Evaluate Robot’s Memory via Human Demonstration

Cartosio Luca, De Mari Lorenzo, Grattarola Alessandro, Robers Maren, Ventura Alessandra

***Abstract*— Two datasets, along with the design of the experiment used to obtain them, has been developed to be used for cognitive-like memory in support of Human Robot Cooperation. The aim is to produce a planning based on human demonstration. For the acquisition of the datasets, 32 people took part in the experiment, which consisted in a simple task: mounting a table. At the end, the human-like memory has been tested and it produced a graph that can be converted into an instruction manual as to accomplish the task based on 100 table configurations that the volunteers showed to the robot.**

I. Introduction

In support to human robot collaboration, a bootstrapping[1] phase is usually considered as an initial interaction for teaching to the robot actions and knowledge that it will requires in future tasks.

For having an effective interaction, the robot and users needs to understand each other and this limit the deployment of Machine Learning techniques[2].

Nevertheless, a state of the art imitation learning algorithm[3] has been used for learning qualitative *scenes*, and build a structure of their similarities[4], which the robot can explain to a supervisor[5]. In particular, this algorithm performs a learning of scenes defined as relationships between objects and it generates a graph structuring scenes, that we want to use with task planning purposes.

We considered the graph as a cognitive human-like memory collecting the experience (i.e., scenes) that the robot made from demonstration. Remarkably, it has also a forgetting behavior which assures that new scenes can be learned with a limited computation complexity, and it allows to use the system in a continuous learning scenarios. In particular, we implement such a mechanism with a score associated to each scene contained in the memory that changes based on the interaction that will make the robot consolidating some configurations more than others.

We want to investigate about experience consolidation, retrieving and forgetting in an efficient manner in order to use our system in long-term human interaction scenarios. In this project, we designed an experiment for collecting data to perform a preliminary evaluation of the memory of the robot after the observation of human demonstrations. With our method it could be possible to confirm that from the memory it is possible to generate plans that accomplish the task without prior knowledge about it, but knowing only objects and relations that can be observed. We considered an Ikea-like table assembling scenario similar to[6], but we focused on a setup that encompass different possibilities for assembling the table. The idea is to generate a robot behavior that is based on different possible demonstration of doing the same task.

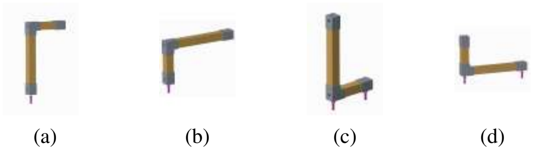


Fig. 1: **Possible configurations of the leg.**

We designed the experiment and collect data for having a ground truth that can be used for compare different techniques for computing the attentive score of the scenes in the robot memory. In particular, our experiment needs to present some precise characteristics. For instance, it needs to be simple and it must produce a dataset where many different configurations are present. This is achieved by making users assembling a table with four identical table legs shaped as an "L".

This way each leg can be connected to the table platform in 4 different configurations, as shown in **Figure 1**. In this scenario, the legs and the pins through which they are attached are considered as objects and are described by the symbolic spatial relationship `connected`. We collected 2 datasets: one with raw perception data and one with symbolic data, elaborated in such a way that the software can process the scene.

II. Material & Methods

A. Experimental Set-up

The dataset has been acquired under the ROS framework (kinetic version for Ubuntu 16.04). This choice has been guided by the possibility of recording data under the *bag* file format, for which ROS provides a simple API. The main characteristic of ROSbags is that they can be easily reproduced for simulation.

The camera device used for recording the experiment is a Microsoft kinectv2 equipped with two cameras, one RGB and one IR, and images have been acquired at 1 Hz.

The bridge between camera drivers and ROS is provided by the ROS package *iai_kinect2*[1] that also offers a tool for calibration.

The table to be assembled is shown in **Figure 2**. It can be noticed the presence of pins that allow to fix the legs in 12 different positions. Those pins only consent to place the leg at 90°, so four possible orientations of the leg are possible.

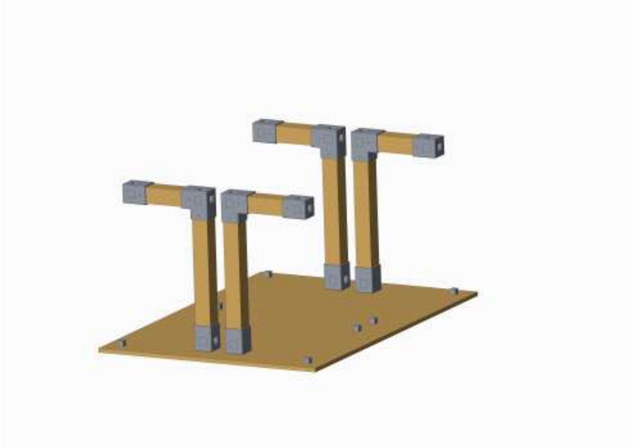


Fig. 2: **Table.** Legs positioned in one possible configuration.

To acquire information about position and orientation of legs the ROS package *ar_track_alvar*[2] has been used. This package consents to track objects with the help of some specific markers. A group of markers (*bundle*) has been attached to each leg, so that, however the leg is positioned, at least one marker can be seen by the camera. This prevents occlusions during the test: when the camera detects one or more markers belonging to the group, a frame is associated to the leg. The same method has been applied to the table platform as to provide a world reference frame that does not depend on the position of the camera.

Using the *tf*[3] package available on ROS, the transformation matrix describing the relation between the world reference frame and each leg's frame has been computed. For the visualization of the frames *RViz* has been adopted (**Figure 3**).

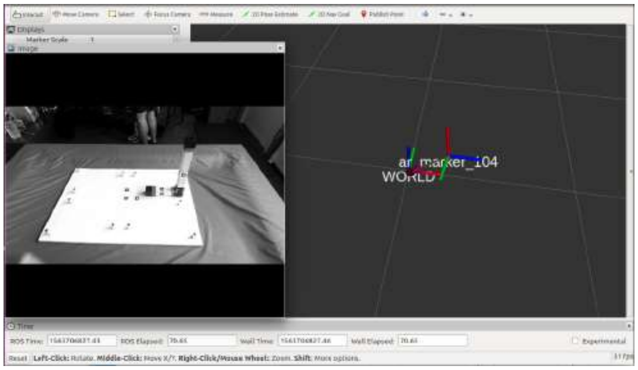


Fig. 3: Example of visualization

B. Table Assembling and Root Observations

For the experiment, the kinect and the table are first positioned as in the figure below **(Figure 4)**.

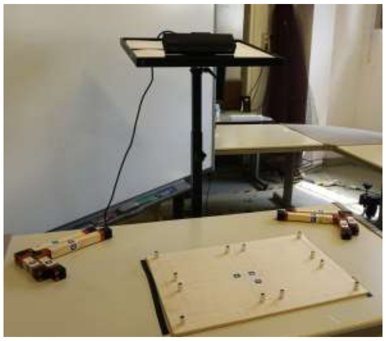


Fig. 4: Experiment setup.

The volunteer has been instructed to place one leg at a time on a pin of their choice as to create a table which could stand stable. An example of assembling session can be seen in **Figure 5**.

Each volunteer has been asked to do the experiment three times, without repeating the configurations.

The number of volunteers that took part in the experiment is 32, for a total of 96 samples. Before the test, everyone was given the same information about the experiment and, in order to avoid any possible bias, they were not aware of the final purpose of the observation. A brief questionnaire was also given to each person, asking, anonymously, information such as age, height and dominant hand.

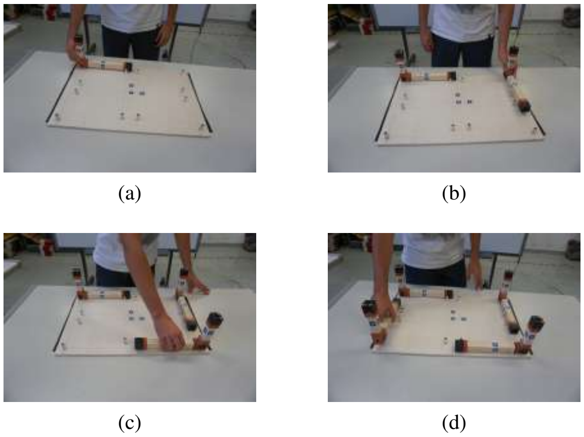
C. Data acquisition and elaboration

The data, as mentioned before, have been acquired using ROSbags.

[1] https://github.com/code-iai/iai_kinect2

[2] http://wiki.ros.org/ar_track_alvar

[3] http://wiki.ros.org/tf



(a) (b) (c) (d)

Fig. 5: **Example of experiment.**

In the first dataset information about the transformation matrices and the video of the experiment have been stored. Notice that from the video it is not possible to identify the volunteer.

A ROS node has been created for data elaboration. From the first dataset, orientation has been used to associate a configuration to each leg and position to identify the pin to which the leg was connected. The computed configuration, along with the leg identifier and the corresponding pin, has been published via custom ROS message on a topic. This information is easily readable and has been very useful for checking the behaviour of the architecture.

The mentioned node performs as publisher on another topic, for which the message is structured as required from the human-like memory software. Those messages are stored in the second dataset.

The creation of a second node has been implemented as interface. This turned out to be necessary since the simulated memory requires a service-client communication, thus the node is blocked during interchange. In practice this means that the same node could not manipulate data and communicate with the software.

The architecture is reported in **Figure 6**

III. Results

The main results of the project are the guidelines for the design of an experiment aimed to test a software architecture. From this experiment the raw dataset and the elaborated dataset have been collected and are now available for future works.

Raw data consist of angles yaw, pitch, roll and position x, y, z of the legs' frames with respect to the world. Elaborated data consist of a series of messages describing each scene. The scene consists of a collection symbols representing legs and pins that are considered to be connected if the euclidean distance (based on their frame) is below 3cm.

From the collected data emerges that the most common configurations are the symmetrical ones.

IV. Discussion

A. General discussion

The decision of asking each person to do exactly three different configurations has been taken as to avoid to collect unusable data. The risk was to have one instance of each configuration or, even more dangerously for the richness of the dataset, all instances of the same scene. More than three configurations, on the other hand, may have been difficult to find.

Before doing the tests the assumption was that most part of the volunteers would do at least one configuration that would be forgotten by the software, and at least one that would be remembered.

What happened in practice is that each volunteer reacted to the test either trying to find the most simple configurations or trying to find the most original ones. However, since the dataset is aimed to test the memory software this discrepancy between expectation and reality does not infer the quality of the data.

An important point to be considered is the meaning of remembering and forgetting. When can something be forgotten? It depends on the planning problem. The definition of a threshold must be tuned by trial and error.

B. Dataset developing guidelines

The designed experiment is intuitive, easy to perform and it takes more or less 5 minutes to complete. This way, it is simpler to have an acceptable number of volunteers and the experiment session is quite fast.

In addition, during the experiment it has been noticed that volunteers were quite fascinated by the test since it was perceived as a challenge, and this helps acquiring consistent data also when users assembly the table in unusual configurations.

Another strong characteristic of the project is that the experimental set-up is robust to the noise produced by the volunteer: the markers are placed on the leg in such a way that hardly all of them will be covered by the volunteer's movements. Moreover, if the leg is not attached to a pin

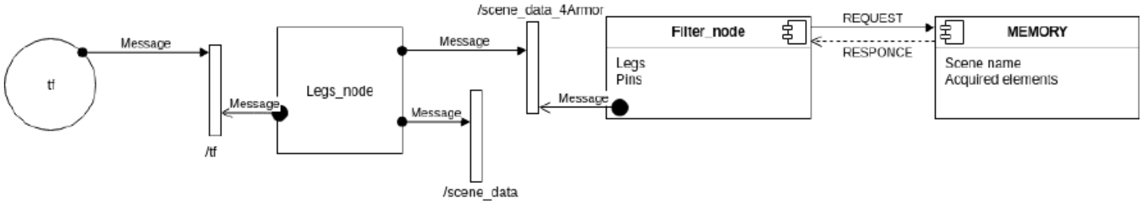


Fig. 6: Architecture.

the information about the leg is disregarded. In addition, even if the platform is slightly displaced during the experiment, the world reference frame adapts to such movement and the data are correctly recorded anyway.

One of the strong points of the architecture is, in fact, its modularity. This could allow to easily change part of the architecture without replacing all the modules that have been developed, provided that the dataset is not used as it is.

The fact of making humans do the experiments instead of producing a synthetic dataset is an important point of the project. This way the configurations have a random distributions but are all stable.

C. Limitations

One of the main limitations of the experiment regards the positioning of the camera and the illumination of the table. In fact, in some conditions the perception part fails to recognize the marker. For this reason, before starting data collection, a thorough tuning session has been executed, mostly by trial and error.

Perception data are not perfectly stable. The error is more evident for the angle orientation, whilst the position's accuracy can be considered good. The overall noise, however, can be believed to be admissible since the angles of interest are discretized at 90° and the tolerance on the position is about 3 cm, half the distance between the closest pins.

Another limitation of the project can be considered to be the small diversity of volunteers that took part in the experiment. Maybe the same test executed by a different sample would have given different results.

V. Conclusion

In conclusion, the dataset resulted being good enough to test the behaviour of the simulated memory.

Even though the project could seem very much oriented to one specific task this is not completely true. The method presented in this document could be easily followed to design similar experiments for different memories.

The final product, which consists in the simulated memory, could be seen as a way to train robots in executing tasks correctly. The advantage in using this method among the others available in literature is that the amount of data used for obtaining good results is not necessarily high.

In the future the same architecture could be adjusted to be working with more precise sensors. Even multi-view methods could be used for the acquisition of raw data, significantly reducing the noise of perception. At this regard, one thing that could be tested is an alternative recognition method, replacing the package *ar_track_alvar* with a different one.

In conclusion, even though the experiment and the datasets themselves are not perfect, they offer a good starting point for further studies. With the exact same setup, for example, a more guided experiment could be developed. In fact, it would be interesting to study how the volunteers would react when, instead of starting with an empty platform, one leg was already placed.

References

[1] Florentin Wörgötter et al. “Structural bootstrapping—A novel, generative mechanism for faster and more efficient acquisition of action-knowledge”. In: *IEEE Transactions on Autonomous Mental Development* 7.2 (2015), pp. 140–154.

[2] David Gunning. “Explainable artificial intelligence (xai)”. In: *Defense Advanced Research Projects Agency (DARPA), nd Web* 2 (2017).

[3] Aude Billard et al. “Robot programming by demonstration”. In: *Springer handbook of robotics* (2008), pp. 1371–1394.

[4] Luca Buoncompagni, Fulvio Mastrogiovanni, and Alessandro Saffiotti. “Scene learning, recognition and similarity detection in a fuzzy ontology via human examples”. In: *arXiv preprint arXiv:1709.09433* (2017).

[5] Luca Buoncompagni and Fulvio Mastrogiovanni. “Dialogue-based supervision and explanation of robot spatial beliefs: a software architecture perspective”. In: *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE. 2018, pp. 977–984.

[6] Luca Buoncompagni et al. “From Collaborative Robots to Work Mates: A New Perspective on Human-Robot Cooperation”. In: *ERCIM NEWS* 114 (2018), pp. 8–9.