On-Line Simultaneous Learning and Recognition of Everyday Activities from Virtual Reality Performances*

Tamas Bates[1], Karinne Ramirez-Amaro[1], Tetsunari Inamura[2], Gordon Cheng[1]

***Abstract*— With the recent rise of virtual reality head-mounted displays and consumer-grade GPUs capable of high quality rendering to stereo displays, we believe virtual reality (VR) is a viable way to collect real information about human behavior without the difficulties often associated with capturing natural performances in a physical environment. For example, estimating whether someone is grasping an object or simply touching it. We present a VR environment for experimentation with household tasks, like washing dishes or doing laundry, paired with a semantic extraction and reasoning system able to utilize data collected in real-time to learn new activities from a human demonstrator. The learning system performs continuous segmentation of the motions of the user's hands and simultaneously classifies known actions while learning new ones on demand. This enhanced system produced more accurate results over previous VR-based training systems, improving the recognition of activities from 80% to 92% while learning unknown activities from more complex and realistic scenarios. The learning system then constructs a graph of all observed activities and their relationships through continuous observations. The resulting activity and task data is abstract enough to allow for easy knowledge transfer from the VR learning environment to a physical robot and is still detailed enough to be useful for the robot's planning process, which was verified by transferring knowledge from several VR training sessions to a PR2 robot. The robot was able to utilize the information learned in VR to carry out multi-step tasks without requiring every step be explicitly given in the instructions.**

I. Introduction

To improve human-robot interaction and the understanding of how to behave in human spaces, it is important to train robots with examples of real human behavior. Constructing physical training environments is often difficult and costly (e.g. embedding accelerometers in objects which can moved, or tagging all the objects with visual markers and calibrating a tracking system covering the space), and there are safety issues which need to be addressed if the space is shared between both humans and robots. One solution is the use of virtual reality (VR) environments, which allow for fast implementation of different scenarios[1]. The system presented in [2] provided improved realism and accuracy over a previous VR-based training system[3], but one of the key challenges in utilizing VR environments is ensuring that information learned by a virtual agent can be transferred to a physical robot and applied in the real world. Building on the work in [2], an experiment was carried out to evaluate the utility of the information learned in VR for a physical robot and to determine whether it was sufficient to aid in a physical robot's task planning process.

*This work has received funding from the European Communitys Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 609206. This work was partially supported by CREST, JST, and the NII International Internship program. Partially funded by the Collaborative Research Center, SFB 1320 Everyday Activity Science and Engineering.

[1]Institute for Cognitive Systems, Technical University of Munich, Arcisstraße 21, 80333 Munich Germany {t.bates, karinne.ramirez, gordon}@tum.de

[2]National Institute of Informatics, 2 Chome-2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 100-0003, Japan inamura@nii.ac.jp

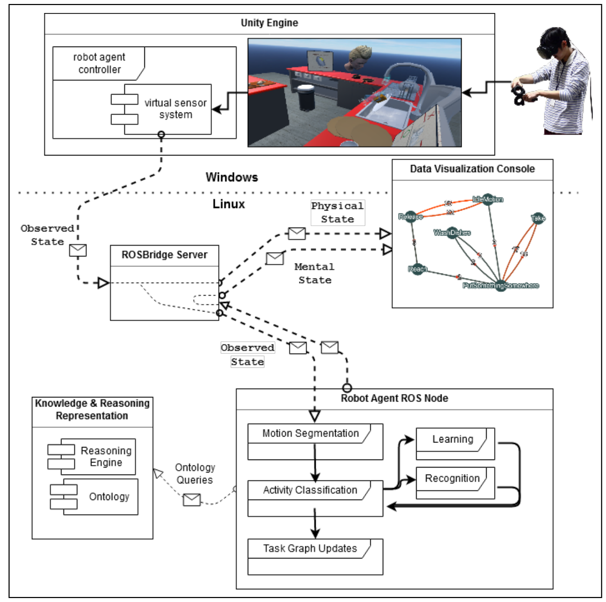

Fig. 1. VR Training Overview. Users demonstrate activities in VR, motion & object information goes to a ROS Node to classify the user's actions.

Several users were recorded performing tasks related to dish washing in VR. Fig 1 shows an overview of the system used to track and classify their activities. This system produces a set of observed activities and the ontology classes of the objects used in each (using the default KnowROB[4] ontology), as well as a graph of the task space explored by the participants. In the task graphs, each node represents an activity, edges represent observed transitions between activities, and edge weights store the number of observations of each transition. Over the course of many observations, the resulting graph represents the task space utilized by observed users during their performance in VR. These results were given to a PR2 robot, which was then commanded to perform some simple tasks related to dish washing.

II. Method and Results

The same motion segmentation and classification system ('Robot Agent ROS Node' in fig 1) used to analyze human motions in VR was used, without modification, on the PR2 to allow the robot to estimate its own state. This enabled

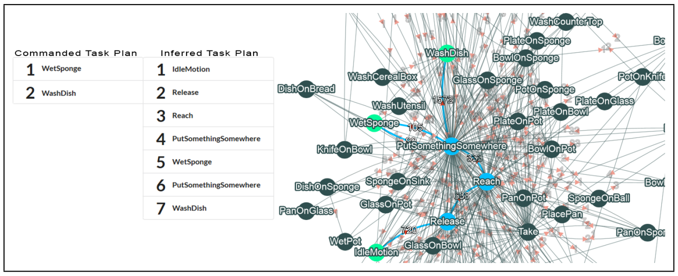

Fig. 2. A generated task plan. Green nodes were in the original command, blue nodes were inferred from the structure of the Task Graph.

the robot to understand how its current actions related to the overall task space, and to know which activities could be performed from its current state. A simple motion planning module built around a state machine was implemented as a proof of concept, with states corresponding to activities and a transition function ensuring the prerequisites for each activity were met (e.g. closing the robot's gripper to transition from Reaching for an object to Taking the object). Complex activities (e.g. Washing) were assigned predefined motions, developed with the aid of MoveIt![5], which were parameterized by the objects involved and associated with the corresponding state in the state machine.

Using the graph of activity relationships, the steps necessary to perform a specific activity can be determined by finding a path from a node matching the robot's current state to one matching the desired activity. Given sparse or incomplete instructions, the robot can then construct a complete task plan by finding paths between each sequential pair of steps in the instructions, as shown in fig 2. Paths with the greatest number of observed transitions between nodes were selected, so the robot would take the most frequently-observed sequence of actions.

This was tested by presenting the robot with an assortment of objects (shown in fig 3), some of which were related to dish washing activities and some which were irrelevant. Given a set of instructions like, ”wash a dish then store it in the drying area” the robot first constructed a plan containing the complete set of necessary steps for the requested task (e.g. reach for and take a sponge, move the sponge to the dish before performing a washing action, etc.), and using the activity definitions learned from the VR training sessions identified the objects to be used by matching their ontology definitions to constraints in the activity definition (e.g. if an activity specifies an object with the ontology class ”FoodVessel” should be used, the robot would be free to choose a bowl or plate, but should not choose a sponge or fork) . As long as the activities in the instructions were in the robot's knowledge base (i.e. they had been learned through observations in VR) the robot was able to successfully carry out tasks regardless of the number of steps involved or whether the task required performance of additional activities not explicitly included in the instructions.

A demonstration of PR2 utilizing the learned data is included in the video from [2]:
https://youtu.be/feF13VZ-rew

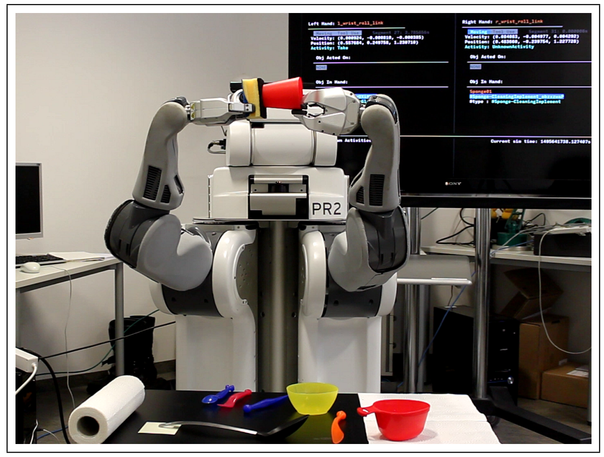

Fig. 3. PR2 ”washing” a dish.

III. Conclusion

The activity definitions and task graphs from individual trials produced by the system described in [2] can be merged through a simple process, allowing for many trials to be aggregated into a single data set describing the demonstrated activities and their relationships in the overall task space. The activity definitions provide a straightforward method of identifying which objects in the environment can be used for an activity, and the task graphs make an effective data structure for high-level task planning (i.e. identifying which activities should be performed, when they should be performed, and which objects should be used). A complete solution also requires a good sensory system (to locate and classify objects in the environment) and a robust motion planning module (to carry out the necessary physical movements), but for the problem of planning this system functions well, even when given incomplete instructions.

References

[1] T. Inamura, T. Shibata, H. Sena, T. Hashimoto, N. Kawai, T. Miyashita, Y. Sakurai, M. Shimizu, M. Otake, K. Hosoda, S. Umeda, K. Inui, and Y. Yoshikawa, “Simulator platform that enables social interaction simulation SIGVerse: SocioIntelliGenesis simulator” in IEEE/SICE Int. Symposium on System Integration (SII), Dec 2010, pp. 212217.

[2] T. Bates, K. Ramirez-Amaro, T. Inamura, G. Cheng. “On-Line Simultaneous Learning and Recognition of Everyday Activities from Virtual Reality Performances.” Intelligent Robots and Systems (IROS 2017), 2017 IEEE/RSJ International Conference on. IEEE, 2017.

[3] K. Ramirez-Amaro, T. Inamura, E. Deal-Leon, M. Beetz, and G. Cheng. “Bootstrapping humanoid robot skills by extracting semantic representations of human-like activities from virtual reality” Humanoid Robots (Humanoids), 2014 14th IEEE-RAS International Conference on. IEEE, 2014.

[4] M. Tenorth, and M. Beetz. “KnowRob: A knowledge processing infrastructure for cognition-enabled robots” The International Journal of Robotics Research 32.5 (2013): 566-590.

[5] Ioan A. Sucan and Sachin Chitta, MoveIt!, [Online] Available: http://moveit.ros.org