

# Benchmarking Protocol for Grasp Planning Algorithms

Yasemin Bekiroglu<sup>\*</sup>, Naresh Marturi<sup>1\*</sup>, Máximo A. Roa<sup>2\*</sup>, Maxime Adjigble<sup>1</sup>  
Tommaso Pardi<sup>1</sup>, Cindy Grimm<sup>3</sup>, Ravi Balasubramanian<sup>3</sup>, Kaiyu Hang<sup>4</sup>, Rustam Stolkin<sup>1</sup>

**Abstract**—Numerous grasp planning algorithms have been proposed since the 1980s. The grasping literature has expanded rapidly in recent years, building on greatly improved vision systems and computing power. Methods have been proposed to plan stable grasps on: known objects (exact 3D model is available), familiar objects (e.g. exploiting a-priori known grasps for different objects of the same category), or novel object shapes observed during task execution. Few of these methods have ever been compared in a systematic way, and objective performance evaluation of such complex systems remains problematic. Difficulties and confounding factors include: different assumptions and amounts of a-priori knowledge in different algorithms; different robots, hands, vision systems and setups in different labs; different choices or application needs for grasped objects. Also, grasp planning can use different grasp quality metrics (including empirical or theoretical stability measures), or other criteria, e.g. computational speed, or combination of grasps with reachability considerations. While acknowledging and discussing the outstanding difficulties surrounding this complex topic, we propose a methodology for reproducible experiments to compare the performance of a variety of grasp planning algorithms. Our protocol attempts to improve the objectivity with which different grasp planners are compared by minimising the influence of key components in the grasping pipeline, e.g. vision and pose estimation. The protocol is demonstrated by evaluating two different grasp planners: a state-of-the-art model-free planner, and a popular open-source model-based planner. We show results from real-robot experiments with a 7-DoF arm and 2-finger hand, and simulation-based evaluations.

## I. INTRODUCTION

Grasp planning has been studied since the 1980s [1], with a recent proliferation of novel approaches. Different methods assume different a-priori knowledge, e.g. model-based [2] versus model-free [3], and adopt widely differing approaches, e.g. analytic [4], [5], data-driven [6], [7], or based on human demonstration [8]. Other work explores mechanically adaptive hands that simplify the grasping process thanks to their inherent mechanical adaptability [9], [10], or combines different sensing modalities for performing grasping [11] and in-hand manipulation [12]. Recent reviews [13]–[16] categorise and discuss these algorithms in terms of their

Manuscript received: August 15, 2019; Accepted: November 4, 2019.

This paper was recommended for publication by Editor Han Ding upon evaluation of the Associate Editor and Reviewers' comments.

This work was supported by CHIST-ERA project EP/S032428/1 Pe-GRoGAM and EP/R02572X/1 (National Centre for Nuclear Robotics).

<sup>1</sup>Extreme Robotics Lab, University of Birmingham, UK. <sup>2</sup>German Aerospace Center - DLR, Germany. <sup>3</sup>Robotics and Human Control Systems Lab, Oregon State University, USA. <sup>4</sup>GRAB Lab, Yale University, USA. [yaseminb@kth.se](mailto:yaseminb@kth.se), [n.marturi@bham.ac.uk](mailto:n.marturi@bham.ac.uk)

\* These authors are identified as the joint first authors of the work.

Digital Object Identifier (DOI): see top of this page.

differences, assumptions and limitations. Competitions such as the Amazon Picking Challenge (APC) [17] and Robotic Grasping and Manipulation Competition at IROS [18] have proposed different tasks to compare the performance of whole robotic systems. While stimulating significant progress, such contests can also engender over-fitting of engineering solutions to the proposed tasks.

Comparing different algorithms on common benchmark datasets is now widespread in the machine learning and vision communities. In contrast, objective benchmarking of grasping appears significantly more challenging, since numerous confounding factors are introduced by the diversity and complexity of mechanical, sensorial and algorithmic innovations combined in any particular solution. One approach is for the community to agree on standardized sets of objects, e.g. the YCB data set [19]. However, even with common objects, two key problems remain. One is how to decouple performance of diverse components in the grasping pipeline, i.e., how do we evaluate the influence of a grasp planning algorithm independently of the vision system, arm and hand? The second issue is a Pandora's box of experimentally confounding factors (arm kinematics, reachability, perception, hand control, etc). In addition, repeatable experiments in some cases, e.g. with deformable objects, or random heaps of objects are difficult.

Different 2D and 3D imaging devices offer greatly varying precision and robustness. Even with a common vision system, we may be confounded by lighting variation, shadows, reflections, surroundings, and other factors. Such factors can be somewhat overcome by benchmarking at the system level, for instance in shelf-picking [20] or bin picking [21] scenarios. Other approaches are focused on evaluating the end-effector performance [22], [23]. There are also approaches focusing on simulation environments for the comparative evaluation of algorithms for grasping and manipulation, [24] offering to use the same visual input and a fixed hardware setup [25]. However, there is still no principled methodology that clearly defines each step in the grasping pipeline, from input sensory data to task execution given any robot setup. Hence the comparative performance of different grasp planning algorithms remains largely uncovered, independently of additional factors such as robot kinematics, gripper design, object perception and pose estimation.

This paper presents a method for comparing the performance of different grasp planning algorithms. We define a protocol for executing grasps under repeatable conditions in terms of objects and their placements, and using a proposed success criterion for the robustness of the grasp. The method

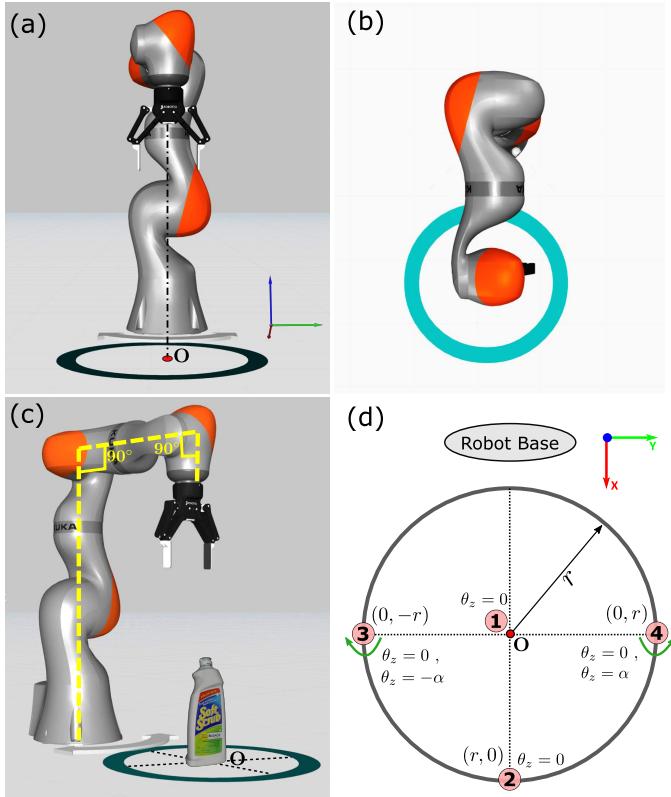


Fig. 1. Illustration of the selected circular workspace for the KUKA iiwa. (a)–(c) Multiple views of the robot and the workspace. The radius of the considered area is selected as 25 cm, and its origin  $O$  is selected by projecting the robot’s tool centre point onto the ground plane when the robot is in a 90–90 configuration as shown in (c). The numbers inside circles shown in (d) represent, respectively, the considered location and order of the experimental object poses. The location coordinates are shown with respect to  $O$ .

can be applied to model-based and model-free approaches. We measure performance on scenes with single objects, and also cluttered heaps of objects. We show how our protocol can be used in experiments with a real robot and vision system, and also in a simulation environment. We evaluate the performance of state-of-the-art model-based [2] and model-free [3] grasp planners using seven different objects from the YCB dataset.

The remainder of the paper is organized as follows. Section II introduces the benchmark protocol, including environment setup, grasp execution procedure, and success criteria. Section III presents example evaluations of two grasp planners, along with some details of the planners and the robotic systems used. Section IV provides concluding remarks and suggestions for future work.

## II. BENCHMARK PROTOCOL

The protocol is defined in terms of the workspace of the robot, and specifies a horizontal surface for placing experiment objects, types and placement of objects, grasp execution procedure, and success evaluation. We describe each of these factors in detail below.

### A. Environment Description

To have generality with any robot setup and facilitate simple workspace construction, we propose that objects should



Fig. 2. Selected experiment objects with different complexities. (A) Simple objects: racquetball (YCB ID: 53), yellow cup 19g (ID: 56); (B) moderately complex objects: flat screwdriver (ID: 43), scrub cleanser bottle (ID: 20); (C) complex objects: spring clamp (ID: 46), assembled toy airplane (ID: 67); and (D) deformable object: chain (ID: 61).

be placed on a horizontal surface, on an area with good reachability. Based on this, the user defines a circular region within the workspace of the arm with radius  $r$  cm,  $r \in \mathbb{Z}^+$ , by placing the tool center point of the arm in a reachable comfortable pose within the robot’s workspace so that the object can be systematically moved. We project the centre of the robot flange/tool onto the ground plane when the robot is in a 90–90 configuration as shown in Fig. 1(c), i.e., the tool is vertically facing down, and the projected point is selected as the centre of the circle. A pictorial description is provided in Fig. 1. Here,  $r$  can be modified based on task requirements, i.e., for grasping large and heavy industrial objects  $r$  can be chosen to fit the objects properly.

For perceiving the scene to build object models (e.g. point clouds, meshes etc.), we consider an *eye-in-hand* scenario, where an RGB-D camera is mounted at the end-effector of the robot. An alternative would be scene camera(s), positioned to view both the arm and the object. *Hand-eye* calibration can be performed to transform the perceived object information into the robot base or world coordinate system so as to maintain generality in grasp hypotheses generation and execution. If the grasp planner uses pre-built models, the process can be initialized with the known object model and pose. In this case, the benchmark makes an isolated evaluation of the grasp planner. However, when the user decides to include a perception pipeline in the loop, the evaluation of the grasp planning algorithm is much closer to the real implementation in the robotic system.

We focus on the YCB object set [19] as it has been widely adopted by the community. We propose using the following items: *Food items*: chips can, coffee can, cracker box, box of sugar, tomato soup can, mustard container, chocolate pudding box, gelatin box, potted meat can, apple, orange. *Kitchen items*: pitcher, scrub cleanser, glass cleaner, plastic wine glass, enamel-coated metal bowl, metal mug. *Tools*: power drill, wood block, screw driver, spring clamp. *Shape items*: mini soccer ball, softball, baseball, racquetball, cups, foam brick, washers (3 distinct sizes), chain. *Task items*: airplane toy. Other objects from the YCB dataset can also be used, using the official name as a reference. In our experiments we use a subset of these objects, namely seven of them with different shape complexities, as seen in Fig. 2.

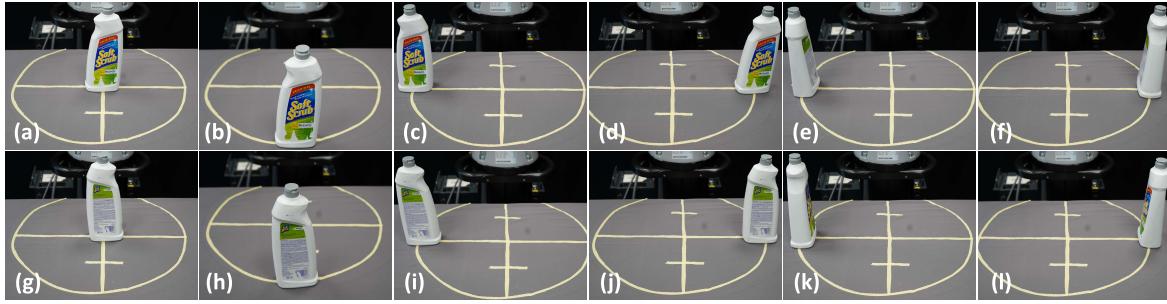


Fig. 3. Object locations and poses considered in this work. (a)–(f) Standard placement; (g)–(l) mirrored placement.

### B. Grasp Execution and Scoring

Each object is placed at a known initial position and orientation with respect to the robot, inside the circular workspace. The origin of the workspace is considered as the initial location of the object, which is then moved systematically within the defined range as follows. We select four locations in the workspace, including the origin, as specified in Fig. 1(d). We consider moving forward (location 2), right (location 3), and left (location 4) by  $r$  cm ( $r = 25$  cm for our experiments), and rotating by  $-\alpha$  and  $\alpha$  for locations 3 and 4 respectively ( $\alpha = 90^\circ$  for our experiments). It is worth noting that the backward location is not considered as it is close to the robot base and may cause reachability issues. For every location we also consider mirrored (rotated) cases where rotating the object around the normal of the horizontal surface on which it is lying does not look the same (non-symmetry case). All of the poses we consider are given in the list below, where  $0^\circ$  rotation corresponds to the initial pose (location 1 in Fig. 1(d)).

- P1**) Workspace origin,  $O: (X_O, Y_O)$  and  $\theta_z = 0^\circ$
- P2**) Forward (location 2):  $(X_O + r, Y_O)$  and  $\theta_z = 0^\circ$
- P3**) Right (location 3):  $(X_O, Y_O - r)$  and  $\theta_z = 0^\circ$
- P4**) Left (location 4):  $(X_O, Y_O + r)$  and  $\theta_z = 0^\circ$
- P5**) Right (location 3):  $(X_O, Y_O - r)$  and  $\theta_z = -\alpha$
- P6**) Left (location 4):  $(X_O, Y_O + r)$  and  $\theta_z = \alpha$

Each case should be repeated  $N$  times, leading to  $6 * N$  cases overall. For non-symmetric objects (where object mirroring is possible) this becomes  $12 * N$  cases, as all the above 6 tests are repeated for the mirrored object pose. Fig. 3 shows an exemplary object placement with the 12 possible poses. After manually placing the object as described above, and given the observed scene (e.g. RGB/RGB-D data) or the known object model and pose, we run the grasp planner and execute the best grasp (e.g. the one with maximum quality/likeness). This process is repeated  $N$  times. Besides, we define an initial object pose as one where the object is standing or lying in a stable manner on a flat surface, and evaluate the planner's performance based on this discrete number of cases. Examples of stable poses for selected test objects are shown in Fig. 4. To create cluttered scene evaluations, i.e., grasping with more than one object present, we randomly throw objects in a bag or a container tray and empty them onto the table top, leading to random object placements within the defined circle, similar to the procedure in the logistics track of [18]. We use the best



Fig. 4. Stable initial poses of the selected test objects.

grasp hypothesis for each trial until the scene is cleared. We repeat  $N$  times the whole clearing up.

Hand control has a direct effect on the successful execution of a planned grasp, and depends on the specific hand embodiment. Therefore, it is hard to define a common standard for the controller. We propose not to restrict controllers or manipulators and let the user define such details. We suggest using the following hand closure technique for a fully-controlled hand: start with an initial fully open configuration, then preshape the hand according to the planned grasp, and close the fingers with equal speeds while limiting the maximum torque of each actuator until reaching a static state where the object does not move, or until reaching a fully closed hand configuration (in the case of an unsuccessful grasp). The joint angles can be set to reach the fully closed configuration of the hand, which can only be reached if the object slips away. For the case of an underactuated hand the wrist is placed at the desired location, and the closing signal is sent to the robot until reaching a state where the object does not move inside the hand, or until a fully closed configuration is achieved.

The object is then lifted 20 cm above the table at a speed of 10 cm/s (the user should report if different values are used). A series of motions are executed as follows to verify grasp robustness, trying to emulate possible motions or perturbations

TABLE I  
RESULT FORMAT FOR GRASPING INDIVIDUAL OBJECTS

Object name (YCB ID) <sup>1</sup>	Stable pose #	Trial #	Object Location <sub>n n=1...4</sub> with Pose <sub>i</sub> : (X, Y), θ <sub>z</sub>				
			C1: Selected (total Feasible)	C2: Time (seconds)	C3: Lift Test	C4: Rotational Test	C5: Shaking Test
Object <sub>j</sub> {string} ({int})	Count {int}	Count {int}	ID (total count) {int} ({int})	{float}	✓   ✗	✓   ✗	✓   ✗

<sup>1</sup> Object names and IDs should match the YCB object list.

<sup>2</sup> ✓ is treated as 1 and ✗ is treated as 0.

TABLE II  
RESULT FORMAT FOR CLEARING A GROUP OF OBJECTS

Trial #	Object name (YCB ID)	Pickup order	C1: Selected (total Feasible)	C2: Time (seconds)	C3: Lift Test	C4: Rotational Test	C5: Shaking Test
Count {int}	Object <sub>j</sub> {string} ({int})	Count {int}	ID (total count) {int} ({int})	{float}	✓   ✗ {binary}	✓   ✗ {binary}	✓   ✗ {binary}

that might occur during transportation of the object. First, a rotational test is performed: the object is rotated at post-grasp position by  $+\beta$  and  $-\beta$  (around the z axis of the last link of the robot) one after the other. A shaking test is conducted afterwards in the sagittal plane (horizontal direction), where the robot shakes the object with a specific amplitude in a sinusoidal pattern for 10 seconds. For our experiments, we used an amplitude of 0.25 m and a peak acceleration of 10 m/s<sup>2</sup>. Note that if the user's robot cannot execute the sine pattern with these values (lifting speed, time, amplitude, and peak acceleration), then the user should clearly report the values used for their tests. The test is stopped if the object falls out of the hand at a given step. Since the shaking test is performed after the rotational test, we do not perform it if the rotational test fails. The results can be presented in terms of  $s/a$ , where  $s$  denotes the number of successful grasps and  $a = 6 * N$  or  $12 * N$  is the number of grasp attempts for a single object for all the object poses. Overall, we define the following conditions as trial failures:

- F1) If the gripper fingers or any used equipment knocks-off the object from its place while reaching-to-grasp;
- F2) If the object slips or rolls away while executing the grasp or while lifting the grasped object;
- F3) If the designed rotational test is failed;
- F4) If the designed shaking test is failed;
- F5) If no feasible hypotheses are found, e.g. due to robot kinematics, or object placement (which only applies to planners with integrated reachability search, [3], [26]);
- F6) If the hardware failed to respond due to communication drops, process timeouts, etc.

For each object the results can be summarized using the format specified in Tables I and II for experiments using single and a group of objects, respectively. The result format includes object name and ID, trial number, stable pose number before grasping (for single object grasping experiments), pickup order (denoting the order of objects being picked up for the experiments with group of objects), and the following data regarding the outcome of the experiments (for both single object and cluttered scenes): selected grasp ID along with total number of grasps generated, time to generate grasp, and binary success outcome of the lift, rotational and shaking tests. A

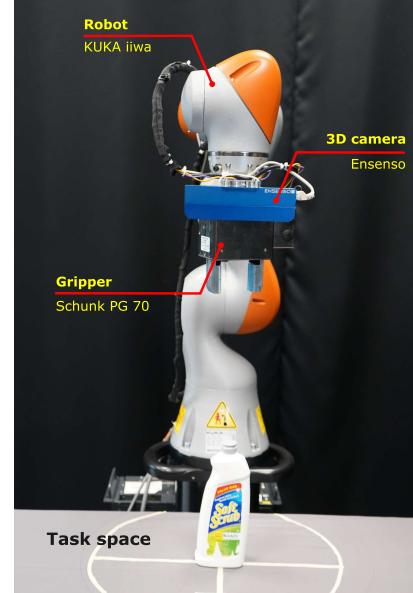


Fig. 5. Experimental setup used for tests.

final success rate (normalized average) based on all trials and poses can be reported for each object or group of objects. Section III presents the application of this protocol to several test cases. The supplementary material to this paper includes the benchmark and protocol templates for the test, and a video demonstrating the experimental procedure.

### III. EXPERIMENTAL EVALUATIONS

In this section, we apply the proposed protocol using specific robot setups and two grasp planners, based on both real and simulated experiments. The details are presented below.

#### A. Robotic Setup

For real tests, the robotic setup consists of a 7 degrees of freedom (DoF) KUKA LBR iiwa arm equipped with a Schunk PG70 parallel jaw gripper with flat fingers (Fig. 5). Maximum gripper stroke is 68 mm. In order to perceive and build the models of scene objects, a wrist-mounted depth camera (Ensenso N35) was used. Object point cloud models

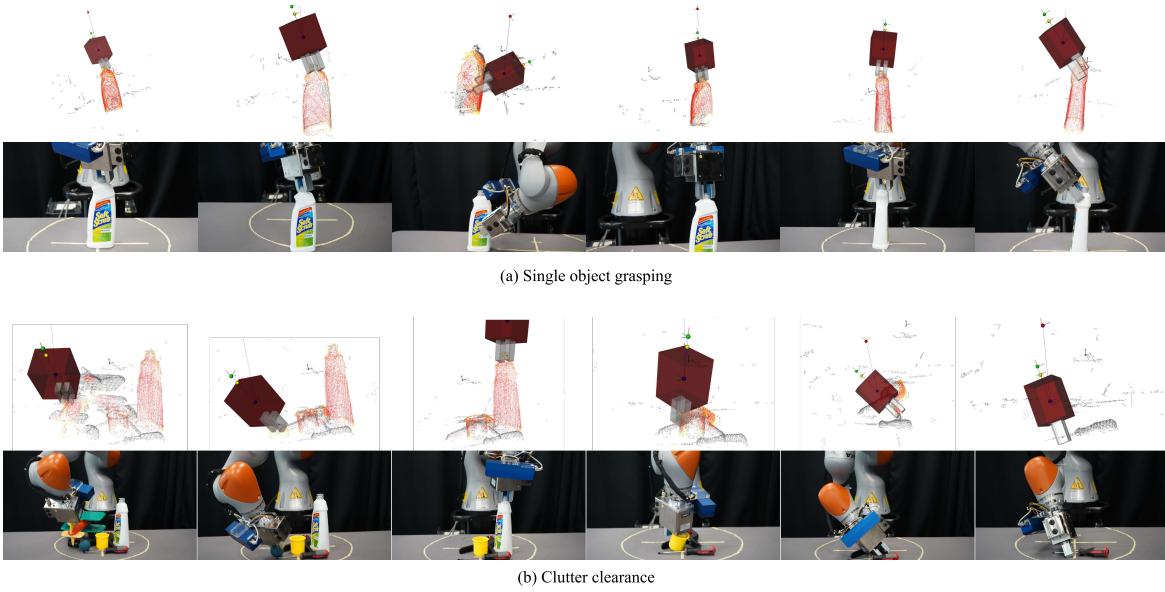


Fig. 6. Application of the LoCoMo-based grasp planner [3] to (a) single object grasping, and (b) clearing a group of objects.

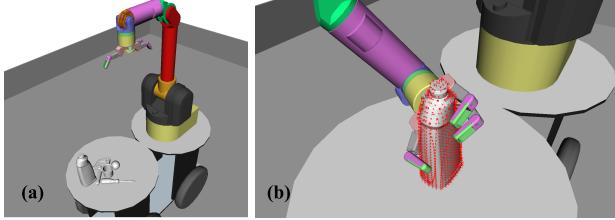


Fig. 7. OpenRAVE setup: (a) A WAM robot with a 3-finger Barrett hand, (b) an example grasp.

are built by moving the robot to four different locations, and the four point clouds acquired are stitched together. Cloud registration is trivial as the camera calibration with respect to the robot base frame is known beforehand. These built clouds, which are pre-processed to segment points lying on the ground plane, are fed to the grasp planners for hypotheses generation. The circular workspace seen in Fig. 5 is selected as explained in Section II-A. On the other hand, the simulation tests are performed using the OpenRAVE environment [2]. Test setup is made up of a WAM robot with a 3-finger Barrett hand, as shown in Fig. 7(a). Object models (in \*.dae file format) downloaded from the YCB website<sup>1</sup> are used for the tests. All experiments are conducted on a PC with an Intel Core i7-4790K CPU @4 GHz and 32 GB RAM.

### B. Grasp Planners

We have applied the protocol explained in Section II to two different grasp planers. The first one is a model- and learning-free planner executed using the real robot setup, and the second one is an off-the-shelf grasp planner available within the OpenRAVE library. Both planners are summarized below.

1) *Model-free LoCoMo-based planner*: This planner, proposed in [3], generates grasp hypotheses on arbitrarily shaped objects based on Local Contact Moment (LoCoMo) matching.

Along with partial point clouds of object/scene, it uses hand and arm kinematics for grasp generation. This method first determines the local similarity between object surface and gripper fingers using zero moment shift features. Highest score is given to the higher shape similarity in terms of their zero moment. Feasible hypotheses are then generated by searching for maximum likelihoods using LoCoMo metric. The generated hypotheses are ranked based on the product of integrals of LoCoMos for each finger of a gripper. From the results presented in that work, the method appears to work well with both individual objects and cluttered scenes. More technical details about the method and its application for a variety of scenes can be found in [3], [27]. A partial point cloud constructed by moving the robot to multiple locations is used for grasp generation. Unlike in [3], we constrain the method to find only the top 100 hypothesis, and the returned top grasp is executed. As mentioned earlier, for each object stable pose and test location, we repeat the grasp planning and execution  $N = 3$  times. For the task of clutter clearance, the whole clearance process is repeated three times. Example grasps from the experiments can be seen in Fig. 6.

2) *OpenRAVE grasp planner*: The simulation framework proposed in [2] provides a collection of tools spanning from a grasp to a motion planner. OpenRAVE implements a model-based grasp planner. It uniformly samples the object's surface to obtain the approach direction for the gripper. It then tests every approaching direction aligning the gripper's palm, closes the fingers and determines the contacts of the fingers on the object. To rank each grasp configuration, OpenRAVE checks whether the grasp is force closure or not, and tests the robustness of the grip under possible misalignment of the contact points. Two parameters drive the number of approach directions to test onto the object's surface, the density of the sampling, and the radius of the sphere (for dealing with contact point uncertainty) at each sampled point. The latter term defines the maximum allowed misalignment between the inward normal to the surface and the gripper's approach

<sup>1</sup><http://www.ycbbenchmarks.com/object-models/>

direction. All objects (shown in Fig. 2) have been tested using the models obtained from the YCB dataset, and we kept the two parameters fixed at 0.005 and 0.7, respectively. To obtain feasible grasps in every case, we let the algorithm span over all grasping candidates. We repeated the experiments three times per object for each placement pose. As for the clutter, we challenged the planner to clear three random scenes using all objects. Example grasp execution on one of the test object models (scrub cleanser) is shown in Fig. 7(b).

### C. Results

In our experiments we use the following parameters (introduced in Section II-B):  $\alpha = 90$ ,  $\beta = 90$  and  $N = 3$ . Results obtained for grasping objects in single and cluttered scenes using the LoCoMo-based method are summarized in Tables III to VII using the format previously presented. Tables III and IV show the results for standard placement, and Tables V and VI show the results for mirrored placement. Due to space limitations we show the results averaged over all three trials. Similarly, grasping of objects in clutter is summarized in Table VII. Detailed results (in extended spreadsheets) with the LocoMo-based grasp planner can be found in <https://git.io/JeVK5> and with the OpenRAVE grasp planner in <https://git.io/JegSH>.

The average results C1–C5 over all trials and poses for each object (IDs: 56, 53, 20, 43, 46, 67, 61) respectively are C1: (78.7, 82.2, 77.4, 73.4, 76.2, 80.1, 88.9), C2: (1.3, 1.2, 3.7, 0.8, 5.4, 8.2, 4.7), C3: (94%, 100%, 87%, 100%, 97%, 90%, 100%), C4: (94%, 100%, 87%, 100%, 97%, 90%, 100%), C5: (94%, 100%, 81%, 100%, 97%, 90%, 100%) for real experiments. In the case of clutter experiments with real scene the averages are C1: 83.66, C2: 4.54, C3, C4 and C5: 100%.

The lifting experiments succeeded in all cases except in one case with yellow cup as marked in Table IV, as the object was pushed away by the fingers while grasping it. Overall, results suggest that the LoCoMo planner performs well in grasping and lifting individual objects, as well as clearing heaps of objects. Few failures were recorded during the trials. Most failures were for the “scrub cleanser” object, because it is slightly wider than the maximum stroke of the gripper, which made it slip/roll away during grasp attempts. For grasping objects in clutter, LoCoMo exhibited 100% success rate. As LoCoMo does not use object models, it removes objects in a random fashion and is therefore unsuitable for tasks requiring specific picking order. Due to limitations encountered with physics engines, we were unable to perform the lift, rotational and shaking tests in our simulation experiments. For the same reason, no grasp failures were observed in OpenRAVE. Furthermore, the grasp generation time is remarkably high (e.g. 206.8 seconds for clutter experiments) due to its method for finding grasps, explained in Section III-B2. Since a fully assembled model of the toy plane object is not available, we were unable to test the OpenRAVE planner for this object.

### IV. CONCLUSIONS

We presented a benchmarking protocol for evaluation of grasp planning algorithms, describing in detail how to setup

the workspace, object choices and placements, grasp execution, and scoring. As future work, we plan to focus on adding obstacles in the scene setup and detailed instructions on how to setup repeatable piles of objects for cluttered scene experiments, extending the work in [21] to larger variety of objects and degree of difficulty in terms of picking order and additional task constraints. We also plan to add more stability checks and levels of difficulty for grasp robustness test, such as linearly increasing/decreasing the acceleration in each cycle, and linearly increasing the maximum velocity between cycles.

### ACKNOWLEDGMENT

The authors would like to express their gratitude to Prof. Matei Ciocarlie for useful discussions and comments since the early stages of the work. The authors also extend their thanks to Dr. Rajasekaran for his help with the experiments.

### REFERENCES

- [1] V.-D. Nguyen, “The synthesis of stable force-closure grasps.” MIT, Tech. Rep., 1986.
- [2] R. Diankov and J. Kuffner, “OpenRAVE: A planning architecture for autonomous robotics,” *Robotics Institute - CMU, USA. Tech. Rep. CMU-RI-TR-08-34*, vol. 79, 2008.
- [3] M. Adjigble, N. Marturi, V. Ortenzi, V. Rajasekaran, P. Corke, and R. Stolk, “Model-free and learning-free grasping by local contact moment matching,” in *IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, 2018, pp. 2933–2940.
- [4] M. Pozzi, A. Sundaram, M. Malvezzi, D. Prattichizzo, and M. A. Roa, “Grasp quality evaluation in underactuated robotic hands,” in *IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, 2016, pp. 1946–1953.
- [5] R. Krug, A. J. Lilenthal, D. Kragic, and Y. Bekiroglu, “Analytic grasp success prediction with tactile feedback,” in *IEEE Int. Conf. Robotics and Automation*, 2016, pp. 165–171.
- [6] S. Kumra and C. Kanan, “Robotic grasp detection using deep convolutional neural networks,” in *IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, 2017, pp. 769–776.
- [7] S. Levine, P. Pastor, A. Krizhevsky, J. Ibarz, and D. Quillen, “Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection,” *Int. J. Robotics Research*, vol. 37, no. 4-5, pp. 421–436, 2018.
- [8] Y. Lin and Y. Sun, “Robot grasp planning based on demonstrated grasp strategies,” *Int. J. Robotics Research*, vol. 34, no. 1, pp. 26–42, 2015.
- [9] M. Catalano, G. Grioli, E. Farnioli, A. Serio, C. Piazza, and A. Bicchi, “Adaptive synergies for the design and control of the Pisa/IIT SoftHand,” *Int. J. Robotics Research*, vol. 33, no. 5, pp. 768–782, 2014.
- [10] A. Dollar and R. Howe, “The SDM hand: A highly adaptive compliant grasper,” in *Experimental Robotics*. Springer, 2009, pp. 3–11.
- [11] M. Prats, P. Martinet, A. P. del Pobil, and S. Lee, “Vision/force control in task-oriented grasping and manipulation,” in *IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, 2007, pp. 1320–1325.
- [12] K. Hang, M. Li, J. A. Stork, Y. Bekiroglu, F. T. Pokorny, A. Billard, and D. Kragic, “Hierarchical fingertip space: A unified framework for grasp planning and in-hand grasp adaptation,” *IEEE Trans. Robotics*, vol. 32, no. 4, pp. 960–972, 2016.
- [13] A. Sahbani, S. El-Khoury, and P. Bidaud, “An overview of 3D object grasp synthesis algorithms,” *Robotics and Autonomous Systems*, vol. 60, no. 3, pp. 326 – 336, 2012.
- [14] S. Caldera, A. Rassau, and D. Chai, “Review of deep learning methods in robotic grasp detection,” *Multimodal Technologies and Interaction*, vol. 2, no. 3, 2018.
- [15] A. Bicchi and V. Kumar, “Robotic grasping and contact: a review,” in *IEEE Int. Conf. Robotics and Automation*, 2000, pp. 348–353.
- [16] J. Bohg, A. Morales, T. Asfour, and D. Kragic, “Data-driven grasp synthesis — a survey,” *IEEE Trans. Robotics*, vol. 30, no. 2, pp. 289–309, 2014.
- [17] N. Correll, K. Bekris, D. Berenson, O. Brock, A. Causo, K. Hauser, K. Okada, A. Rodriguez, J. Romano, and P. Würman, “Analysis and observations from the first Amazon Picking Challenge,” *IEEE Trans. Automation Science and Engineering*, vol. 15, no. 1, pp. 172–188, 2018.

TABLE III  
RESULTS FOR GRASPING SINGLE OBJECTS USING LoCoMo-BASED PLANNER FOR OBJECT POSES P1 TO P3: STANDARD PLACEMENT

Object	Stable Pose	Location and pose														
		P1					P2					P3				
		<sup>2</sup> C1	C2	C3	C4	C5	C1	C2	C3	C4	C5	C1	C2	C3	C4	C5
Yellow cup (56)	1	77	1.39	100	100	100	65.6	1.37	100	100	100	98.3	1.29	100	100	100
	2	73	1.49	100	100	100	82.3	1.41	100	100	100	42.3	1.27	100	100	100
Racquetball (53)	1	84.6	1.20	100	100	100	73	1.18	100	100	100	71.33	1.23	100	100	100
Scrub Cleanser (20)	1	96.6	3.20	100	100	100	93	3.6	100	100	100	1	0.68	<sup>3</sup> (F5, F6)	33	33
	2	67.3	2.89	66 (F2)	66	66	100	3.39	100	100	66 (F4)	97.66	2.74	100	100	100
Screwdriver (43)	1	100	1.10	100	100	100	94.3	1.06	100	100	100	16	0.32	100	100	100
Big clamp (46)	1	56.3	1.83	100	100	100	73	2.06	66 (F1)	66	66	51.3	2.63	100	100	100
Toy plane (67)	1	69.3	16	100	100	100	100	21.7	100	100	100	26	2.03	<sup>3</sup> (F5, F6)	33	33
Chain (61)	1	76	4.28	100	100	100	100	4.62	100	100	100	81.6	4.75	100	100	100

<sup>1</sup>P1 to P3 correspond to the placement poses (1)–(3) mentioned in Section II-B.

<sup>2</sup>C1 to C5 correspond to respectively the final five columns mentioned in Table I. C1 shows only the average of total grasps of all three trials for an object. C3, C4 and C5 are given as percentages.

<sup>3</sup>F# (in red) indicates the failure case (mentioned in Section II-B) that occurred during execution.

TABLE IV  
RESULTS FOR GRASPING SINGLE OBJECTS USING LoCoMo-BASED PLANNER FOR OBJECT POSES P4 TO P6: STANDARD PLACEMENT

Object	Stable Pose	Location and pose														
		P4					P5					P6				
		C1	C2	C3	C4	C5	C1	C2	C3	C4	C5	C1	C2	C3	C4	C5
Yellow cup (56)	1	100	1.243	66 (F1)	66	66	<sup>2</sup> NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	2	90.6	1.17	100	100	100	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Racquetball (53)	1	100	1.03	100	100	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Scrub Cleanser (20)	1	100	3.8	100	100	100	96.3	4.34	100	100	100	100	4.61	100	100	100
	2	36.3	2.01	100	100	100	44.3	1.56	66 (F5)	66	66	66	1.84	66 (F5)	66	66
Screwdriver (43)	1	100	0.97	100	100	100	53.6	0.72	100	100	100	100	28.3	0.39	100	100
Big clamp (46)	1	69	2.07	100	100	100	98.3	3.24	100	100	100	100	65	3.78	100	100
Toy plane (67)	1	99.3	4.85	100	100	100	100	8.01	100	100	100	100	100	4.83	100	100
Chain (61)	1	100	5.19	100	100	100	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

<sup>1</sup>P4 to P6 correspond to the placement poses (4)–(6) mentioned in Sec. II-B.

<sup>2</sup>NA corresponds to “Not Applicable”. Used when the trial is not performed due to object symmetry.

TABLE V  
RESULTS FOR GRASPING SINGLE OBJECTS USING LoCoMo-BASED PLANNER FOR OBJECT POSES P1 TO P3: MIRRORED PLACEMENT

Object	Stable Pose	Location and pose														
		P1					P2					P3				
		C1	C2	C3	C4	C5	C1	C2	C3	C4	C5	C1	C2	C3	C4	C5
Yellow cup (56)	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Racquetball (53)	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Scrub Cleanser (20)	1	88.3	3.87	100	100	100	90.6	5.34	100	100	100	2.3	2.18	<sup>33</sup> (F5)	33	33
	2	45	3.11	100	100	100	93	5.85	100	100	66 (F2)	81.3	3.02	100	100	66 (F2)
Screwdriver (43)	1	100	1.06	100	100	100	49.3	0.51	100	100	100	93.6	0.97	100	100	100
Big clamp (46)	1	100	3.05	100	100	100	69	2.8	100	100	100	80	3.2	100	100	100
Toy plane (67)	1	100	5.8	100	100	100	100	10.6	66 (F1)	66	66	30.67	2.77	100	100	100

TABLE VI  
RESULTS FOR GRASPING SINGLE OBJECTS USING LOCOMO-BASED PLANNER FOR OBJECT POSES P4 TO P6: MIRRORED PLACEMENT

Object	Stable Pose	Location and pose														
		P4					P5					P6				
		C1	C2	C3	C4	C5	C1	C2	C3	C4	C5	C1	C2	C3	C4	C5
Yellow cup (56)	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
	2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
Racquetball (53)	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
Scrub Cleanser (20)	1	80.6	3.13	100	100	100	100	14.12	100	100	66 (F2)	100	5.23	33 (F2, F6)	33	33
	2	74.6	2.31	100	100	100	96.6	2.47	100	100	100	87.6	7	4.30	100	100
Screwdriver (43)	1	98.3	1.02	100	100	100	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Big clamp (46)	1	100	2.53	100	100	100	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Toy plane (67)	1	76	5.12	100	100	100	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

TABLE VII  
RESULT FORMAT FOR CLEARING A GROUP OF OBJECTS USING LOCOMO-BASED PLANNER

Trial #	Object name (YCB ID)	Pickup order	Selected (total Feasible)	Time (seconds)	Lift Test	Rotational Test	Shaking Test
1	Yellow Cup (56)	4	19 (100)	2.874	✓	✓	✓
	Racquette ball (53)	2	6 (60)	6.083	✓	✓	✓
	Scrub Cleanser (20)	3	7 (69)	6.665	✓	✓	✓
	Flat Screwdriver (43)	6	1 (10)	0.229	✓	✓	✓
	Big Clamp (46)	5	21 (100)	2.36	✓	✓	✓
	Toy plane (67)	1	1 (82)	5.891	✓	✓	✓
2	Yellow Cup (56)	2	54 (100)	6.121	✓	✓	✓
	Racquette ball (53)	3	65 (73)	5.185	✓	✓	✓
	Scrub Cleanser (20)	4	51 (52)	5.106	✓	✓	✓
	Flat Screwdriver (43)	6	1 (90)	1.073	✓	✓	✓
	Big Clamp (46)	5	63 (87)	4.36	✓	✓	✓
	Toy plane (67)	1	54 (100)	6.424	✓	✓	✓
3	Yellow Cup (56)	5	60 (100)	3.951	✓	✓	✓
	Racquette ball (53)	2	2 (83)	6.173	✓	✓	✓
	Scrub Cleanser (20)	4	56 (100)	5.173	✓	✓	✓
	Flat Screwdriver (43)	3	39 (100)	5.837	✓	✓	✓
	Big Clamp (46)	6	4 (100)	1.521	✓	✓	✓
	Toy plane (67)	1	2 (100)	6.768	✓	✓	✓

- [18] Y. Sun, J. Falco, N. Cheng, H. Chosi, E. Engeberg, N. Pollard, M. Roa, and Z. Xia, "Robotic Grasping and Manipulation competition: task pool," in *Robotic grasping and manipulation*, Y. Sun and J. Falco, Eds. Springer-Verlag, 2018, pp. 1–18.
- [19] B. Calli, A. Walsman, A. Singh, S. Srinivasa, P. Abbeel, and A. M. Dollar, "Benchmarking in manipulation research: Using the Yale-CMU-Berkeley object and model set," *IEEE Robotics Automation Magazine*, vol. 22, no. 3, pp. 36–52, 2015.
- [20] J. Leitner, A. W. Tow, N. Stürnerhauf, J. E. Dean, J. W. Durham, M. Cooper, M. Eich, C. Lehnert, R. Mangels, C. McCool, P. T. Kujala, L. Nicholson, T. Pham, J. Sergeant, L. Wu, F. Zhang, B. Upcroft, and P. Corke, "The ACRV picking benchmark: A robotic shelf picking benchmark to foster reproducible research," in *IEEE Int. Conf. Robotics and Automation*, 2017, pp. 4705–4712.
- [21] P. Triantafyllou, H. Mnyusiwalla, P. Sotiropoulos, M. A. Roa, D. Russell, and G. Deacon, "A benchmarking framework for systematic evaluation of robotic pick-and-place systems in an industrial grocery setting," in *IEEE Int. Conf. Robotics and Automation*, 2019, pp. 6692–6698.
- [22] P. Sotiropoulos, M. A. Roa, M. F. Martins, W. Fried, H. Mnyusiwalla, P. Triantafyllou, and G. Deacon, "A benchmarking framework for systematic evaluation of compliant under-actuated soft end effectors in an industrial context," in *IEEE-RAS Int. Conf. Humanoid Robots*, 2018, pp. 280–283.
- [23] J. Falco, K. Van Wyk, S. Liu, and S. Carpin, "Grasping the performance: Facilitating replicable performance measures via benchmarking and standardized methodologies," *IEEE Robotics & Automation Magazine*, vol. 22, no. 4, pp. 125–136, 2015.
- [24] S. Ulbrich, D. Kappler, T. Asfour, N. Vahrenkamp, A. Bierbaum, M. Przybylski, and R. Dillmann, "The OpenGRASP benchmarking suite: An environment for the comparative analysis of grasping and dexterous manipulation," in *IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, 2011, pp. 1761–1767.
- [25] G. Kootstra, M. Popović, J. A. Jørgensen, D. Krägic, H. G. Petersen, and N. Krüger, "Visgrab: A benchmark for vision-based grasping," *Paladyn*, vol. 3, no. 2, pp. 54–62, 2012.
- [26] N. Marturi, M. Kopicki, A. Rastegarpanah, V. Rajasekaran, M. Adjigble, R. Stolk, A. Leonardis, and Y. Bekiroglu, "Dynamic grasp and trajectory planning for moving objects," *Autonomous Robots*, vol. 43, no. 5, pp. 1241–1256, 2019.
- [27] M. Adjigble, N. Marturi, V. Ortenzi, and R. Stolk, "An assisted telemanipulation approach: combining autonomous grasp planning with haptic cues," in *IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, 2019.