HUMAN-TO-ROBOT HANDOVER PROTOCOL

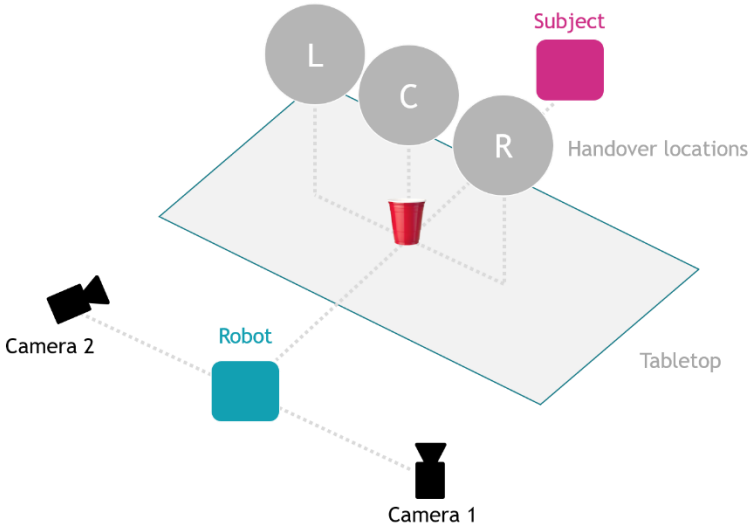
<table>
<tr><td>Reference No / Version</td><td>“RAL-SI-2020-P19-0835 -V1.0” for the latest versions of the protocol, please refer to http://corsmal.eecs.qmul.ac.uk/benchmark.html</td></tr>
<tr><td>Authors</td><td>Ricardo Sanchez-Matilla, Konstantinos Chatzilygeroudis, Apostolos Modas, Nuno Ferreira Duarte, Alessio Xompero, Pascal Frossard, Aude Billard, and Andrea Cavallaro</td></tr>
<tr><td>Institution</td><td>Centre for Intelligent Sensing, Queen Mary University of London, UK;
LASA&LTS, Swiss Federal School of Technology in Lausanne, Switzerland;
VisLab, Institute of Systems and Robotics, Lisbon, Portugal</td></tr>
<tr><td>Contact information</td><td>ricardo.sanchezmatilla@qmul.ac.uk, konstantinos.chatzilygeroudis@epfl.ch</td></tr>
<tr><td>Purpose</td><td>Assessing the ability of a robot to perform handovers from a human with unseen objects whose physical properties such as dimensions, shape, rigidity and fillings are unknown</td></tr>
<tr><td>Task Description</td><td>A subject grasps a (filled) object from a table and hands the object over to a robot arm that must infer the object properties on-the-fly. After the handover, the robot arm delivers the object to a pre-defined location on the table. Any individual that is asked to perform the handover task is referred to as subject.</td></tr>
<tr><td>Setup Description</td><td><u>List of objects and their descriptions:</u>
The task is composed of 288 configurations.

<u>Four</u> subjects are required to complete the handovers

We use <u>four</u> types of cups of different shapes, sizes, and materials, and with different degrees of deformability and transparency. These cups are inexpensive and available worldwide.
<table>
<tr><th></th><th>Deformability</th><th>Transparency</th><th>Purchase link</th></tr>
<tr><td>Cup 1</td><td>High</td><td>Medium</td><td>http://bit.ly/2N6n3tm</td></tr>
<tr><td>Cup 2</td><td>Average</td><td>Low</td><td>https://amzn.to/2QrsXH5</td></tr>
<tr><td>Cup 3</td><td>Average</td><td>High</td><td>https://amzn.to/2JwRk3l</td></tr>
<tr><td>Cup 4*</td><td>No</td><td>High</td><td>https://amzn.to/33zw4AY</td></tr>
</table>
* wine cup from YCB

The subjects are instructed to grasp the cup in <u>three</u> different ways
<table>
<tr><th></th><th>Grasp 1</th><th>Grasp 2</th><th>Grasp 3</th></tr>
<tr><td>Cup1</td><td>  </td><td>  </td><td>  </td></tr>
</table></td></tr>
</table>

The subjects are instructed to perform the handover at <u>three</u> predefined handover areas.



Cups should be filled with different amounts of filling to vary the mass and deformability of the cup. We use <u>two</u> amounts of fillings. Empty or rice, which is easy to purchase and - unlike liquids - harmless for the hardware.
We fill the cups as

Properties	**Unit**	**Cup 1**	**Cup 2**	**Cup 3**	**Cup 4**
Volume	ml	179	497	605	354
Filling amount	ml	125	400	450	300

	Initial and target poses of the objects: For each execution of the task (configuration), the cup is initially placed at the center of the table. Subjects will grasp and carry the cup to different approximate handover areas above the table, which should be constrained by the reachability of the arm (i.e., 40% - 50%), such that it is comfortable for the subjects, and for the handover to be performed as naturally as possible. After the handover, the robot places the cup back at the initial location at the center of the table. Participants are free to choose the best way to measure the final error (e.g. using printable stencils as for YCB). Note that participant is any group or individual that evaluate their method(s) on the benchmark and submit some results.
	Description of the manipulation environment: Conduct the experiment on a table, which can be covered with a white tablecloth (e.g. white tablecloth used in YCB) Purchase link: https://amzn.to/2X5qnaN (LinenTablecloth 90 x 132-Inch Rectangular Polyester Tablecloth with Rounded Corners, White)
Robot/Hardware/Software/Subject Description	**Targeted robots/hardware/software:** Robotic arm with at least 6 degrees of freedom (e.g. KUKA, UR5) equipped with a gripper (e.g. a 2-finger gripper or a complete robotic hand). Two cameras. In addition, a third camera to record the overall scene shall be used for documenting the experiments
	Initial state of the robot/hardware/subject with respect to the setup: Any initial robot pose can be chosen with respect to the environment setup; however, we expect the subject to stand on the opposite side of the table with respect to the robot. Subjects are previously instructed with all the handover configurations that they must perform (e.g. cup type, filling amount, grasping type, and handover areas)
	Prior information provided to the robot: Only the high-level semantic category of the object is provided. Databases of the specific objects or their 3D models are not provided prior to the task, as well as any other partial knowledge (e.g. object dimensions, shape or weight). Information of each object should be estimated on-the-fly via perception (e.g. vision or robotic sensors)
Procedure	For each configuration: 1. Prepare the cup with its filling 2. Place the cup at the center of the table 3. The subject grasps the cup from its location 4. The subject carries the cup with the intention of handing it over to the robot 5. The robot should track and predict the pose of the object to move the arm towards the handover area 6. The subject hands the cup over to the robot 7. The robot closes the end effector and grasps the object 8. The robot delivers the cup to the initial location.

Execution Constraints	• Participants are not allowed to be subjects as well: the subjects should be external to the design and development of the system. • Each subject is instructed to use only one hand and always the same. • No prior knowledge of the objects is available apart of its category (i.e. cup). • The location to deliver the cup must be inferred using perception from the initial location and not hard-coded. • Learning across executions of configurations is not allowed: participants must not update/fine-tune the vision/robotic algorithms using the data/measurements captured during the execution of the configurations of the benchmark by the subjects (test time).