Benchmark for Human-to-Robot Hand-Overs of Glasses with Different Fillings

Konstantinos Chatzilygeroudis[*2†], Ricardo Sanchez-Matilla[*1], Apostolos Modas[3], Kaspar Althoefer[4], Véronique Perdereau[5], Pascal Frossard[3], Aude Billard[2], and Andrea Cavallaro[1]

[1] Centre of Intelligent Sensing - Queen Mary University of London - QMUL, UK
[2] LASA Laboratory, Swiss Federal School of Technology in Lausanne - EPFL, Switzerland
[3] LTS4 Laboratory, Swiss Federal School of Technology in Lausanne - EPFL, Switzerland
[4] Centre for Advanced Robotics - Queen Mary University of London - QMUL, UK
[5] Institut des Systèmes Intelligents et de Robotique - UPMC/Sorbonne Université, France
[*] Equal contribution
[†]Corresponding author: konstantinos.chatzilygeroudis@epfl.ch

Motivation

In most of our everyday activities, such as handing over some previously unseen objects (e.g., a cup) to others, we use our perception to get an initial estimation of the properties of the object and then use tactile and force feedback from the interactions in order to improve our initial estimation and perform the activity.

Having robots and humans to share the same room (and even workspace) with safety opens up a diverse and big set of applications. Examples include robots performing household chores, robots helping people working in elderly services, or robots helping employees in factories to lift heavy loads. In particular, handing-over objects from and to a human is one task that can find applications in a wide variety of fields/sectors. However, even though the execution accuracy of robotic systems exceeds human capabilities, the perception, dexterity and manipulation abilities of robots still fall way behind those of humans.

The objective of the benchmark is to facilitate the development of algorithms for scenarios where a robot observes the manipulation of an object by a human and infers how to robustly and safely grasp the object during a hand-over from the human. In more detail, the proposed benchmark aims at providing a means for evaluating algorithms that perceive the environment and control robotic systems that have to perform hand-overs of unknown and previously unseen objects that are filled in different percentages of their capacity. The properties of the objects (shape, size, rigidity), the type and amount of content are unknown and have to be inferred through multi-modal perception. The algorithms should only have access to the general semantic category that the objects belong to (e.g., a cup or a food box).

This benchmark is inspired by the CORSMAL[1] project that is exploring the fusion of multiple sensing modalities (touch, sound, and vision) to accurately and robustly estimate the physical properties of objects in noisy and potentially ambiguous environments.

Task Description

The high-level overview of the task can be described with the following steps:

- an empty object is placed on the table and filled with some content by the human,
- the human grasps the object and moves it closer to the robot,
- the hand-over from human to robot is performed.

[1] http://corsmal.eecs.qmul.ac.uk

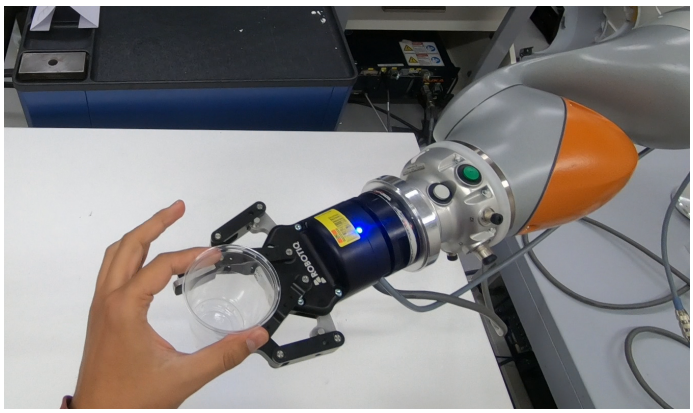

Fig. 1. An example of hand-over of a glass from a human to a robot.

An example of the setup is given in Fig. 1 where a human is performing the hand-over of a plastic glass to a KUKA manipulator.

The core challenge of the proposed task is the intra-class variability of the objects to be handed over. The shape and material variations (e.g., glasses of different sizes, shapes and materials), the content and the filling percentage (e.g., empty, half and full glasses), make the problem particularly challenging. These changes are compounded by the variability in human grasping that will lead to different degrees of occlusions of the object to be handed-over.

Objects The benchmark includes a set of physical objects that can be easily purchased and a multi-modal data collection on those objects. In particular, the objects to be considered are glasses of different shapes and materials (different rigidity, color, transparency). A few examples of the glasses that will be part of the benchmark are shown in Fig. 2. We will also use some of the glasses from the **YCB object database**. Finally, we also propose to add the new glasses into the YCB object database (i.e., RGB-D point clouds etc.).

Sensor Setup We propose to use three cameras. Two static third-person view cameras and one first-person camera worn by the human (see a sample frame in Fig. 1). We will provide detailed descriptions of how to replicate the camera setup. An illustration of a tentative setup is shown in Fig. 3.

To encourage participation from the wider research community, we also allow the usage of RGB-D cameras (in place of the

Fig. 2. The set of glasses that will be part of the benchmark.

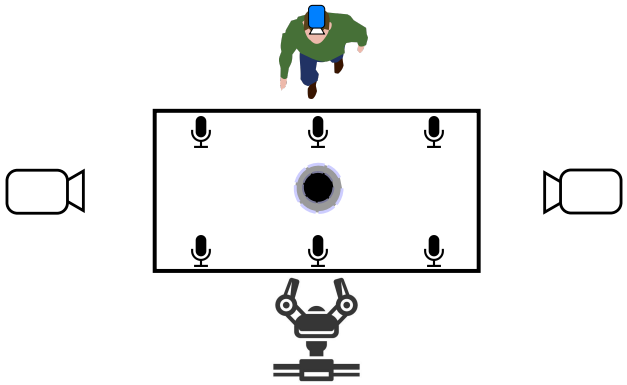

Fig. 3. Illustration of a sensor setup.

RGB ones) as well as microphone arrays placed on the table. Furthermore, the robot gripper is allowed (and even encouraged) to be equipped with touch or pressure sensors in order to facilitate the inference of the object's material properties.

Finally, our goal is to provide, along with the benchmark, a detailed dataset with ground-truth annotations of the physical characteristics of the objects, descriptions of the contents (including content type and weight) as well as pre-trained computer vision (and possibly multi-sensor) models and/or features that will help robotic groups (without any computer vision expertise) to participate in the benchmark.

Fillings We will use different types of content and five filling percentages: 0% (empty), 25%, 50%, 75%, and 95%.

Protocol Let us consider the setting with an initial state where an empty object (e.g., a glass) is on the table and the content that will be used to fill the object is rice. The human pours the rice into the glass while the action is captured by the two fixed cameras, the body camera, and the microphone array. Then the human picks up the object and moves it closer to the manipulator and the hand-over is performed. The hand-over is successful if the object stays in the gripper of the robot for more than t_{grasp} seconds and the human motion was not disrupted. The robotic system has access to the raw sensor inputs the whole period of the experiment.

Practical details The experiments should be performed on a table with the white table cloth from the YCB database. For each percentage of filling, the initial pose the object is defined randomly for ten different scenarios. Printable layouts will be provided for all these configurations. The robot does not know the models of the objects nor whether they are filled or not (or how much). The system only has access to the semantic category that the object belongs to; in particular, the system will know that the objects belong in the general category of cups.

Performance evaluation

The goal of the proposed benchmark is to evaluate the ability of a system to perform the same task using different objects and conditions. The score of the benchmark, S, is defined as follows (the larger, the better):

$$S = \frac{\sum_{k=0}^{K} \sum_{l=0}^{L} W_{kl} \sum_{i=0}^{N} t_i^{kl} \frac{M_i^{kl}}{M^{kl}}}{NKL}$$

where K is the number of objects, L is the number of filling percentages, N is the number of repetitions, W_{kl} is the weight for the k-th object with the l-th filling (depicting the difficulty level of each combination[2]), M^{kl} is the mass of the k-th object with the l-th filling, M_i^{kl} is the mass of the k-th object with the l-th filling after the end of the i-th repetition, and t_i^{kl} is the time (in seconds) that k-th object with the l-th filling stayed in the gripper of the robot in the i-th repetition. If there is no hand over then t_i^{kl} is set to zero and thus the score of this repetition is zero.

This scoring scheme has the following interesting properties:

- Puts pressure to not spill the content of the objects,
- Enforces the hand over behavior,
- Measures robustness by measuring the ability to generalize over objects, fillings for multiple instances of initial poses.

Conclusion

We believe that the proposed benchmark and the sharing of data and methods will encourage collaboration across research disciplines that can contribute to the robotics community. We also hope that the benchmark will underpin the development of accurate and safe algorithms that will emerge from solving the proposed task that is fundamental for supporting everyday human-robot interactions.

[2] We will refine the weights for the final submission.