Specification of Physical Interaction through Vision and Force-based Demonstration

Mario Prats[1,2], Angel P. del Pobil[1,2] and Sukhan Lee[1]
[1]Department of Interaction Science, Sungkyunkwan University
Seoul, South Korea
[2]Computer Science and Engineering Department, University of Jaume-I
Castellón, Spain
Email: {mprats,pobil}@uji.es

***Abstract*— This paper shows work in progress towards a new approach for building an abstract task representation from a single human-guided demonstration: the *Specification by Demonstration* approach. Guided by a human instructor, a robot extracts a set of key task references and relates them to a visual model of the object. With this information, a physical interaction task representation is built and stored in a database for its future use. The robot makes use of visual and force feedback both during the teaching and on future autonomous operation. Preliminary experiments are performed using a mobile manipulator interacting with a dishwasher.**

I. Introduction

Future robot companions will have to work in human spaces and deal with objects that they have never seen before. Many of these objects will be home appliances such as diswashers, washing machines, TV sets, etc. that need a specific interaction procedure for its use. It is very likely that new robots introduced in our homes need a training period in order to teach them how to operate the different appliances. Therefore, methods for teaching physical interaction tasks are needed. In addition, it is very convenient that the interaction procedures learnt by one robot can be easily shared with others.

Most approaches to *Programming by Demonstration* (PbD) focus on task representations at the joint level [1], which cannot easily adapt to wide variations in the working scenario. In order to solve this problem, other approaches focus on abstract task representations that are independent of the robot configuration. Most of these works are based on qualitative descriptions of the robot environment, e.g. trying to reach a desired pose of the objects relative to each other, mostly for applications involving pick and place actions [3].

In this work we focus on building abstract task representations for interacting with home appliances through vision and force feedback. We introduce the *Specification by Demonstration* approach, based on our previous work on a framework for specification of physical interaction tasks (FPI) [4]. The main idea is to automatically build an abstract robot-independent representation of the task from a single user demonstration. First, the user introduces a new object to the robot and indicates a visual reference for its localization. Then, the user shows a new task on this object to the robot, by manually guiding the robot hand. After that, the robot reproduces the same motion by its own, and the user can either validate it or teach the task again. When the robot performance is validated by the user, a physical interaction task specification is built from the position, vision and force feedback logged during the teaching process. This abstract information is structured in an XML format and stored in a database for its future use, and for sharing it with other robots.

II. Human-guided task specification

The following steps are performed for showing a new task to the robot:

1) First, the user indicates the name of the object. If the object already exists in the database, the robot loads the tasks that have been already specified and the visual reference used for tracking the object. The user has the option to add a new task, modify an existing task, or specify a visual reference. The visual reference is currently specified by clicking on four points of a rectangular region with clearly visible edges. The robot then asks for the dimensions of the rectangular patch. With this information, a visual model is built and the patch pose is retrieved and tracked using the Virtual Visual Servoing method [2], [5]. By stereo visual processing, the need for specifying the patch dimensions could be avoided.
2) For teaching a new task on an object, the user first introduces the task name, and then guides the robot hand through the different steps of the task. This is done thanks to a force sensor placed at the robot wrist. During this step the robot stores the joint trajectories.
3) After that, the robot tries on its own and performs the task by reproducing exactly the same joint trajectories while logging the end-effector trajectory, the forces generated and the object pose. We assume that the object pose does not change between steps 2 and 3. The goal of this step is to log the interaction forces, which is not possible in step 2, due to the fact that the human guidance introduces forces that cannot be distinguished from the interaction forces.
4) From the end-effector trajectory, forces and object pose, a set of frames and velocity/force references relative to the object pose are set, and a XML file describing the task is generated. Everything is computed

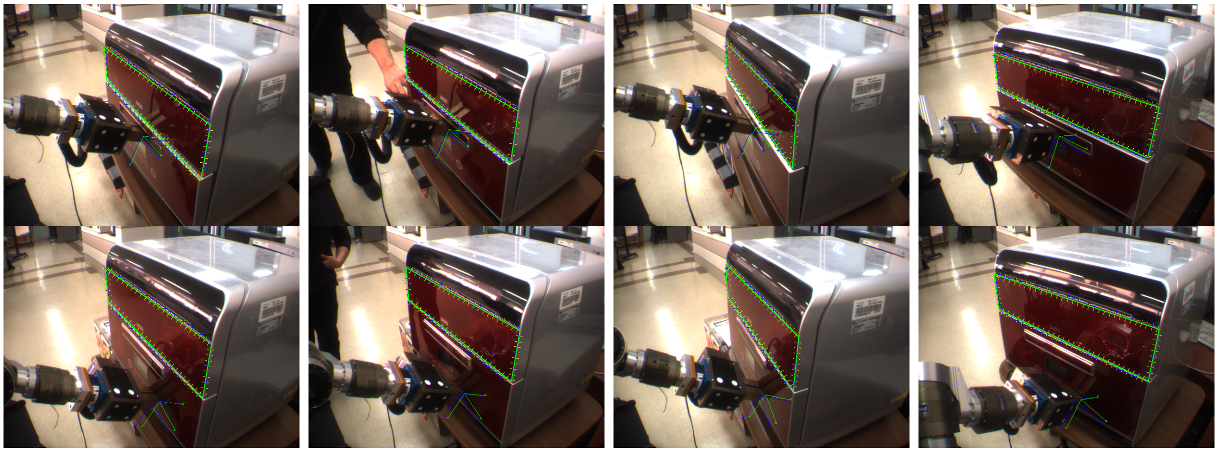

Fig. 1. The mobile manipulator reproducing the task under different orientations of the object. The top row shows the *opening* task, done by pushing on a large button. The bottom row shows the *switch on* task, that requires pushing a smaller button. The images show only the moment in which the robot makes contact.

relative to the visual reference pose, and, therefore, the same task can be always reproduced as long as the visual reference has been localized. The task is specified according to our FPI framework [4].

Below is an example of an XML file generated after showing the *switch on* task on the dishwasher of Figure 1:

```
<?xml version="1.0" ?>
<object>
  <name>dishwasher</name>
  <visualmodel>
    <point3d>0.000000 0.000000 0.000000</point3d>
    <point3d>0.540000 0.000000 0.000000</point3d>
    <point3d>0.540000 0.120000 0.000000</point3d>
    <point3d>0.000000 0.120000 0.000000</point3d>
  </visualmodel>
  <task>
    <name>switchon</name>
    <physint>
      <preshape>1</preshape>
      <graspframe>0.353996 0.307992 0.019425
                 -0.513390 0.066517 0.108215</graspframe>
      <pushdir>0.000000 0.000000 0.000000 0.000000
               0.000000 0.000000</pushdir>
      <fd>1.753352 -2.443886 -5.067749 2.417732
          3.510157 -0.249084</fd>
      <itime>6495</itime>
      <dograsp>0</dograsp>
    </physint>
  </task>
</object>
```

III. Task reproduction

When the user asks the robot to perform a task on a object, the corresponding XML file is loaded, and the following steps are performed:

1) The robot first loads the visual reference model and asks for a pose initialization. Currently the user has to click on the visual reference, although we plan to automate this process in the future. After the visual reference is initialized, its edges are tracked and the full pose is continuously estimated.
2) After the pose is estimated, the robot goes to the contact point (relative to the visual reference) and applies the required force.

Figure 1 shows four different reproductions of the two tasks under different poses of the dishwasher.

IV. Conclusion

This paper has shown our work in progress towards the automatic specification of physical interaction tasks from a single demonstration assisted by a human. The main application concerns showing our future robot companions how to interact with household appliances in a manner that can be easily transferred to other robots. We adopt an approach similar to what is done for showing this kind of interaction to other humans: first we show how to perform the task, then we let the other do the task without our intervention and either approve it or show the task again. The proposed approach relies on our previously published physical interaction framework, that allows to specify grasping and interaction tasks in a robot-independent manner. Therefore, tasks shown to one robot can be reused by other robots, without the need of showing the task again. This work is still preliminary and needs further development, specially concerning reducing the instructor intervention in aspects like reporting the visual reference dimensions or making the pose initialization. In addition, further experiments have to be performed with different objects and considering tasks that require several sequential actions.

References

[1] S. Calinon. *Robot Programming by Demonstration: A Probabilistic Approach*. EPFL/CRC Press, 2009.

[2] A. I. Comport, E. Marchand, and F. Chaumette. Robust model-based tracking for robot vision. In *In IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS04*, pages 692–697, 2004.

[3] S. Ekvall and D. Kragic. Grasp recognition for programming by demonstration. In *IEEE International Conference on Robotics and Automation*, pages 748–753, 2005.

[4] M. Prats, P. J. Sanz, and A. del Pobil. A framework for compliant physical interaction - the grasp meets the task. *Autonomous Robots*, 28(1):89–111, 2010.

[5] J. Sorribes, M. Prats, and A. Morales. Visual robot hand tracking based on articulated 3d models for grasping. In *IEEE International Conference on Robotics and Automation*, Anchorage, Alaska, May 2010.