Learning to Grasp Objects in Virtual Environments through Imitation

Alexandre Filipe, Alexandre Bernardino and Plinio Moreno

Motivation

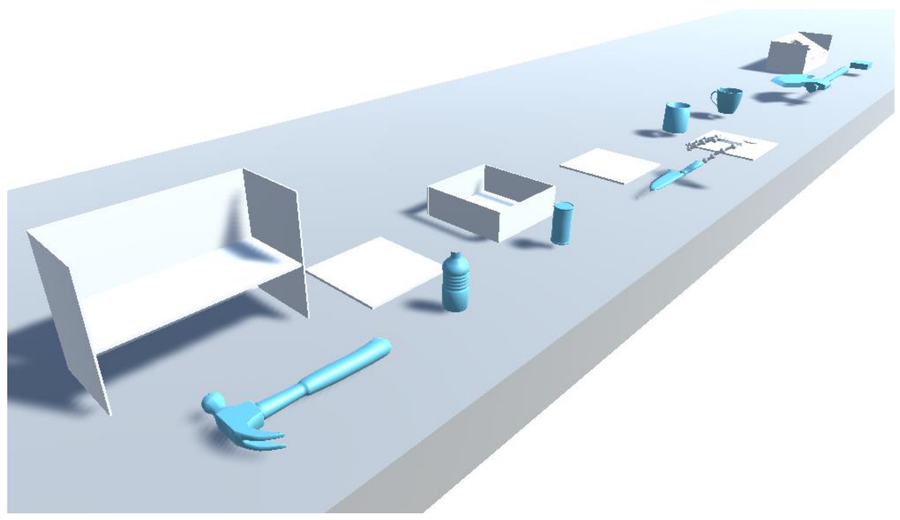
Even with today's standards of what robots can achieve, robotic manipulation still stands as a huge challenge, leading to the common use of grippers/claws to avoid having to deal with the full complexity of a human hand.

In this work we will record demonstrations of simple manipulation tasks in a virtual environment, capturing the human hand with a glove with sensors, train LSTMs with these demonstrations and finally test the quality of the reproduction of the tasks performed by the trained LSTMs. Our method will also segment the demonstrations/reproductions in 2 phases to ensure the task will only continue if the object is well grasped.

Finally, to allow anyone intending to try and test their own imitation algorithms, we will provide our virtual environment and complete dataset of demonstrations freely.

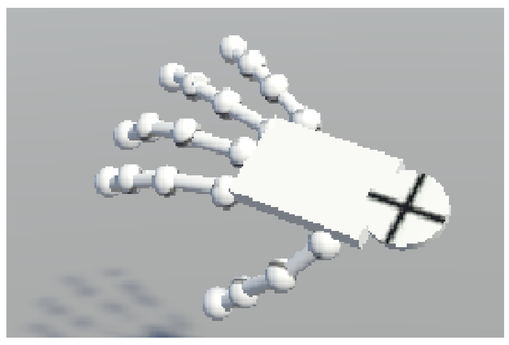
Setup

The tasks are demonstrated and reproduced on a virtual environment hosted in Unity (v2019.1). The virtual environment consists of a table with several objects to interact with.



Virtual Environment

To capture the human hand the VMG35-Haptic glove is used. This glove contains 2 gyroscopes to indicate the angular position of the hand and wrist and sensors placed on the joints of the fingers to measure how much each is bent. The hand will be represented in the virtual environment by a skeletal representation of the human hand.



VMG35-Haptic (left) and virtual representation of the hand (right)

The demonstrations/reproductions are represented by a sequence of iterations (captured every fifth of a second). Each iteration contains the information of how much each joint is bent and the relative spatial and angular position of the hand to the desired object.

The demonstrations/reproductions are also segmented into 2 phases, before and after the selected object is grabbed. This was added to avoid the hand continuing the task without the object being well grasped (usually dropping it in the process). This way the reproduction will continue to try to grasp correctly the object, advancing only after doing so.

Process

The reproductions start by choosing a starting position for the hand. The object to be grasped will also spawn at a random place (the same happens during the demonstrations). The first LSTM, using the starting position, will try to predict the following iteration. This will happen recurrently, creating a string of iterations that represent an action. Meanwhile, in the virtual environment, the virtual hand is following the values outputted by the LSTM.

The moment the virtual environment detects the object is grabbed it sends a flag to the LSTM program. The LSTM program after receiving the flag switches to the second LSTM, continuing this one to predict the following iterations that will, hopefully, recreate the trained task.

Experimental results and conclusions

To test for each task 20 random starting positions were chosen. If from a position the LSTMs could accomplish the task it was considered a success. The tasks were:

- Grab a hammer and place it on a shelf;
- Grab a bottle and place it on a base;
- Grab a can, rotate it and place it in a box;
- Grab a mug and bring it to a base.

The success rate was:

Bottle	85%
Can	80%
Mug	75%
Hammer	75%

In general the reproductions made using this method were a success.

To anyone that wants to test their own manipulation training methods our virtual environment and our complete dataset of demonstrations can be found in github.com/alexamor/thesis.

This work was supported by FCT with the LARSyS - FCT Project UIDB/50009/2020.