Construction of a 3D Object Recognition and Manipulation Database from Grasp Demonstrations

David Kent and Sonia Chernova
{davidkent, soniac}@wpi.edu
Worcester Polytechnic Institute, Worcester, MA, USA

Abstract— **Object recognition and manipulation are critical for enabling robots to operate in household environments. Object model databases can improve manipulation by accounting for non-visual object informtion often missed by shape-based grasp planners, but existing methods for database construction are time and resource intensive. We present an easy-to-use system for constructing object models for 3D object recognition and manipulation made possible by advances in web robotics. The database consists of point clouds generated using a novel iterative point cloud registration algorithm, and demonstrated grasp poses learned through an online training algorithm. We validate the system with data collected from both a crowdsourcing user study and expert demonstration. We show that the demonstration approach outperforms purely vision-based grasp planning approaches for a wide variety of object classes.**

Index Terms— **Object Manipulation, Grasp Demonstration, Crowdsourcing**

I. INTRODUCTION

The abilities to recognize and manipulate a wide variety of objects are critical for robots operating autonomously in household environments. As such, a robot must automatically determine useful properties of encountered objects, or make use of a database of known objects. Automatic grasp detection can lead to difficulties in determining usability characteristics for arbitrary objects, but a database of known objects enables this type of information to be encoded into the object models.

A major disadvantage of object recognition and grasping databases is that they require a significant amount of work to construct. Adding new objects to a database first requires obtaining a detailed model of each new object. The current state of the art is to use a turntable with sensors mounted around it in a specific setup [1]. This is time consuming, however, and cannot be used for learning in a real-world environment where such a detailed setup is impractical. Once the object model is constructed, grasp points must be calculated. This is often accomplished through geometric analysis [2], or through exhaustive testing in simulation [3].

We turn to grasp demonstration and crowdsourcing to address these challenges. Grasp demonstration by human users provides a natural way to convey information about important hard-to-observe physical object characteristics, such as weight distribution or fragility of materials, which are not accounted for by the geometric analysis planners. Furthermore, the developing field of web robotics, such as the Robot Management System [4], allows us to collect these demonstrated grasps on a large scale, as well as training data for object recognition.

In this work, we first present a web-based system for the collection of point cloud data and demonstrated grasps for a desired set of objects. We also present a graph-based algorithm for constructing 3D object models from the sets of point clouds for each object gathered by the web-based data collection system, and verify that the generated models are suitable for object recognition. Lastly, we present a system for incorporating demonstrated grasp data into the 3D object model database. We compare the effectiveness of the crowdsourced grasps to expert demonstrated grasps, and we show the advantages that this grasping system has over a grasp planner based solely on object geometry.

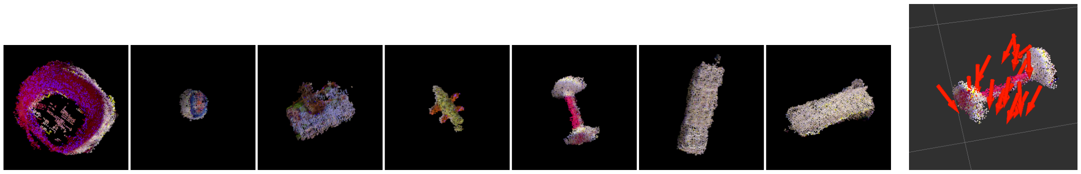
II. CROWDSOURCING EXPERIMENT

Leveraging the Robot Management System, we conducted a crowdsourced user study to gather object recognition and grasping data for the database. The study setup consisted of a PR2 facing a table containing ten household objects arranged in random positions and orientations. Participants remotely connected to the robot from their home computers using only a web browser. Using interactive markers to change the position and orientation of the physical robot's end effectors, participants picked up as many objects as possible within an alloted twenty minutes. Upon a successful pickup, the system stored the segmented point cloud of the grasped object, the pose of the PR2's gripper, and an object label.

III. MODEL CONSTRUCTION

Given the object and grasping data gathered from the crowdsourcing user study, the next step of our process was to construct the database. Combining point clouds collected in an unstructured environment into complete 3D models is a challenging problem since the object angles and relationships between multiple viewpoints are unknown. This problem can be solved with iterative point cloud registration, but the method's high susceptibility to error propagation often results in inaccurate models.

To address this, we propose a new graph-based point cloud registration algorithm which determines an effective order for successive pairwise merging of point clouds. We evaluate whether merges are successful or failed using a decision tree trained from a set of metrics for evaluating potential merges



(a) Examples of object models constructed from the user study data. Left to right: bowl, hacky sack ball, dinosaur book, dragon toy, plastic dog bone, phone (front view), phone (back view)

(b) Object model with grasps

Fig. 1. Object model examples

based on overlap, distance error, and similarity of color. A graph constructed of nodes representing each point cloud and edges representing successful merges is then collapsed by merging two nodes at a time until no edges remain. This results in a set of disparate 3D object models for each object. Figure 1(a) shows examples of models constructed from the crowdsourcing data using this approach.

The registration process also transforms the demonstrated grasp poses to the final 3D object models (shown in Figure 1(b)), but due to sensor and executional errors, not all of the grasps are useful. We developed a system which determines the effectiveness of each grasp, and eliminates unsuccessful grasps. The grasp training process first removes any clear outliers offline by filtering any grasp poses that are significantly outside of the object's space. The remaining grasps are then tested with an online epsilon-greedy exploration algorithm, which learns success rates for each grasp through trial-and-error. The online algorithm initially prioritizes the testing of unexplored grasps, and later refines the experimentally determined grasp rates of previously explored grasps. The algorithm terminates when a sufficient number of successful grasps are found, and any grasps with a zero success rate are eliminated from the database.

IV. Summarized Results

We first evaluated the usefulness of the 3D object models for object recognition. Using the database constructed from the crowdsourcing user study data, the system correctly classified objects at a rate of 88.7%. To test the scalability of the system, we also collected recognition data for a larger set of thirty objects. The system correctly classified this supplemental object set at a rate of 83.93%.

To evaluate the successfulness of the grasping system, we compared the results of a series of grasping trials using the grasps produced by the database trained from the user study data, grasps produced by a database trained using expert-demonstrated grasps, and grasps produced by the PR2's geometric partial-view grasp planner. The results are shown in Figure 2. In general, the grasps learned from crowdsourced data performed comparably with the grasps learned from expert demonstration and from geometric analysis, and in some cases outperformed the geometric planner. We also found that the expert demonstrations produced high-quality grasps at a much greater rate than that of the crowdsourced demonstrations, but due to the grasp learning process the poor-quality grasps are filtered out and the final grasping results are comparable.

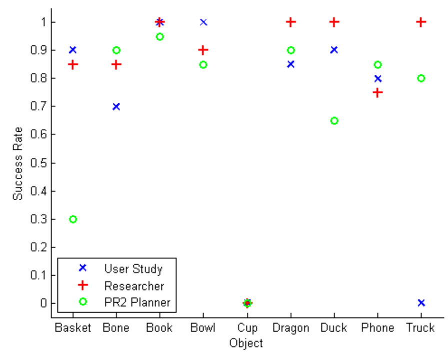


Fig. 2. A comparison of grasp rates for the user study object set

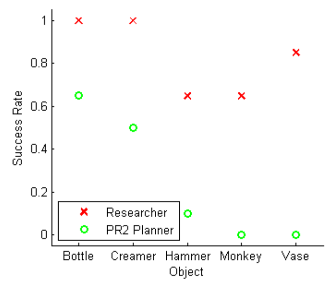


Fig. 3. A comparison of grasp rates for the supplemental object set

To better show the advantages of the demonstrated grasps, we also performed the same evaluation on a supplemental set of special-case objects, which contained unusual materials, extreme weight distributions, and removable parts. The results of these test are shown in Figure 3, where the grasps from the database significantly outperform the grasps based solely on object geometry.

References

[1] A. Kasper, Z. Xue, and R. Dillmann, "The KIT object models database: An object model database for object recognition, localization and manipulation in service robotics," *The International Journal of Robotics Research*, vol. 31, no. 8, pp. 927–934, 2012.

[2] K. Hsiao, S. Chitta, M. Ciocarlie, and E. Jones, "Contact-reactive grasping of objects with partial shape information," in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, Oct 2010, pp. 1228–1235.

[3] C. Goldfeder, M. Ciocarlie, H. Dang, and P. K. Allen, "The Columbia grasp database," in *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*. IEEE, 2009, pp. 1710–1716.

[4] R. Toris, D. Kent, and S. Chernova, "The robot management system: A framework for conducting human-robot interaction studies through crowdsourcing," *Journal of Human-Robot Interaction*, 2014.