DOI: 10.2390/biecoll-saga2013_7

In Pfeiffer and Essig (2013): Proceedings of the First International Workshop on Solutions for Automatic Gaze-Data Analysis 2013 (SAGA 2013). DOI: 10.2390/biecoll-saga2013_0

Studying joint attention and hand-eye coordination in human-human interaction: A model-based approach to an automatic mapping of fixations to target objects

Patrick Renner, Thies Pfeiffer

1 Introduction

If robots are to successfully interact in a space shared with humans, they should learn the communicative signals humans use in face-to-face interactions. For example, a robot can consider human presence for grasping decisions using a representation of peripersonal space (Holthaus & Wachsmuth, 2012). During interaction, the eye gaze of the interlocutor plays an important role. Using mechanisms of joint attention, gaze can be used to ground objects during interaction and knowledge about the current goals of the interlocutor are revealed (Imai et al., 2003). Eye movements are also known to precede hand pointing or grasping (Prablanc et al., 1979), which could help robots to predict areas with human activities, e.g. for security reasons.

We aim to study patterns of gaze and pointing in interaction space. The human participants' task is to jointly plan routes on a floor plan. For analysis, it is necessary to find fixations on specific rooms and floors as well as on the interlocutor's face or hands. Therefore, a model-based approach for automating this mapping was developed. This approach was evaluated using a highly accurate outside-in tracking system as baseline and a newly developed low-cost inside-out marker-based tracking system making use of the eye tracker's scene camera.

2 Experiment

The study is based on an adapted receptionist scenario: A map of a building with three floors is located between two participants. For each participant, there is one floor close-by, one in an intermediate distance within reaching space, and one in the distance which cannot be reached directly (Figure 1). Besides patterns of joint

Patrick Renner e-mail: prenner@techfak.uni-bielefeld.de
SFB 673, Faculty of Technology, Bielefeld University, 33615 Bielefeld

Thies Pfeiffer e-mail: tpfeiffe@techfak.uni-bielefeld.de
Cognitive Interaction Technology Center of Excellence (CITEC), Bielefeld University, 33615 Bielefeld

attention and patterns of hand-eye coordination that may predict pointing gestures, we are interested in differences in gaze behaviour and the use of modalities for the three distances. Head and hand movements of both participants are tracked using the high end system. Eye movements of one participant are tracked and both participants are recorded on two video cameras (audio and video).

The procedure of the main task is as follows: Firstly, one of the participants draws a miniature floor plan on which a starting point and a target room are marked. She is asked to show both of them on the real plan and describe the route. Afterwards, the partner draws a card with blockings drawn in. She is supposed to mark these with small tokens. Then, both participants jointly plan the fastest remaining route. The roles of the task are alternated each two of twelve trials altogether.

As we suppose that friends and acquaintances will act differently than strangers, in this stage of the experiment, each pair of participants is required to know each other. The participants firstly fill out a questionnaire which is meant to retrieve information about gender, age, debility of sight and prophecy in order to be able to account for possible behavioural differences.

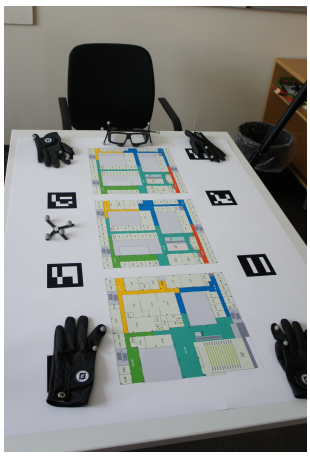

Fig. 1 Setup of the experiment: Floor plans, tracking markers, gloves and glasses.

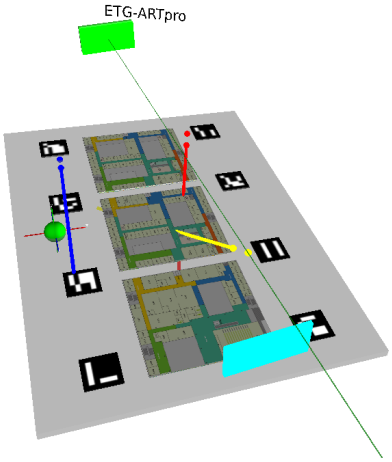

Fig. 2 Virtual reality simulation of the experiment including eye- (green), head- (green and blue) and hand (red and yellow) tracking.

3 Method: Model-based annotation of fixations

In experiments like ours, users move their head freely and gaze analysis requires a manual mapping of all fixations from the scene camera video to real world objects. To enable an automatic fixation mapping, the flexible head positions have to be

tracked. Pfeiffer (2012) proposed a method for measuring attention in 3D space by using an optical tracking system and a tracking target attached to the eye tracker.

Here, we use two approaches of tracking the position of the eye tracker in space. An ART optical tracking system is a highly accurate and fast outside-in solution. Additionally, the built-in scene camera of the eye tracker is used as a low-cost inside-out tracking approach: By calibrating the intrinsic camera parameters, it is possible to transform recorded images of geometries of known size to their corresponding 3D pose in the real world. For that purpose, we chose fiducial augmented reality markers (see Figure 2 for examples) that can easily be detected in the camera images.

We then have to map gaze directions and fixations to our target stimuli: Room positions and the counterpart's head and hands. For this we modelled the floor plan in virtual reality and fed eye tracking and marker tracking data into the system. By reconstructing the line of sight in 3D, we can then automatically detect fixations on the virtual floor plan. For detecting fixations on the interlocutor, we use a face detection algorithm. Pointing directions are detected using the ART tracking system: The participants wear light gloves with markers attached.

4 Results and Conclusion

Using our approach (see Figure 3), the interaction of the participants can be simultaneously simulated in the virtual scene (Figure 2): Gaze directions are cast as rays into the scene. By testing for collision of those rays with the target stimuli, higher-level events can be output, e.g. *Fixation of 300ms on room 101*, which can be analysed easily without the need for manual annotation.

The whole interaction can be analysed using either tracking solution, except for the finger tracking. Both systems are capable of real-time analysis: The outside-in approach has a frame rate of 60 Hz, but is restricted to the 30 Hz input of gaze di-

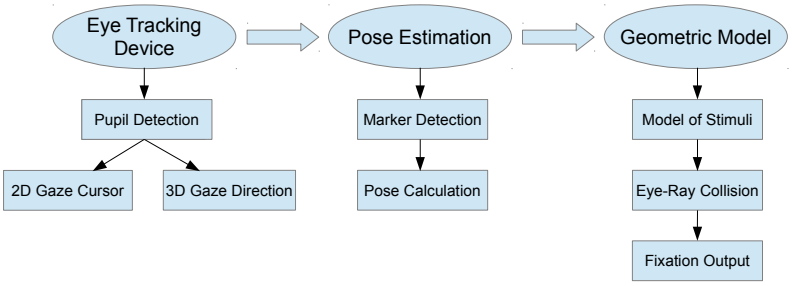

Fig. 3 The different steps of our inside-out tracking based approach: All gaze-related computations are done by the eye tracking device. The scene-camera image is used to estimate the pose of the device using marker tracking. Gaze and pose information are combined to identify fixations on target stimuli by means of ray-casting in a 3D scene model.

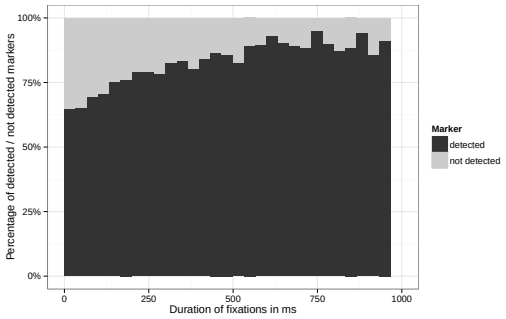

Fig. 4 The figure shows the percentage of detected markers depending on the duration of fixations.

rections from the eye tracker. The inside-out approach runs at the 24 Hz of the eye tracker's scene camera. It has, however, a mean delay of 379 ms (sd: 90 ms) compared to the outside-in approach, which is not sufficient for real-time interaction, but gaze and pose data are synchronized and thus the accuracy of the fixation classification is not affected by latency. In a comparison study, the inside-out solution could estimate the pose in 75.96% of the analysed frames, the outside-in solution covers all. In the remaining frames no markers were present or the image of the scene camera was smeared, caused by quick head movements. During fixations, however, the head remains relatively stable and losses of markers are less likely. Indeed, Figure 4 shows that the percentage of detected markers increases during longer fixation. The accuracy of the inside-out tracking is good, compared to the highly accurate outside-in tracking we have an average deviation of 1.11 cm (sd: 0.69 cm) in the 3D position and 1.39 degrees (sd: 0.68 degrees) in the orientation.

To conclude: The presented model-based approach combined with the inside-out tracking, which does not require additional expensive equipment except for the eye tracker itself, allows for an automatic analysis of gaze data in our scenario.

ACKNOWLEDGMENTS: This work has partly been funded by the SFB 673 "Alignment in Communication", in the project "Interaction Space".

References

1. Holthaus, P., Wachsmuth, S. (2012). Active Peripersonal Space for More Intuitive HRI. International Conference on Humanoid Robots: 508513.
2. Imai, M., Ono, T., & Ishiguro, H. (2003). Physical relation and expression: joint attention for human-robot interaction. IEEE Transactions on Industrial Electronics, 50(4):636643.
3. Prablanc, C., Echallier, J., Komilis, E., Jeannerod, M. (1979). Optimal response of eye and hand motor systems in pointing at a visual target. Biological cybernetics 124:113124.
4. Pfeiffer, T. (2012). Measuring and visualizing attention in space with 3D attention volumes. Proceedings of the Symposium on Eye Tracking Research and Applications:18.