

A dataset of head and eye gaze during dyadic interaction task for modeling robot gaze behavior

Mirko Raković^{1,2,*}, Nuno Duarte¹, Jovica Tasevski², José Santos-Victor¹ and Branislav Borovac²

¹ University of Lisbon, Instituto Superior Tecnico, Institute for Systems and Robotics, Av. Rovisco Pais, 1049-001, Lisboa, Portugal

² University of Novi Sad, Faculty of Technical Sciences, Trg Dositeja Obradovića 6, Novi Sad, Serbia

Abstract. In this work is presented a dataset of humans' head and eye gaze acquired with Pupil Labs gaze-tracking glasses and Optitrack motion capture system. The dataset contains recordings of adult subjects in dyadic interaction task. During the experiment, the subjects are asked to pick up an object and, based on the randomly defined instructions, to place it on the table in front of her/him or to give the object to a person sitting across the table. If the object is handed over, the second person takes the object and places it on the table in front of her/him. The dataset is intended to be used to model the behavior of the human's gaze while interacting with another human and implement the model in a controller of a robot for dyadic interaction with a humans.

1 Introduction

Understanding human movements, actions, and intentions are important when two actors (human or robot) share a common workspace. In a human-human interaction (HHI) this process is called nonverbal communication in a dyadic scenario. This form of nonverbal communication is partially enabled by action observation. One of the main cues in action understanding is gaze which can influence observer's selection of motor activities. This explains why the attention should be paid to proper modeling of the robot's gaze behavior that needs to be easily understandable by humans.

Nonverbal signals are an integral part of all our communicative endeavors. In some cases, they are the most significant part of our message. In HHI, communication through nonverbal channels also influences the coordination of joint activity. We believe this is the same in the case of human-robot interaction (HRI). So the design of readable human-like behavior will support efficient and robust teamwork. Motivated by a desire to develop effective robot teammates for people, one of our goals is to model human-like gaze behavior as an integral part of nonverbal communication in HRI.

In [1] is shown that implicit nonverbal communication positively impacts human-robot task performance with respect to efficiency, and robustness to errors that arise from miscommunication. Authors in [2, 3] validated the use of gaze in addition to body pose cues as means of predicting human action. Thus understanding and effectively using nonverbal cues is of great importance to success for robots in dyadic tasks.

The initial step to model gaze behavior is to acquire quantitative data of humans' movement during dyadic interaction task. For this purpose, we prepared the experiment with two humans that have to interact in order to accomplish their tasks (Fig. 1.). In this experiment, we measured head and eye gaze.



Fig. 1. Interaction in dyadic scenario

Paper [4] presents a data set that captures the gaze patterns of humans solving a recognition task. Eye movements were recorded using static gaze tracker, and the participants were looking at the content displayed on the LCD display. In [5] is presented the MPIIGaze dataset that contains 213,659 images collected from 15 participants during natural everyday laptop use over more than three months. Another publicly available dataset, called Eyediap [6], contains 94 video sequences of 16 participants looking at three different targets (discrete and continuous markers displayed on a monitor, and floating physical targets) under both static and free head motion. All the existing dataset are missing the gaze information during the interaction scenario. With this work, we tend to bridge this gap and

* Corresponding author: rakovicm@ist.tecnico.ulisboa.pt

to use it as a base for modeling the behavior of a humanoid robot.

Mutual gaze awareness is important in communication and collaboration in group activities. Prior works were mainly showing different ways of how to analyze gaze cues visually. Some recent studies are analyzing the behavior of two users in different scenarios by measuring the gaze using head mounted gaze tracker devices. Paper [7] gives early work-in-progress that explores the effects of gaze awareness on gameplay, in particular when the gaze of one or more players is augmented over the game and revealed to others. Authors in the paper [8] presented an exploratory study to understand how gaze cues can enhance collaboration between two users in front of a large shared display.

This paper presents the dataset of the head and eye gaze together with image sequences of eyes and scene view collected during the HHI tasks. In Section II is explained the task humans have to accomplish during which the movements are recorded. Section III describes hardware and software setup used to collect the data, while Section IV explains the details of the collected dataset. The conclusion and the future work is given in Section V.

2 Dyadic interaction task

Findings in neuroscience [9] suggest that the human motor control combines state estimation, feed-forward, and multiple feedback loops operating at different speeds. Its structure is highly modular and is believed to rely on the combination of motor primitives to generate complex movements.

Authors in [11] used coupled dynamical systems for the realization of coordinated complex movements of the robot's upper body when performing reaching and grasping motion in the presence of obstacles. The computational model of the eye–arm–hand coupling is based on human motion data collected with subjects performing a prehensile motion with obstacle avoidance. The coupling between dynamical systems is learned from the experiment with a single human performing reaching and grasping tasks. Paper [11] shows the experiment with a human performing bimanual movements and with two humans each performing reaching movements at the same time. In the case of dyadic interactions, the results indicated that co-actors synchronize the timing of their movements, although the task in itself is discrete and non-rhythmic.

With this experiment, we want to create a basis for research on how to integrate this coupling in robot's motor control system, in scenarios where both human and robot, share the same space and objects during task execution. For that purpose, the participants are asked to assemble a pair of towers inside a circle on the paper in front of them. Both towers are assembled from 3D printed objects of different shape or color as shown in Figure 2. The objects are marked with numbers 1-3.

The numbers are used to define the position of the object in the tower.



Fig. 2. Objects for assembling the tower.

In the beginning, two stacks of three objects are placed next to each participant. A stack of objects is positioned below the table top in order to occlude them from the other person. Next to the stack of objects is given a paper with the desired order of the objects to build the tower. (Fig. 3.).

When the assembly of towers starts, the participants are asked, one at a time, to pick the first object from the stack. If the number of the object matches the number in their next level of the tower, they should use the object for their tower. Otherwise, they are instructed to give, i.e. handover the object to a teammate. Thus, there are two types of actions the participant can execute: (i) intra-personal action (pick and place an object on its tower, i.e. placing action) or (ii) inter-personal action (pick and handover an object, i.e. giving action).

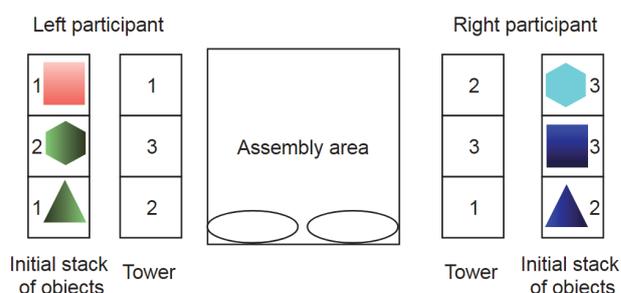


Fig. 3. Illustration of the initial stack of objects and the task given to the participants.

The towers are defined such that in the case of a handover, the object given to another participant is always the matching object for her/his next level in the tower. After an object is positioned in one of the towers, the turn is taken by a second participant. The actions are repeated until all the objects are used and both towers are assembled. Illustration of the progress of the task with the order and the type of action is given in Figure 4.

Once the assembly is finished, the new task and the new initial stack of objects is prepared and given to the participants. Each pair of participants had to repeat the task four times, i.e. to assemble four different pairs of towers.

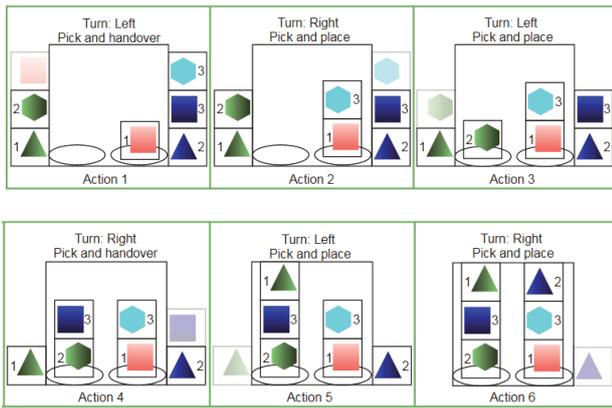


Fig. 4. Example of turn-taking order (left and right participant) and type of actions (pick and place or pick and handover) for assembling two towers.

The four tasks are defined in a way that there is always a different number of giving actions. This is to prevent the subjects to predict the action ahead of time. The goal here is to record a natural, unbiased human gaze behavior. The first task has two giving and four placing actions, the second task has six giving and no placing actions, the third task has no giving and six placing actions and the fourth task has four giving and two placing actions. Thus, during the experiment two participants performed together twelve giving and twelve placing actions.

3 The Data Acquisition Setup

When observing or scanning immediate surroundings, human eyes make jerky saccadic movements and stop several times, moving very quickly between each stop. The speed of movement during each saccade cannot be controlled, and the eyes move as fast as they are able [12]. To capture such eye movements, in this experiment both participants were wearing Pupil-Labs binocular gaze trackers [13].

During the performed actions, participants’ head gaze was recorded using Optitrack motion tracking system [14]. Hardware and software setup used to acquire dataset is illustrated in Figure 5.

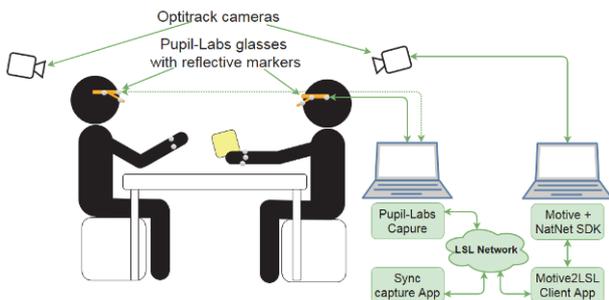


Fig. 5. Experiment hardware and software setup.

The Pupil Labs binocular gaze tracker is in the form of glasses equipped with three cameras. Two cameras are recording eyes at ~120Hz. A video stream of the egocentric view is recorded at 60Hz. The pupil detection algorithm does not depend on corneal reflection

technique [15] and as reported in [13] the gaze tracker should work with users who wear contact lenses and eyeglasses. However, we experienced difficulties in calibrating the glasses with such participants, and we had to choose the participants not wearing glasses and contact lenses. Before the recording starts, each participant first calibrates his/her gaze tracker using the screen calibration method.

Optitrack motion capture system captures passive opto-reflective spherical markers at 120Hz. To record head gaze we fixate five opto-reflective markers on each glasses. Each group of five markers represented one rigid body whose position and orientation in the reference frame is being recorded.

Software setup is composed of following applications. For gaze data recording we used Pupil Labs Capture. For recording the body movements the Motive software platform is used. Since we want to capture synchronous data of head and eye gaze it was necessary to merge the input from two sensory systems. For that purpose is used Lab streaming layer (LSL) library [16]. LSL is designed to be a system for unified collection of measurement time series of various sensing equipment that handles both the networking, time-synchronization, (near-) real-time access and optionally the centralized collection.

In order to use LSL, we developed a Motive2LSL application that captures the broadcasted position of the markers and rigid bodies tracked within Motive software platform. Another application we developed is Sync capture application that receives the data measurements from two Pupil-Labs glasses and Optitrack cameras and records those data together with timestamps of the measurements into a file with synchronization timestamps.

4 The Dataset

We have acquired the data of three pairs, i.e. six participants. Participants were adults between 25 and 40 years of age. Acquisitions were performed at the Institute for Systems and Robotics, IST, University of Lisbon, during January 2018. The dataset contains:

- a video stream of the egocentric view with associated timestamps for both glasses,
- pupil data with gaze positions, pupil positions and its timestamps,
- position and orientation of rigid bodies representing head gaze with its timestamps and
- synchronization file with timestamps of gaze tracking and motion tracking data.

In each recording, subjects had to perform 6 actions in one task, and each pair of subjects had 4 tasks. Thus, we collected the 24 actions in dyad scenario, i.e. 48 gaze and head motion for each pair (24 observer’s movements and 24 performer’s movements). The recordings are repeated for three different pair of humans, and thus we collected 72 actions and 144 gaze and head movements.

Figure 6 illustrates the recorded gaze movement of the eyes during the experiment for two different actions (placing and giving) and for two different roles (performer and observer) of the subject.

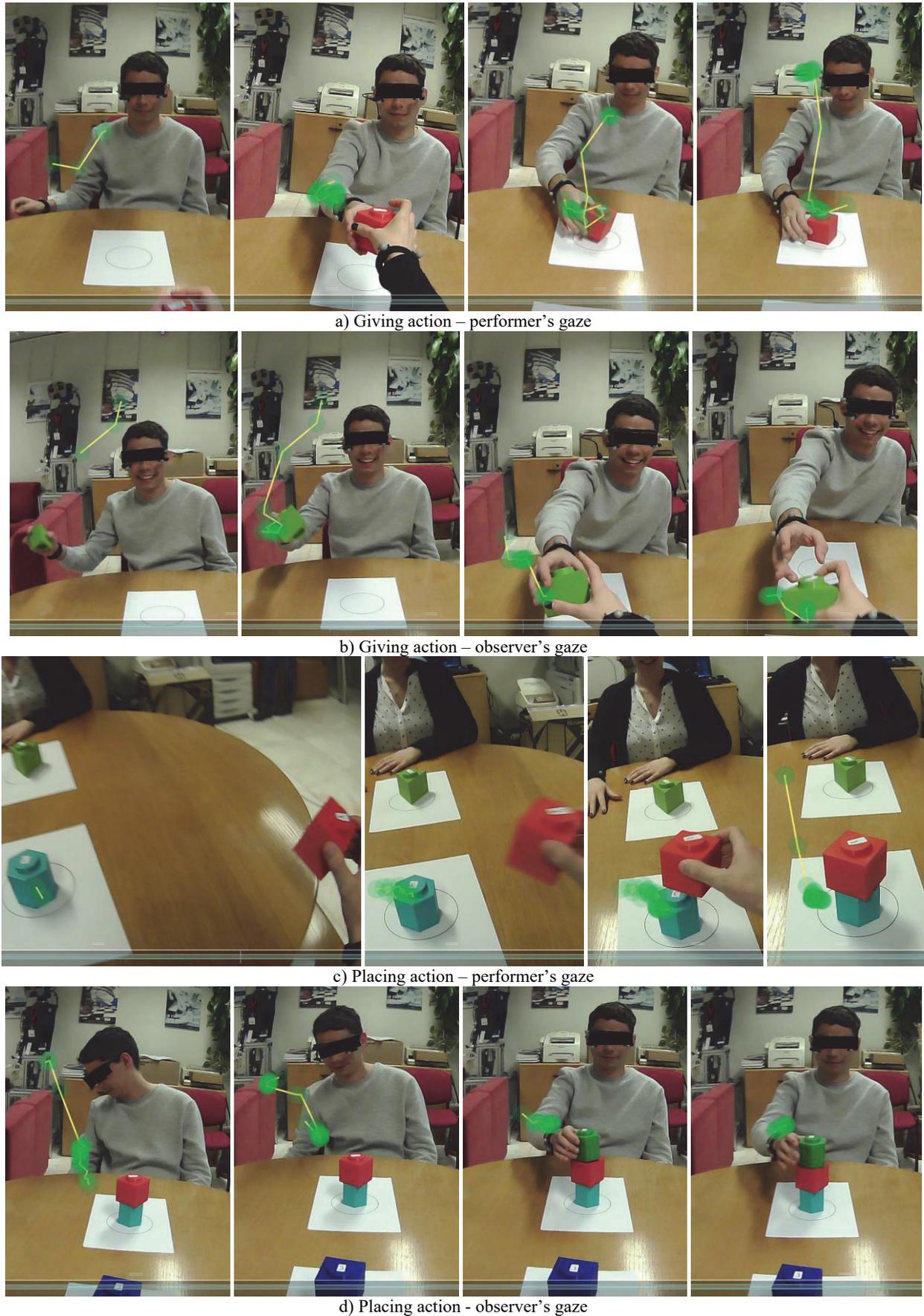


Fig. 6. Illustration of data set with an example given by an image sequences showing the gaze of a performer/observer during placing and giving actions (green circle represent the recorded gaze points, yellow line represent interpolation between recorded gaze points).

5 Conclusion

A multimodal dataset containing eye gaze and head gaze movement during placing and giving actions has been acquired using gaze tracking glasses integrated with motion capture system. Six adult subjects have participated in the recording sessions and made their data available for research. The acquired data has sufficient quality to investigate the behavior of gaze during dyadic interaction task. Our next steps will consist in annotating specific timestamps in the recorded movement so temporal correlations between important events can be established. After annotating the data, we will focus on modeling of the gaze behavior in interaction scenarios, and implement the model on a humanoid robot platform.

This work was partially supported by EU H2020 project 752611-ACTICIPATE, UID/EEA/50009/2013 FCT project and MNTR project III44008. The authors would like to thank all the volunteers that participated in the experiment.

References

1. C. Breazeal, C.D. Kidd, A.L. Thomaz, G. Hoffman, M. Berlin, IEEE/RSJ International Conference on Intelligent Robots and Systems, (IROS 2005), 708–713 (2005)
2. P. Schydlo, M. Raković, J. Santos-Victor, IEEE International Conference on Robotics and Automation, (ICRA 2018) (2018)
3. N. Duarte, J. Tasevski, M. Coco, M. Raković, J. Santos-Victor, arXiv preprint arXiv:1802.02788 (2018)
4. S. Mathe, C. Sminchisescu, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), **37** (2015)
5. X. Zhang, Y. Sugano, M. Fritz, A. Bulling, Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 4511–4520 (2015)
6. K.A. Funes Mora, F. Monay, J.M. Odobez, Proceedings of the Symposium on Eye Tracking Research and Applications, 255–258 (2014)
7. J. Newn, E. Velloso, M. Carter, F. Vetere, Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts, 239–244 (2016)
8. Y. Zhang et al, Personal and Ubiquitous Computing, **21** (1), 173–186 (2017)
9. F.A. Mussa-Ivaldi, S.F. Giszler, E. Bizzi, Proceedings of the National Academy of Sciences, **91** (16), 7534–7538 (1994)
10. L. Lukic, J. Santos-Victor, A. Billard, Biological cybernetics, **108** (2), 223–248 (2014)
11. C. Jung, A. Holländer, K. Müller, W. Prinz, Experimental brain research, **211** (3-4), 471–482 (2011)
12. R.C. Neil, *Psychology: the Science of Behaviour* (Toronto: Pearson, 2010)
13. M. Kassner, W. Patera, A. Bulling, Proceedings of the 2014 ACM International Joint Conference On Pervasive and Ubiquitous Computing: Adjunct publication, 1151–1160 (2014)
14. Point Natural. "Optitrack." Natural Point, Inc., Available at: <http://www.naturalpoint.com/optitrack/>
15. C. Jian-Nan et al, WRI Global Congress on Intelligent Systems, (GCIS'09), 133–138 (2009)
16. Kothe C., Lab streaming layer (LSL), Available at: <https://github.com/sccn/labstreaminglayer>