Eye Gaze and Body Motion Synchronization in Dyadic Interaction

Danilo Nikić, Nikola Ilić, Darko Todorović, Nuno Ferreira Duarte, José Santor-Victor, Branislav Borovac, *Member, IEEE* and Mirko Raković, *Member, IEEE*

***Abstract* — Understanding the behavior alignment in dyadic human-human interaction and human-in-the-loop control in human-robot interaction relies on reliable tracking of the human motion. The gaze tracking and motion capture are common techniques that are used nowadays, but they are usually used separately. In this work we combined two Pupil-labs gaze tracker with a Vicon optical motion capture system. To synchronize the recordings of all devices we developed the solution that utilized Lab Streaming Layer for unified collection of measurement time series in research experiments that handles both the networking, time-synchronization and (near-) real-time access of the data. The aim of the experimental setup is to study the interaction of two humans while performing a joint task that requires interpersonal motion coupling.**

***Index Terms*—human-human interaction, human-robot interaction, eye gaze, motion capture.**

I. Introduction

THE synchronized measurement of eye gaze, head gaze, trunk, and arm/hand movement enables insight into the action-perception process. This is especially important in dyadic interaction [1], for understanding interpersonal behavior coupling, mechanisms to read each other's intentions and anticipate the action of a co-worker. Such data quantifies the nonverbal communication that is comprehensive only when both (i) body parts motion capture and (ii) eye gaze tracking are merged and timely aligned.

Experimental setup (Fig. 1.) is prepared to gather the data for modeling intra- and interpersonal coupling and for developing a solution for action anticipation and consequently action planning in human-human (HHI) and human-robot interaction (HRI). The question of interest is whether the human is coupling eye gaze and body movements with his/her co-worker when picking the objects from the box, handing over the object, and placing them on the table. In HRI, eye gaze behavior, together with the activity of the hands and other parts of the body, has been argued to be important for anticipation and mutual alignment in the behavior [2-3].

D. Nikolić, N. Ilić, B. Borovac and M. Raković are with the Faculty of Technical Sciences, University of Novi Sad, Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia (e-mail: danilonikic4@gmail.com, nilic993@gmail.com, borovac@uns.ac.rs, rakovicm@uns.ac.rs).

D. Todorović is with the Faculty of Electronic Engineering, University of Niš, Serbia (e-mail: Darko.Todorovic@elfak.ni.ac.rs).

J. Santos-Victor, and N.F. Duarte are with the Institute for System and Robotics, Insituto Superior Tecnico, Av. Rovisco Pais 1, 1049-001 Lisboa, Portugal, University of Lisbon (e-mail: jasv@isr.tecnico.ulisboa.pt, tp.aobsilu.ocincet.rsi@etraudarierrefn).

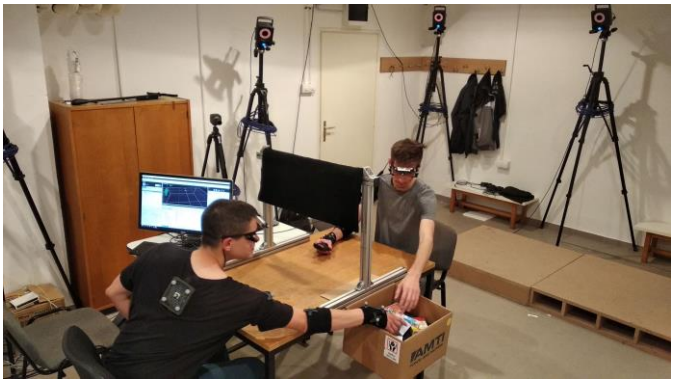

Fig. 1. Two human co-workers are picking the items from a box, scanning the barcode and handing over or placing the item on the table. During the experiment, eye gaze, head gaze, trunk, and arm motion is tracked with a Vicon motion capture system and Pupil-Lab gaze-tracking glasses.

Gaze tracking and motion capture are commonly used in many areas from psychology [4] and sports [5] to biomechanics [6] and robotics [7], but separately. The difficulty in fusing data from the different measuring system (i.e. motion capture system and gaze tracking), is time alignment and synchronization. Data acquisition and precise synchronization are especially significant for time-critical analysis of the data and for relating the different data streams to each other [8]. Interpersonal coupling, due to the nature of humans, is a stochastic process with asynchronous events, and precise time alignment is of paramount importance.

II. State Of the Art and Contribution

A. Motion capture

Motion capture systems are available in different technologies [9]. Systems based on inertial measurement units are measuring the acceleration and angular velocity in three dimensions of a sensor attached to body parts. Another type of system is motion capture that is measuring position and orientation of the object in the magnetic field. In this research, we are using infrared camera-based system that is detecting the light reflected by markers attached to body parts and the objects. Optical motion capture system uses direct linear transformation to determine the exact position and orientation of cameras with respect to the reference coordinate frame. Then the minimum two cameras are needed to capture one marker in order to determine three-dimensional representation [6]. In order to track the position and orientation of the body, a minimum of three markers are needed to be attached to each body part. Different manufacturers are choosing between two

approaches for data labeling. Qualisys, for example, is recording the raw marker data and reconstructs the body motion after post-processing [9]. Manufacturers such as Vicon and OptiTrak allow the user to define the body model prior to the recording [10, 11]. The main challenge of optical systems is that occlusions of markers during the recording cause marker loss and gaps in the data. Thus, such occlusions should be prevented by careful marker placement and camera positioning before and during the recording.

B. Eye gaze tracking

Most of existing gaze tracking systems can be divided into two groups, screen-based or head mounted gaze tracking glasses [12]. In this work we are using head mounted gaze tracker from Pupil-Labs [13]. It is a system of two cameras with an infrared light source for capturing eye movement and RGBD camera for recording the scene. Before the recording starts, it is necessary to calibrate the glasses, to match the actor's eyes to the model, for the estimation of the gaze point in the egocentric view. During the calibration, the actor needs to focus on the calibration marker on the screen while keeping the head still. Pupil-Labs gaze tracker can capture data at up to 200Hz and is able to estimate the position of the gaze in 3D.

C. Synchronization of gaze tracking and motion capture

Synchronized gaze tracking and motion capture represent a multimodal information human behavior. Precise time alignment is important to extract relevant time events in the change of the behavior and its modeling. To synchronize multiple data streams, one can rely on sync boxes and plug-ins [14] or analog claps [15] equipped with motion capture markers captured by the gaze tracking glasses egocentric view camera. In [8] authors presented a method to synchronize Ergoneers Dikablis gaze-tracking glasses with Qualisis Oculus motion capture.

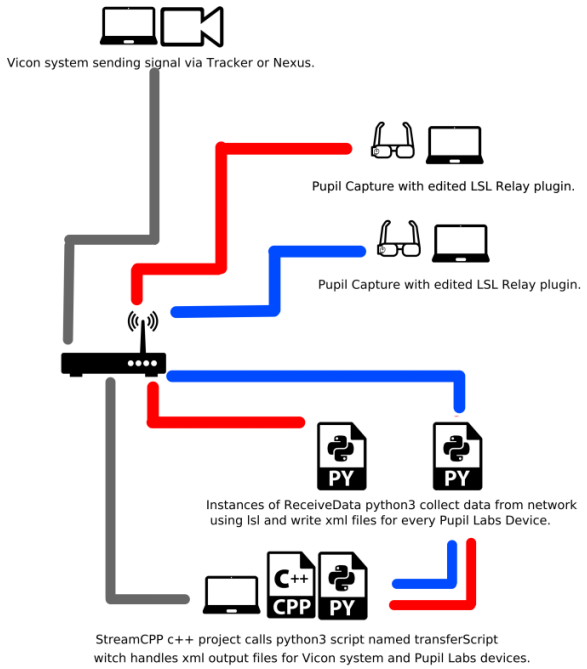

Fig. 2 Diagram of the experimental setup

In our work, as mentioned, we are syncing two Pupil-Labs gaze-tracking glasses with a Vicon motion capture system. The synchronization leverages on Network Time Protocol [16] and its implementation in lab-streaming-layer (LSL) [17]. The block diagram of the hardware and software components of our solution is given in Fig. 2.

Hardware-wise it is composed of Vicon cameras connected to Vicon Giganet MX box and dedicated PC, two Pupil-Labs glasses connected to laptops, a PC for the acquisition of data from different streams and a switch for establishing TCP/IP connection between all sensory systems. Software-wise, the system uses Vicon Nexus (Fig. 3.) for motion capture recording/streaming, Pupil capture for eye gaze recording/streaming, an application developed to stream data from Vicon system to LSL and application developed to capture and synchronize all data.

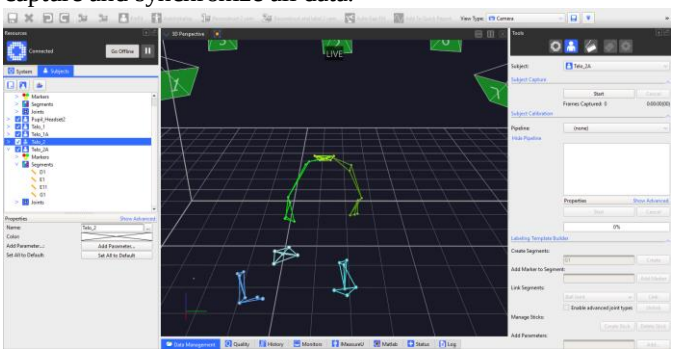

Fig. 3. Illustration of recording from Vicon Nexus while an actor (with markers placed on the glasses and both arms) interacting with a co-worker and manipulating the objects

First developed application (called StreamCPP) records the data on position and orientation of an object (rigid body) or subject (kinematic chain of rigid bodies) defined by markers in working environment whereas matching acquisition of raw data is done directly in Vicon Nexus. Subjects and objects in the workspace are recognized as defined instances. The data on their motion is streamed in real-time at 500Hz. Data acquisition is performed with the use of Vicon DataStream SDK (used version is 1.8.0). SDK can be used for the implementation of data streaming with different programming languages and operating systems. In this case, we used the C++ version of libraries for Windows operating system.

StreamCPP application establishes a connection with the PC that runs Vicon Nexus software through defined IP address. After the connection is successfully established, StreamCPP iterates through identified objects and subjects, enumerates them and extracts the information on the number of markers, and their assigned names. The main part of the application is a loop that is constantly updating the acquired frame (data package in specific time instance) streamed from Vicon Nexus. Acquired data are written in XML files („subject1.xml", „subject.xml", etc.). Another functionality of StreamCPP is to stream the acquired data to the LSL network for the purpose of synchronization with gaze tracking systems. Thus, for that purpose, StreamCPP creates LSL outlet and streams acquired data for each frame with an assigned time stamp from the Vicon system as well as LSL timestamp.

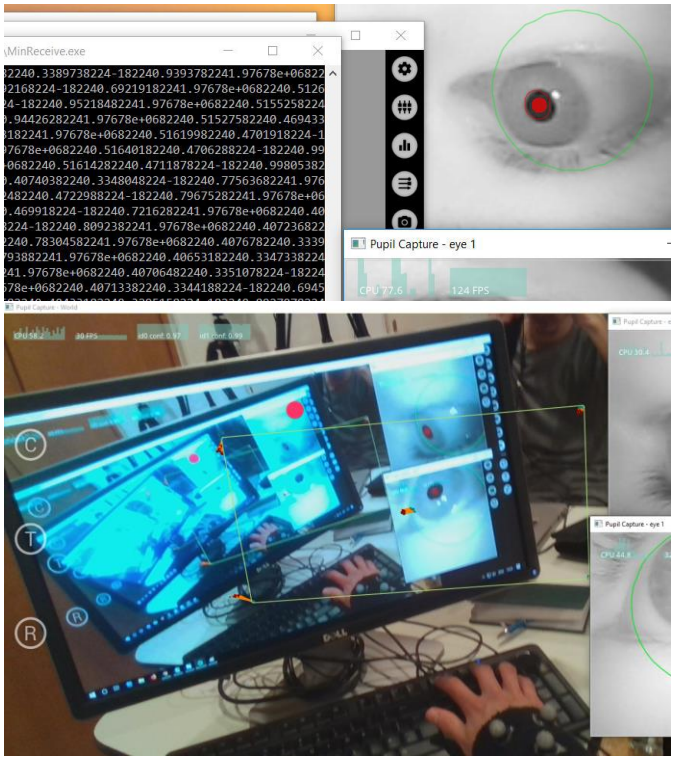

Fig. 4. Snapshots of the two eyes capture and egocentric view during data streaming

Gaze tracking and streaming are performed with Pupil Capture (Fig. 4). It captures eye movements with two infrared cameras and matches the eyeball, iris and eye gaze direction to the eyes model for estimation of gaze in 2D egocentric view camera and in 3D. Comparing to Vicon Nexus, Pupil Capture application does not provide data streaming directly with the installation. For this purpose, we used a plugin Pupil LSL Relay (pylsl) that enables streaming of the eye gaze data on LSL network. Pupil LSL Relay plugin opens outlet for each glass that is on LSL network. To acquire gaze-tracking glasses on LSL network, ReceiveData python script is developed. It writes the data to an XML file in the same format as in the case of StreamCPP application for motion capture data.

In order to synchronize all the streams, a Transferscript python script is developed. It is being called in each iteration of StreamCPP. Transferscript checks the total number of pupilOutput.xml files to determine the number of active gaze-tracking glasses. It also merges the timestamps of Vicon motion capture with Pupil-Labs glasses. That way, time synchronization of different sensing devices is performed[1].

III. Interaction Experiment Setup

The purpose of this experimental setup is to prepare and conduct the human-human interaction experiment for evaluating the necessity of eye contact between pairs of subjects in a scenario which involves co-workers randomly picking objects from a box manipulating and placing or handing over to a co-worker. An opaque blind is placed between co-workers with three different visibility conditions: first, the subjects can't see each other at all, next they can see up to shoulder height (head motion and eye gaze is occluded) and lastly the blind is removed completely. This is done with the intent of finding out how eye contact and/or arm motion visibility influence the subjects' choice of object picking, asynchronous request for involvement in interaction and frequency of collisions when picking the objects.

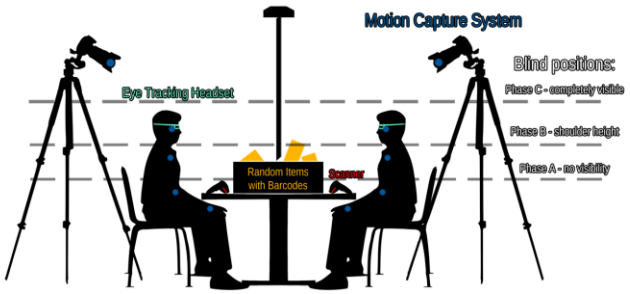

Fig. 5. Sketch of three different phases of the human-human interaction experiment

Pairs of subjects are seated facing each other, with a desk between them. On the desk is a box containing a number of items (product packaging with barcodes). Both subjects (simultaneously) pick out items from the box, scan the item's barcode with a scanner and then place it aside or handover it to a co-worker. The decision between placing and giving the item is made on the parity of bar code number (odd number is for placing and even for giving the item). The scanning in itself is significant as an additional step (action) for the subjects so that they have to focus on item manipulation and extract information on their task. The subjects are equipped with head-mounted eye trackers and surrounded by a motion tracking camera system which tracks markers placed on their arms and on eye tracking glasses. The box contains 24 items, 12 with odd and 12 with even bar code numbers. The experiment involves a total of 18 people divided into pairs. Each pair will be involved in one out of three visibility conditions of the experiment.

In our previous work [2, 3] we used a similar setup to study the importance of eye gaze for understanding and anticipating the interaction and to develop gaze behavior model to control the eye gaze and arm movements of the robot. The experiment involved questioner given to human subjects to anticipate the action that the actor will perform based on different available cues. Here we are directly disturbing the interaction with the cover placed in between the co-workers and we will be able to better understand the role of different non-verbal cues in human-human interaction and transfer that knowledge to the models for controlling humanoid robot co-worker.

A. *Procedure for experiment data acquisition*

At the beginning of the experiment, two subjects are first equipped with gaze tracking glasses and markers placed on their arms. Before each data acquisition, first, it is necessary to calibrate all the systems, i.e. motion capture, as well as two gaze-tracking glasses. Next step requires the creation of the subjects to match the body parts of the experiment

[1] All the scripts and projects developed are accessible on the GitLab repository on the following link: https://gitlab.com/vicon-pupil-data-parser

participants. Vicon Nexus and Pupil Capture have to be validated if they are recognizing the subjects and are properly estimating the gaze. In case of to many false measurements, the calibration procedure should be repeated. Once the motion and gaze tracking are reliable, the internal recording (in Vicon Nexus and Pupil Capture) can start. After starting the internal data recording, all the scripts and application for LSL data streaming and synchronization has to be started. It is necessary to check if all the devices are sending streams on LSL network. When all the data are streamed and recorded properly, the experiment can start. Once the experiment is finished, all applications and scripts have to be stopped, and it is necessary to check if all the output files are properly saved and closed.

IV. Conclusion

In this work, we presented the experimental setup for acquiring a multimodal dataset containing eye gaze and body motion during human-human interaction. The setup involves a Vicon Motion tracking system and two Pupil-Labs gaze-tracking glasses. The acquired data are saved in raw format in manufacturer's recording software. For synchronizing the data obtained from different measuring equipment we developed applications and scripts that are streaming the data to LSL network and capture that data at the same place where timestamps of different data sources are matched. Our next steps will involve preparation of a dataset for the experimental setup that is described in Section III. that includes annotating specific events and fixations during interaction so that temporal correlations between important events can be modeled. Further on we will focus on improving the models for upper body humanoid robot control based on the derived model.

Acknowledgment

This work was partially supported by MNTR project III44008 and TR35003, EU H2020 project ACTICIPATE and FCT project RBCog-Lab research infrastructure.

References

[1] M. Gallotti, M.T. Fairhurst, C.D. Frith, "Alignment in social interactions", *Consciousness and cognition*, 48, pp. 253-261., 2017

[2] N. F. Duarte, M. Raković, J. Marques, J. Santos-Victor, "Action Alignment from Gaze Cues in Human-Human and Human-Robot Interaction" In Proceedings of the European Conference on Computer Vision (ECCV), 2018.

[3] N. F. Duarte, M. Raković, J. Tasevski, M.I. Coco, A. Billard, J. Santos-Victor, "Action anticipation: Reading the intentions of humans and robots", *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 4132–4139, Oct 2018. https://doi.org/10.1109/LRA.2018.2861569

[4] D. Preissmann, C. Charbonnier, S. Chagué, J.P. Antonietti, J. Llobera, F. Ansermet, P. J. Magistretti, "A motion capture study to measure the feeling of synchrony in romantic couples and in professional musicians.", *Frontiers in psychology*, vol. 7, 2016.

[5] E. van der Kruk, M. Reijne, "Accuracy of human motion capture systems for sport applications; state-of-the-art review.", *European journal of sport science*, nol.18, no. 6, pp. 806-819, 2018.

[6] G.E. Robertson, G.E. Caldwell, J. Hamill, G. Kamen, S. Whittlesey, "Research methods in biomechanics", Human kinetics, 2018

[7] C. Mandery, Ö. Terlemez, M. Do, N. Vahrenkamp, T. Asfour, "The KIT whole-body human motion database", *In 2015 IEEE International Conference on Advanced Robotics (ICAR)*, pp. 329-336, 2015

[8] B. Burger, A. Puupponen, T. Jantunen, "Synchronizing eye tracking and optical motion capture: How to bring them together.", *Journal of Eye Movement Research*, vol. 11, no. 2, 2018.

[9] Qualisys. http://www.qualisys.com

[10] Vicon. http://www.vicon.com.

[11] Optitrack, https://optitrack.com/

[12] A. Duchowski, "Eye tracking methodology", Springer, 2007.

[13] M. Kassner, W. Patera, A. Bulling, "Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction", In: Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing: Adjunct publication, ACM, pp. 1151–1160, 2014

[14] L. Bishop, W. Goebl, "Mapping visual attention of ensemble musicians during performance of "temporally-ambiguous" music." *In Conference of Music & Eye-Tracking*. 2017.

[15] F. Marandola, "Eye-Hand synchronization and interpersonal interaction in xylophone performance: A comparison between African and Western percussionists", *Journal of Eye Movement Research*, vol.11, no.2, 2019,

[16] D.L. Mills, "Internet time synchronization: the network time protocol", *IEEE Tran. on communications*, Vol. 39, no. 10., pp 1482-1493, 1991

[17] C. Kothe, "Lab streaming layer (LSL)", Accessed on April 26., 2019., https://github. com/sccn/labstreaminglayer