Making Virtual Pancakes — Acquiring and Analyzing Data of Everyday Manipulation Tasks through Interactive Physics-based Simulations

Lars Kunze[1], Andrei Haidu[1], and Michael Beetz[2]

[1] Intelligent Autonomous Systems, Technische Universität München, Germany
kunzel@cs.tum.edu
[2] Artificial Intelligence, University of Bremen, Germany
beetz@tzi.de

Abstract. Teaching robots everyday tasks like making pancakes by instructions requires interfaces that can be intuitively operated by non-experts. By performing manipulation tasks in a virtual environment using a data glove task-related information of the demonstrated actions can directly be accessed and extracted from the simulator. We translate low-level data structures of these simulations into meaningful first-order representations, called timelines, whereby we are able to select data segments and analyze them at an abstract level. Hence, the proposed system is a powerful tool for acquiring examples of manipulation actions and for analyzing them whereby robots can be informed how to perform a task.

1 Introduction

In their daily routines personal robot assistants are supposed to accomplish novel tasks for which they have not been pre-programmed in advance. In [6], it is demonstrated how robots can extend their task repertoire by extracting natural language step-by-step descriptions from the Web and translating them into well-defined executable plans. For example, the instructions for making a pancake read as follows: 1) pour the pancake mix into a pan, 2) flip the pancake using a spatula, 3) place the pancake onto a plate.

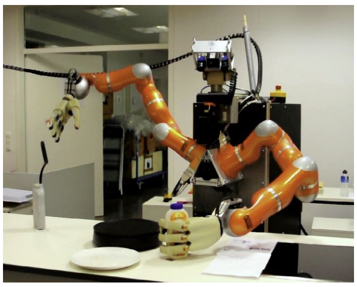

Fig. 1. Rosie preparing a pancake.

These instructions are descriptive enough for humans to understand the task. However, for robots these instructions are highly under-specified. That is, a robot has to infer the appropriate parameters of these actions by other means. By observing humans performing the task the robot can estimate some of the missing parameters. For example, the robot could estimate parameters like height and angle of the container while the pouring action is performed. Also the duration of this action could be estimated. Such information could be extracted from instruction videos retrieved from

S. Wölfl (Ed.): Poster and Demo Track of the 35th German Conference on Artificial Intelligence (KI-2012), pp. 140-144, 2012.

the Web or from a human tracking system [2]. Since our goal is to acquire a deep understanding of the physical effects of such manipulation actions, we propose a virtual manipulation environment based on a physics-based simulation. Objects within this virtual environment can be manipulated using a data glove and a 3D position sensor where the sensor information is directly translated into a pose and articulation of the simulated hand model. Since we have complete knowledge about the simulated world state we are able to extract different kinds of information of the task-related objects. These information include, for example, an object's position, orientation, linear and angular velocities as well as its bounding box. Also contacts between objects are reported in each time step. In contrast to vision-based systems we do not have to deal with occlusions and other typical problems like the recognition of transparent objects. The virtual manipulation framework, that we have designed and implemented, can be used as a tool for the acquisition of task-related information by logging the internal states of the simulator. The logged simulations are then translated into interval-based first-order representations, called timelines, as described in [5]. By formulating logical queries we can extract task-related information from these timelines semantically. For example, we can ask for the series of poses of the container while it was held in the hand. Then, further methods can be applied on the trajectory data to analyze the manipulation action with respect to various aspects.

2 Virtual Manipulation Environment

The virtual environment is based on Gazebo[3], a 3D multi-robot simulator with rigid-body physics. In the environment a user wearing a data glove controls a robotic hand which allows him/her to interact with various objects. Figure 2 show the hardware equipment, a user controlling the robot and a screenshot from the virtual environment. The virtual robotic hand (DLR/HIT) has four fingers with four joints, except the thumb which has an extra degree of freedom for easier manipulation. The hand is controlled with the help of a proportional-integral (PI) force controller acting on the wrist. For easier control the gravity acting on the hand is disabled. The data glove we use (X-IST Dataglove) is equipped with 15 bend sensors (three per finger, one for each joint). To get the pose of the

[3] `http://gazebosim.org`

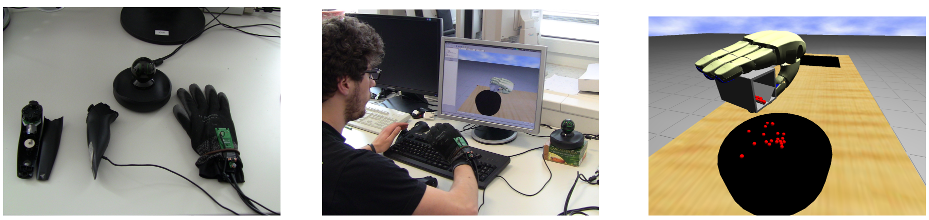

Fig. 2. Virtual Manipulation Environment.

hand within six degrees of freedom we use Razer Hydra, a game controller using a weak magnetic field to detect its absolute position and orientation. The sensor was disassembled from the game controller and attached to the data glove.

3 Preliminary Experimental Results

A user performed two tasks related to the *pancake* scenario: pouring some mix onto a pancake maker and flipping a pancake. We have monitored and logged the data structures of the simulator and translated them to first-order representations (timelines). Figure 3 illustrates steps from both tasks.

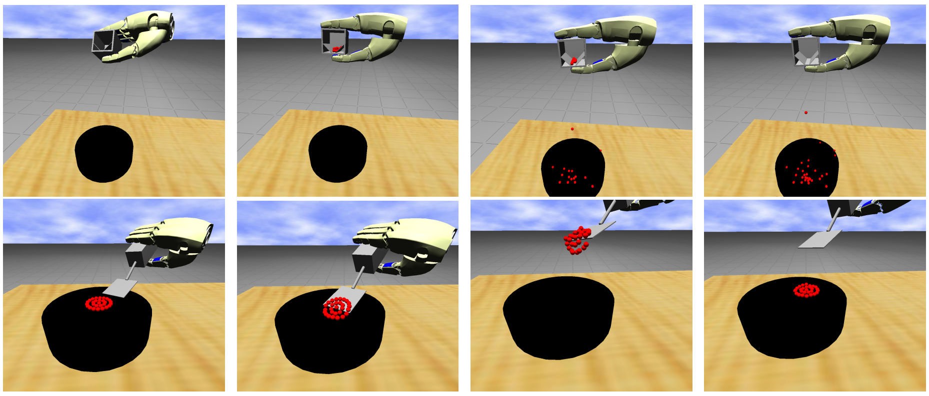

Fig. 3. Virtual Manipulation Tasks: Pouring liquids and flipping a pancake.

By translating the data structures of the simulator into timelines we can use first-order logic to query task-related data semantically. We access the timelines by using predicates similar to those in the Event Calculus [3]. The notation is based on two concepts, namely fluents and events. Fluents are conditions that change over time, e.g., a mug contains a pancake mix: *contains(mug,mix)*. Events (or actions) are temporal entities that have effects and occur at specific points in time, e.g., consider the action of pouring the mix from the mug onto the pancake maker: *occurs(pour(mix,mug,pancake_maker))*. Logical statements about both fluents and events are expressed by using the predicate: *Holds(f,t,tl)* where *f* denotes a fluent or event, *t* simply denotes a point in time, and *tl* a timeline. Using the predicate *Holds_tt* we can query for a time interval throughout the fluent holds. For example, we can ask for pose, velocities, and bounding box of the mug in a time interval where there was a contact between mug and the robotic hand as follows:

```
?- holds_tt(contacts(mug,hand),I,TL), simulator_values(position(mug,Ps),I,TL),
                                      simulator_values(orientation(mug,Os),I,TL),
                                      simulator_values(linear_velocities(mug,LVs),I,TL),
                                      simulator_values(angular_velocities(mug,AVs),I,TL),
                                      simulator_values(bboxes(mug,BBs),I,TL).
```

where I denotes a time interval, TL a timeline, and the other variables denote lists of their respective data types. Similarly, we can get the last pose of the mug in that interval, e.g., to analyze where the user has placed it after the pouring.

In the experiments liquid was poured from different heights which can be seen by clustering the trajectories (Figure 4). First, we applied dynamic time warping to align the trajectories and then we clustered the trajectories as in [1].

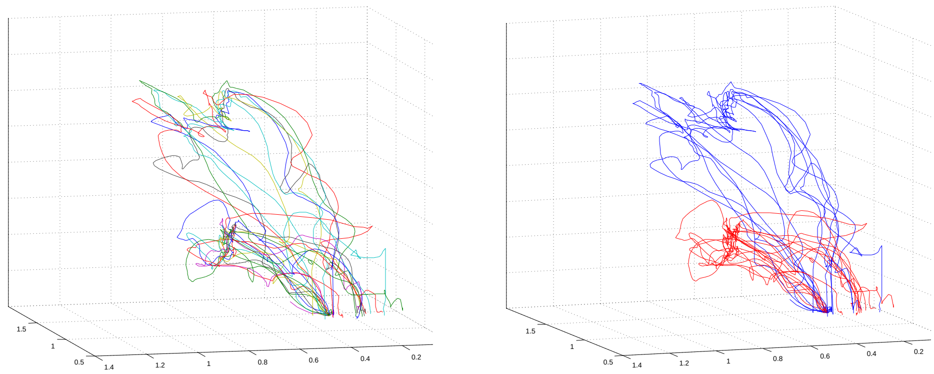

Fig. 4. Trajectories of the mug when it was in contact with the hand. Raw (left) and clustered (right) trajectories after aligning them using dynamic time warping.

Logical queries allow us to select data segments of the logged simulations on an abstract level. For example, we can select only data when the mug is over the pancake maker or when it is tilted at an angle in a certain range.

4 Conclusions and Future Work

In this paper we have presented a system for acquiring data of manipulation actions by controlling a robotic hand in a virtual environment using a data glove. By translating the data to timelines we are able to analyze and interpret the performed actions at a semantic level. In future work, we will deliberately tweak the underlying physics of the simulation to produce behaviors that deal with various physical phenomena such as the viscosity of liquids. We will also apply the found parameter values as seeds in our envisioning system for robots [4]. In the long run, we would like to integrate a vision-based tracking system using a physics-based simulation to acquire examples of manipulation actions more naturally.

Acknowledgments

This work has been supported by the EU FP7 Project RoboHow (grant number 288533) and the cluster of excellence Cognition for Technical Systems (Excellence Initiative of the German Research Foundation (DFG)).

References

1. S. Albrecht, K. Ramirez-Amaro, F. Ruiz-Ugalde, D. Weikersdorfer, M. Leibold, M. Ulbrich, and M. Beetz. Imitating human reaching motions using physically inspired optimization principles. In *11th IEEE-RAS International Conference on Humanoid Robots*, Bled, Slovenia, October, 26–28 2011.
2. M. Beetz, J. Bandouch, D. Jain, and M. Tenorth. Towards Automated Models of Activities of Daily Life. In *First International Symposium on Quality of Life Technology – Intelligent Systems for Better Living*, Pittsburgh, Pennsylvania USA, 2009.
3. R. Kowalski and M. Sergot. A logic-based calculus of events. *New generation computing*, 4(1):67–95, 1986.
4. L. Kunze, M. E. Dolha, and M. Beetz. Logic Programming with Simulation-based Temporal Projection for Everyday Robot Object Manipulation. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, San Francisco, CA, USA, September, 25–30 2011.
5. L. Kunze, M. E. Dolha, E. Guzman, and M. Beetz. Simulation-based temporal projection of everyday robot object manipulation. In Yolum, Tumer, Stone, and Sonenberg, editors, *Proc. of the 10th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2011)*, Taipei, Taiwan, May, 2–6 2011. IFAAMAS.
6. M. Tenorth, D. Nyga, and M. Beetz. Understanding and Executing Instructions for Everyday Manipulation Tasks from the World Wide Web. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 1486–1491, Anchorage, AK, USA, May 3–8 2010.