Bin Packing Demonstrations Dataset

André Santos andrejfsantos@tecnico.ulisboa.pt Atabak Dehban adehban@isr.tecnico.ulisboa.pt José Santos-Victor jasv@isr.tecnico.ulisboa.pt

Abstract—This paper introduces a new multifaceted dataset of humans packing supermarket-like objects into a box. It is the first to collect demonstrations of this task with such diversity: in total, 245 boxes were packed by 27 different participants, yielding 4342 everyday object manipulations, all in virtual reality.

I. INTRODUCTION

Recent developments in virtual reality (VR) technology have encouraged its adoption in many fields of scientific research, including for data generation. These include improvements in automatic data annotation capabilities and increased realism, which were key factors when choosing to collect our dataset in a virtual environment. The most similar dataset available was collected from real demonstrations by [1] and required a complex setup for tracking the gripper and each grasp. Our approach only requires a VR-capable computer and headset.

The dataset enables learning multiple aspects of this task, inferring grasp poses appropriate for a parallel gripper, understanding how humans manipulate objects, how objects influence each other when placing and stacking them, and more.

II. METHODOLOGY AND PRELIMINARY RESULTS

The virtual environment shown in Fig. 1 was developed in Unity and contains a static table and box, as well as 24 different objects which behave as rigid bodies. 20 of these are from the YCB objects set and the others from other public sources. The participants interacted with the virtual world via the controller, which controls a virtual gripper. They packed 245 boxes, each with a different random subset of the aforementioned objects. We recorded more than 4000 6-DoF grasp poses, averaging 181 grasp poses per object. The data also includes the same number of object trajectories (sampled at 20 Hz) and placements inside the box (both the position and orientation). Finally, it includes the sequences in which the objects were packed and a snapshot of their initial layout.

Fig. 1 *right* shows a plot of the box filling duration as a function of the number of objects that were packed, thus obtaining a measure of the difficulty of the task. The pink line was obtained with a linear regression and indicates that on average, the duration increases 4 seconds per object.

Furthermore, histograms of the placement positions of each object inside the box showed that larger objects tend to be placed further away from the participant. This is likely because

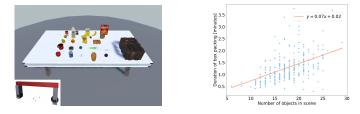


Fig. 1. (left) Virtual environment and gripper. (right) Task duration against number of objects.

otherwise they would obstruct subsequent placements. Fig. 2 shows an analysis of the placement orientations, revealing that *e.g.*, symmetric objects, such as the tall bleach bottle from YCB, are almost always placed aligned with the edges of the box, likely to optimize the limited box space. Finally, by extracting frequent ordered pairs of objects from the packing sequences we built a directed graph, shown in Fig. 2 *right*, that reflects the order in which the objects are packed, thus encoding intrinsic object properties such as fragility and size.

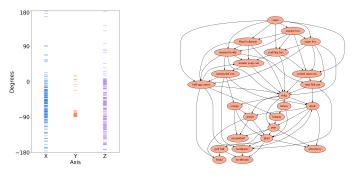


Fig. 2. (left) Histogram of the placing orientations of a bleach cleanser (Y axis points up). (right) Frequent ordered pairs of objects.

III. CONCLUSIONS

With this new dataset we hope to facilitate further research into automating the task of filling a box with everyday objects and understanding how humans manipulate objects in complex sequential scenarios. We plan on building a model based on Fig. 2 to predict the order in which objects should be packed, and another model to predict optimal object placement poses. We are developing a framework to handle the dataset, which will soon be made available along with the dataset. Interested researchers are welcome to request the data from the authors.

REFERENCES

 S. Song, A. Zeng, J. Lee and T. Funkhouser, "Grasping in the wild: learning 6DoF closed-Loop grasping from low-cost demonstrations," IEEE Robot. Autom. Lett., vol. 5, no. 3, pp. 4978-4985, July 2020.

This work was supported by the Fundação para a Ciência e a Tecnologia (FCT) through the ISR/LARSyS Grant UID/EEA/50009/2020, the RBCog-Lab Research Infrastructure; the H2020 FET-Open project *Reconstructing the Past: Artificial Intelligence and Robotics Meet Cultural Heritage* (RePAIR) under EU grant agreement 964854; and by the Lisbon Ellis Unit (LUMLIS).