualization of the recorded data using PESAO. The subject's movement is plotted as a dashed-line in black, and fixations of either object are illustrated as a fixation frustum in the corresponding colour of the fixated object. Some randomly chosen fixation frusta are annotated with snapshots of the subject's first-person view and the gaze at a particular fixation.

motion tracking in full 3D with gaze tracking at microsecond resolution while the subject is entirely untethered and can move freely and naturally. Furthermore, PESAO provides tracking and recording of 6D head motion, gaze, eye movement-type, first-person video, head-mounted IMU sensor, birds-eye video, and experimenter notes. Fig. 1 (Left) shows a subject performing an experiment and visualized data (Right) using PESAO. The project page can be found under <https://data.nvision.eecs.yorku.ca/PESAO/>.

2 PESAO

The core hardware pieces are a motion tracking system, an eye-tracking system, and glasses tracking body. The motion tracking system is used to track the pose of the head, as well as the pose of other objects in the scene, such as 3D visual stimulus. PESAO uses a motion tracking system by OptiTrack with six motion tracking cameras recording at 120Hz. The tracking volume spans over 4m x 4m x 2m and can be increased by adding more cameras to the system. PESAO's modular design allows the integration of more markers easily. If more body parts besides the head are wished to be tracked or more objects, this could be achieved by using more tracking markers. In terms of accuracy, the motion tracking system can achieve an accuracy of up to 0.2mm in 97% of the capture volume [Aurand et al. 2017]. Only very near the edges of the capture volume (3%) would the error exceed 0.2mm. However, this depends on multiple parameters such as the number of cameras, camera type, size of the volume, interferences of the environment, quality of the

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike International 4.0 License.

ETRA '21 Adjunct, May 25–27, 2021, Virtual Event, Germany

© 2021 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-8357-8/21/05.

<https://doi.org/10.1145/3450341.3458496>

Figure 2: PESAO and its connectivity overview.

calibration, etc. In our configuration, we achieved an accuracy of 0.4mm in 97% of the capture volume. PESAO reports an estimated error and provides it alongside the tracking data.

As an eye-tracking solution, we chose the Tobii Pro Glasses 2. The glasses are capable of recording precise gaze information at either 50 or 100Hz (dependent on the model), first-person video, and motion information (accelerometer and gyroscope). The mean accuracy for this kind of glasses is 1.42 degrees error in visual angle [MacInnes et al. 2018]. Similar to the motion tracking error, the eye-tracking error is provided through the data as well. Furthermore, we used the Tobii Pro Lab software to export gaze information and detect eye events. To integrate the eye-tracking data with the motion-tracking data, we developed a custom tracking body for the Tobii Pro Glasses 2. The tracking body is a lightweight (21.6g incl. markers), snap-on solution that works with most hairstyles and head-decorations. While it adds to the size of the glasses and, therefore, its inertia, none of our subjects reported it to be disruptive. Furthermore, the body uses eight tracking markers assembled in a rotation variant configuration for reliable tracking.

In order to estimate the additive error of motion- and eye tracker specific to an experiment, we recommend performing a calibration trial(s) for each subject following a known pattern of movements and observations.

For several visual perception experiments, it is also of interest to track the object's position under investigation. We have designed two custom object tracking bodies that can be used to track objects in 3D space. Our visual system cannot function without light; hence to study the effect of light, PESAO offers controllable light settings and light measuring capabilities with a light sensor. For our tracking area, we have used five light sources, one in each corner and one above functioning as a ceiling light.

Fig. 2 gives an overview of how to integrate the hardware. Important to note is that PESAO relies on two different connectivity standards; USB and WiFi. Components able to be connected over USB should be connected over a wired connection for robustness. However, the eye-tracking system, to allow unconstrained experiments, is connected over WiFi. The WiFi connection is used to control and live view the eye-tracking system. Specifically, we enable the set up, start and stop recordings and acquire synchronization timestamps of the system. The recording itself (gaze information, first-person video, calibration data) is stored on an SD-card in the recording unit of the eye-tracking system and is copied after the experiment for better performance. Fig. 1 (Left) shows a subject wearing the glasses with the glasses tracking body attached.

3 PESAOLIB

Figure 3: Overview of PESAOLib. Showing the dataflow and corresponding software modules.

PESAO's software to control and execute experiments, record data with precise, accurate to microsecond-level timestamps, as well as synchronize and analyze the recorded data is packaged as PESAOLib. It is implemented as a processing pipeline that provides temporary files along its execution to help the user understand intermediate steps better and modularize their workflow. The software library is subdivided into three modules; PESAOLib Recording, PESAOLib Processing, and PESAOLib Evaluation. PESAOLib Recording, as the name suggests, deals with connecting to and recording from devices and builds the first execution step in the pipeline. PESAOLib Processing uses the recorded files to synchronize them based on a global timestamp and stores them in a consolidated file. Lastly, PESAOLib Evaluation provides tools and examples to evaluate and visualize the data. The software is written in Python and C++ and uses the networking and synchronization functionalities of the lab-streaming layer by Swartz Center for Computational Neuroscience [Kothe 2014] as a recording framework. Software parts that might vary from experimental design to another experimental design are provided in Python Fig. 3 displays a diagram showing the dependencies between PESAO's modules. All programs are executed on the Workstation (Fig. 2). However, as PESAO is designed, programs can be run on multiple connected workstations. This might be of interest if more sensors or a higher bandwidth is needed.

A strength of this system is that the Subject is entirely untethered, allowing free and natural motion. To compensate for transmission loss and lag, the eye-tracking device records on its local SD-Card. As mentioned earlier, adjustable PESAOlib modules are written in Python (marked by the green indicator in Fig. 3). However, modules that require high performance and are foreseen to remain unchanged for most setup scenarios are implemented in C++ (magenta indicator in Fig. 3). Nevertheless, the C++ source code is provided and can be changed as desired. Lastly, to allow the best interoperability with the physical devices, we use the manufacturer's programs (blue indicator in Fig. 3). An example of a complete PESAOlib pipeline execution with an evaluation of the data can be seen in Fig. 1 (Right).

4 CONCLUSION

In conclusion, PESAO provides precise and synchronized head and eye-tracking that allows for the investigation of active visual observation in a 3D world. Building mostly upon source software, we hope it finds many use-cases and different instantiations in and outside laboratories worldwide. The example given in Fig. 1 shows ongoing research investigating the same-different task [Carroll 1993] but in 3D and including active observers. Already published work using PESAO can be found in [Tsotsos et al. 2021].

In the future, we hope to extend PESAO to different hardware vendors, making it accessible cross-platform, and perhaps integrate more sensors, such as full-body tracking, including legs and hands, and EEG.

ACKNOWLEDGMENTS

This research was supported by grants to the senior author (JKT) from the following sources: Air Force Office of Scientific Research USA (FA9550-18-1-0054), The Canada Research Chairs Program (950- 231659) and Natural Sciences and Engineering Research Council of Canada (RGPIN-2016-05352). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

- J Aloimonos, I Weiss, and A Bandyopadhyay. 1988. Active vision. *International Journal of Computer Vision* 1, 4 (1988), 333–356. <https://doi.org/10.1007/BF00133571>
- Alexander M. Aurand, Jonathan S. Dufour, and William S. Marras. 2017. Accuracy map of an optical motion capture system with 42 or 21 cameras in a large measurement volume. *Journal of Biomechanics* 58 (2017), 237–240. <https://doi.org/10.1016/j.jbiomech.2017.05.006>
- Ruzena Bajcsy. 1985. Active Perception vs. Passive Perception. In *Proceedings of 3rd IEEE workshop on computer vision: Representation and Control*. Washington DC: IEEE Computer Society Press, Bellare, MI., 55–62.
- Ruzena Bajcsy. 1988. Active Perception. *Proc. IEEE* 76, 8 (1988), 966–1005. <http://psycnet.apa.org/psycinfo/1993-98908-000>
- Ruzena Bajcsy, Yiannis Aloimonos, and John K. Tsotsos. 2017. Revisiting active perception. *Autonomous Robots* 42, 2 (2017), 177–196. <https://doi.org/10.1007/s10514-017-9615-3> arXiv:1603.02729
- John B Carroll. 1993. *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press.
- S J Dickinson, H I Christensen, J K Tsotsos, and G Olofsson. 1997. Active Object Recognition Integrating Attention and Viewpoint Control. *Computer Vision and Image Understanding* 67, 22 (1997), 239–260. <https://doi.org/10.1006/cviu.1997.0532>
- Christian Kothe. 2014. Lab streaming layer (LSL). <https://github.com/sccn/labstreaminglayer>
- Eric Krotkov and Chris Brown. 1992. *Active Computer Vision*. (1992).
- Jeff J MacInnes, Shariq Iqbal, John Pearson, and Elizabeth N Johnson. 2018. Wearable Eye-tracking for Research: Automated dynamic gaze mapping and accuracy/precision comparisons across devices. *bioRxiv* (2018).

John K Tsotsos, Omar Abid, Iuliia Kotseruba, and Markus D Solbach. 2021. On the control of attentional processes in vision. *Cortex* 137 (2021), 305–329. <https://doi.org/10.1016/j.cortex.2021.01.001>