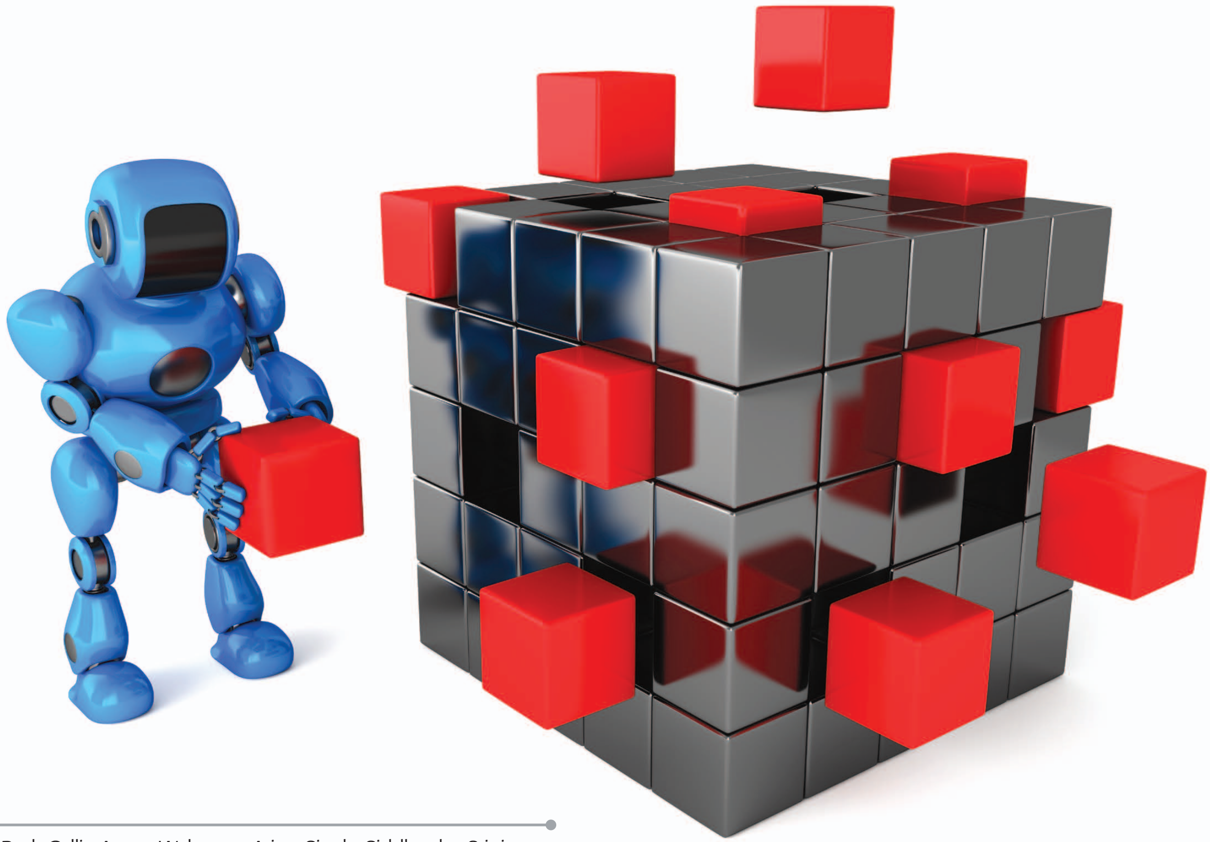
Benchmarking in Manipulation Research

Using the Yale–CMU–Berkeley Object and Model Set


©ISTOCKPHOTO.COM/ALENGO

By Berk Calli, Aaron Walsman, Arjun Singh, Siddhartha Srinivasa, Pieter Abbeel, and Aaron M. Dollar

In this article, we present the Yale–Carnegie Mellon University (CMU)–Berkeley (YCB) object and model set, intended to be used to facilitate benchmarking in robotic manipulation research. The objects in the set are designed to cover a wide range of aspects of the manipulation problem. The set includes objects of daily life with different shapes, sizes, textures, weights, and rigidities as well as some widely used manipulation tests. The associated database provides high-resolution red, green, blue, plus depth (RGB-D) scans, physical properties, and geometric models of the objects for easy incorporation into manipulation and planning software platforms. In addition to describing the objects and models in the set along with how they were chosen and derived, we provide a framework and a number of example task protocols, laying out how the set can be used to quantitatively evaluate a range of manipulation approaches, including planning, learning, mechanical design, control, and many others. A comprehensive literature survey on the existing benchmarks and object data sets is also presented, and their scope and limitations are discussed. The YCB set will be freely distributed to research groups worldwide at a series of tutorials at robotics conferences. Subsequent sets will be, otherwise, available to purchase at a reasonable cost. It is our hope that the ready

Digital Object Identifier 10.1109/MRA.2015.2448951
Date of publication: 11 September 2015

1070-9932/15©2015IEEE

availability of this set along with the ground laid in terms of protocol templates will enable the community of manipulation researchers to more easily compare approaches as well as continually evolve standardized benchmarking tests and metrics as the field matures.

Benchmarking in Robotics Research

Benchmarks are crucial for the progress of a research field, allowing performance to be quantified to give insight into the effectiveness of an approach compared with alternative methods. In manipulation research, particularly in robotic manipulation, benchmarking and performance metrics are challenging due largely to the enormous breadth of the application and task space researchers are working toward. The majority of research groups have, therefore, selected for themselves a set of objects and/or tasks that they believe are representative of the functionality that they would like to achieve/assess. The chosen tasks are often not sufficiently specified or general enough such that others can repeat them. Moreover, the objects used may also be insufficiently specified and/or not available to other researchers (e.g., they may have been custom-fabricated or are only available for purchase in certain countries). Unfortunately, such an approach prevents the analysis of experimental results against a common basis and, therefore, makes it difficult to quantitatively interpret performance.

There have been a limited number of efforts to develop object and model sets for benchmarking in robotic manipulation. Most of these have focused on providing mesh model databases of objects, generally for object-recognition or grasp-planning purposes (see [1]–[4], with a thorough overview provided in the "Related Work" section). There have, however, been a few instances of proposed object/task sets for which the physical objects are available to researchers. Access to the objects is crucial to performance benchmarking as many aspects of the manipulation process cannot be modeled, thereby requiring experiments to demonstrate success or examine failure modes.

Overview

In this article, we present an object set for robotic manipulation research and performance evaluation, a framework for standard task protocols, and a number of example protocols along with experimental implementation. The object set is specifically designed to allow for widespread dissemination of the physical objects and manipulation scenarios. The objects were selected based on a survey of the most common objects utilized in the robotics field as well as the prosthetics and rehabilitation literature (in which procedures are developed to assess the manipulation capabilities of patients) along with a number of additional practical constraints. Along with the physical objects, textured mesh models, high-quality images, and point-cloud data of the objects are provided together with their physical properties (i.e., major dimensions and mass) to enable realistic simulations. These data are all available online at http://rll.eecs.berkeley.edu/ycb/. The models are integrated into the MoveIt motion-planning tool [5] and the robot operating system (ROS) to demonstrate their use. The set will be freely distributed to research groups worldwide at a series of tutorials at robotics conferences and will be, otherwise, available at a reasonable purchase cost. Our goal is to do as much as possible to facilitate the widespread usage of a common set of objects and tasks to allow easy comparison of results between research groups worldwide.

In choosing the objects in the set, a number of issues were considered. The objects, many of which are commercial household products, should span a variety of shapes, sizes, weights, rigidities, and textures as well as a wide range of manipulation applications and challenges. In addition, several practical constraints were considered, including ease of shipping and storage, reasonable overall cost, durability, perishability, and product longevity (the likelihood that the objects/products will be available in the future).

In addition to the object and model set, we provide a systematic approach to define manipulation protocols and benchmarks using the set. The protocols define the experimental setup for a given manipulation task and provide procedures to follow, and the benchmarks provide a scoring scheme for the quantification of performance for a given protocol. To facilitate the design of well-defined future protocols and benchmarks, guidelines are provided through a template. The protocols and benchmarks are intended to generally be platform-independent to allow for comparisons of approaches across platforms. Along with the template and guidelines, we present a number of preliminary protocols and benchmarks. These serve both as examples of how to utilize the template and as useful procedures for quantitatively evaluating various aspects of robotic manipulation. The implementation of these benchmarks on real robotic systems is also provided to demonstrate the benchmarks' abilities to quantitatively evaluate the manipulation capabilities of various systems.

The models are integrated into the MoveIt motion-planning tool [5] and the robot operating system (ROS) to demonstrate their use.

We expect to continually expand on this work not only by our own efforts (adding more objects' properties and additional benchmarks) but also, more importantly, via our web portal: http://www.ycbbenchmarks.org/. Through this web portal, the user community can engage in this effort, with users proposing changes to the object set and putting forth their own protocols, benchmarks, and so on.

Related Work

For benchmarking in manipulation, specifying an object set is useful for the standardization of experimental conditions. Table 1 summarizes the object sets that have been proposed for manipulation tasks in the fields of robotics, prosthetics,

Table 1. Object data sets in the literature (sorted by year).

	Data Set Name	Year	Data Type	Purpose	Number of Objects/ Categories	Physical Objects Available	Website
1	BigBIRD [1]	2014	Meshes with texture + HQ images	Object recognition	100	No	http://rll.eecs.berkeley.edu/bigbird
2	Amazon Picking Challenge [7]	2014	Shopping list	Grasping	27	Yes	http://amazonpickingchallenge.org/
3	SHREC'14 [2]	2014	Mesh models	Object retrieval	8,987/171	No	http://www.itl.nist.gov/iad/vug/sharp/contest/2014/Generic3D/
4	SHREC'12 [21]	2012	Mesh models	Object retrieval	1,200/60	No	http://www.itl.nist.gov/iad/vug/sharp/contest/2012/Generic3D/
5	The KIT object models database [19]	2012	Mesh with texture, stereo images	Recognition, localization, and manipulation	100	No	http://i61p109.ira.uka.de/ObjectModels WebUI/
6	VisGraB [22]	2012	Stereo images	Manipulation	18	No	http://www.robwork.dk/visgrab/
7	The object segmentation database [17]	2012	RGB-D images	Object segmentation	N/A	No	http://users.acin.tuwien.ac.at/arichtsfeld/?site=4
8	Toyohashi shape benchmark [23]	2012	Mesh models	Object retrieval	10k/352	No	http://www.kde.cs.tut.ac.jp/benchmark/tsb/
9	The Willow Garage object recognition challenge [24]	2012	RGB-D images	Object recognition	N/A	No	http://www.acin.tuwien.ac.at/forschung/v4r/mitarbeiterprojekte/willow/
10	SHREC'11 [25]	2011	Mesh models	Object retrieval	600	No	http://www.itl.nist.gov/iad/vug/sharp/contest/2011/NonRigid/
11	Berkeley 3-D object data set [26]	2011	RGB-D data set of room scenes	Object detection	N/A	No	http://kinectdata.com/
12	RGB-D object data set [27]	2011	RGB-D Data set	Object detection and recognition	300/51	No	http://rgbd-dataset.cs.washington.edu/
13	The open GRASP benchmarking suite [20]	2011	Mesh with texture, stereo images	Grasping	Uses KIT database	No	http://opengrasp.sourceforge.net/benchmarks.html
14	SHREC 2010 [28]	2010	Mesh models	Object retrieval	3168/43	No	http://tosca.cs.technion.ac.il/book/shrec_robustness2010.html
15	The Columbia grasp database [3]	2009	Mesh models	Grasping	~8,000	No	http://grasping.cs.columbia.edu/
16	Benchmark set of domestic objects [6]	2009	Shopping list	Robotic manipulation	43	Yes	http://www.hsi.gatech.edu/hrl/object_list_v092008.shtml
17	Bonn architecture benchmark [29]	2009	Mesh models	Object retrieval	2,257	No	ftp://ftp.cg.cs.unibonn.de/pub/outgoing/ArchitectureBenchmark
18	Engineering shape benchmark [30]	2008	Mesh models	Object retrieval	867	No	https://engineering.purdue.edu/PRECISE/shrec08
19	3-D object retrieval benchmark [31]	2008	Mesh models	Object retrieval	800/40	No	http://www.itl.nist.gov/iad/vug/sharp/benchmark/
20	McGill 3-D shape benchmark [32]	2008	Mesh models	Shape retrieval	N/A	No	http://www.cim.mcgill.ca/~shape/benchMark/

(Continued)

Table 1. Object data sets in the literature (sorted by year). (*Continued*)

	Data Set Name	Year	Data Type	Purpose	Number of Objects/ Categories	Physical Objects Available	Website
21	The Toronto Rehabilitation Institute hand-function test [33]	2008	Commercial kit/no model data	Prosthetics and rehabilitation	14	No	http://www.rehabmeasures.org/Lists/RehabMeasures/PrintView.aspx?ID=1044
22	GRASSP [9]	2007	Commercial kit/no model data	Prosthetics and rehabilitation	N/A	Yes	http://grassptest.com/
23	AIM@SHAPE shape repository [16]	2006	Mesh models	General	1,180	No	http://shapes.aim-atshape.net/viewmodels.php
24	The Princeton shape benchmark [18]	2004	Mesh models	Shape-based retrieval	1,814	No	http://shape.cs.princeton.edu/benchmark/
25	Mesh deformation data set [34]	2004	Mesh models	Mesh transformation	N/A/13	No	http://people.csail.mit.edu/sumner/research/deftransfer/data.html
26	NTU 3-D model benchmark [35]	2003	Mesh models	Shape retrieval	1,833	No	http://3d.csie.ntu.edu.tw/
27	SHAP [8]	2002	Commercial kit/no model data	Prosthetics and rehabilitation	—	Yes	http://www.shap.ecs.soton.ac.uk/
28	Action research arm test [10]	1981	Commercial kit/no model data	Prosthetics and rehabilitation	19	Yes	http://saliarehab.com/actionresearcharmtestarat.html
29	Jebsen–Taylor hand-function test [11]	1969	Commercial kit/no model data	Prosthetics and rehabilitation	N/A	Yes	N/A
30	The ITI database [36]	N/A	Mesh models	Object retrieval	544/13	No	http://vcl.iti.gr/3d-object-retrieval/
31	Model bank library [37]	N/A	Mesh with texture	General	1,200	No	http://digimation.com/3dlibraries/model-bank-library/
32	SketchUp [4]	N/A	Mesh with and without texture	General	N/A	No	https://3dwarehouse.sketchup.com/
33	RoboCup at home [38]	Multiple	No data	Manipulation	N/A	No	http ://www.robocupathome.org/

and rehabilitation. Even though there have been many efforts that provide data sets of object mesh models that are useful for many simulation and planning applications as well as for benchmarking in shape retrieval, these data sets have limited utility for manipulation benchmarking for several reasons.

- Since most of them are not designed specifically for manipulation benchmarking, the selected objects do not usually cover the shape and function variety needed for a range of manipulation experiments.
- None of these databases provides the objects' physical properties, which are necessary to conduct realistic simulations.
- Most importantly, the vast majority of objects in these sets are not easily accessible by other researchers, preventing their use in experimental work.

Exceptions to this include [6], which provides an online shopping list (though it is now outdated, with many dead links), and the recently announced Amazon Picking Challenge [7], which provides a shopping list to purchase objects meant for a narrow bin-picking task. In the prosthetics and rehabilitation field, commercial kits are available for upper-limb assessment tests [8]–[11]. While demonstrating the benefits of utilizing a standard set for manipulation assessment, the scope of these kits is limited for benchmarking in robotics as they are not representative of a wide range of manipulation tasks. Our effort is unique in that it provides a large amount of information about the objects necessary for many simulation and planning approaches, makes the actual objects readily available for researchers to utilize experimentally, and includes a wide range of objects to span many different manipulation applications.

We provide a detailed overview of prior related benchmarking efforts, discussing their scope and limitations. For organization purposes, we first discuss work primarily related to robotic manipulation (including vision and learning applications), then efforts in rehabilitation, including prosthetics.

Robotic Manipulation

The necessity of manipulation benchmarks is highly recognized in the robotics community [12]–[14] and continues to be an active topic of discussion at workshops on robotic manipulation (see [15]). As mentioned earlier, the majority of prior work related to object sets has involved just object images and models (with varying degrees of information, from purely shape information to textural plus shape). Such work has often been created for research in computer vision (see [2], [16], and [17]). There have also been a number of shape/texture sets designed for/by the robotics community, particularly for applications such as planning and learning. The Columbia grasp database [3] rearranges the object models of the Princeton shape benchmark [18] for robotic manipulation and provides mesh models of 8,000 objects together with a number of successful grasps per model. Such a database is especially useful for implementing machine-learning-based grasp synthesis algorithms in which large amounts of labeled data are required for training the system. A multipurpose object set, which also targets manipulation, is the Karlsruhe Institute of Technology (KIT) object models database [19] which provides stereo images and textured mesh models of 100 objects. While there are a large number of objects in this database, the shape variety is limited, and like the previously mentioned data sets, the exact objects are typically not easily accessible to other researchers due to regional product differences or variation over time, and they generally cannot be purchased in one place as a set.

The object set is specifically designed to allow for widespread dissemination of the physical objects and manipulation scenarios.

There have only been two robotics-related efforts in which the objects are made relatively available. The household objects list [6] provides good shape variety that is appropriate for manipulation benchmarking as well as a shopping list for making the objects more easily accessible to researchers. Unfortunately, the list is outdated, and most objects are no longer available. The three-dimensional (3-D) models of objects in [6] are not supplied, which prevents the use of the object set in simulations. Very recently, the Amazon Picking Challenge [7] also provides a shopping list for items, but those were chosen to be specific to the bin-picking application and do not have models associated with them.

In terms of other robotic manipulation benchmarking efforts, a number of simulation tools have been presented in the literature. The OpenGRASP benchmarking suite [20] presents a simulation framework for robotic manipulation. The benchmarking suite provides test cases and setups and a standard evaluation scheme for the simulation results. So far, a model-based grasp synthesis benchmark has been presented using this suite. VisGraB [22] provides a benchmark framework for grasping unknown objects. The unique feature of this software is its utilization of real stereo images of the target objects for grasp synthesis as well as execution and evaluation of the result in a simulation environment. For gripper and hand design, benchmark tests [39], [40] are proposed for evaluating the ability of the grippers to hold an object, but only cylindrical objects are used.

Prosthetics and Rehabilitation

In the general field of rehabilitation and upper-limb prosthetics, there are a number of evaluation tools used by therapists to attempt to quantify upper-limb function in humans. Some of these are commercially available, clinically verified, and have been substantially covered in the literature, including normative data to compare a patient's performance to baselines. While some tools are commonly used, other tests have only been proposed in the literature and not (yet, at least) been widely utilized. Many of these tests aim to evaluate the ability of patients to perform tasks that contribute to activities of daily living.

The tests that are commercially available are the box-and-blocks test [41]; the nine-hole peg test [42]; the Jebsen–Taylor hand-function test [11]; the action research arm test (ARAT) [10]; the graded redefined assessment of strength, sensibility, and prehension (GRASSP) test [9]; and the Southampton hand-assessment procedure (SHAP) [8]. The setups for the box-and-blocks and nine-hole peg tests are very specific, with evaluation based on timed movements of simple objects. The setup for the Jebsen–Taylor hand-function test includes objects for manipulation actions, such as card turning, and moving small (paper clips, bottle caps), light (empty cans), and heavy objects (1-lb weighted cans), but it utilizes a small number of objects of limited shape and size variety. The ARAT assesses upper-limb function, and its commercial set [43] contains objects such as wooden blocks of various sizes, glasses, a stone, a marble, washers, and bolts. The test proposes actions like placing a washer over a bolt and pouring water from one glass into another. The GRASSP measure has also been proposed for the assessment of upper-limb impairment. It is based on a commercial kit available in [44]. Apart from a specialized manipulation setup, the kit also includes the nine-hole peg test, jars, and a bottle. The SHAP setup includes some objects of daily living, such as a bowl, a drink carton, and a jar, together with some geometrical shapes. Patients are requested to perform a variety of manipulation tasks, mostly involving transporting objects but also including pouring a drink, opening the jar, and so on. Considering manipulation benchmarking in robotics, the box-and-blocks, nine-hole peg, and Jebsen–Taylor hand-function tests are far from providing an adequate object variety for deriving new benchmarks. Despite enabling a larger possibility of manipulation tasks than the previously mentioned setups, the GRASSP and SHAP setups are still bounded to a limited number of tasks, and both are pricey (currently around US$1,300 and US$3,000, respectively).

Some well-known tests that do not provide a commercial setup are the grasp-and-release test [45], the Toronto Rehabilitation Institute hand-function test [33], and the activities measure for upper-limb amputees (AM-ULA) [46]. The grasp-and-release test is proposed for evaluating the performance of neuroprosthetic hands. For this test, detailed descriptions of the objects are given, but the objects are not easily obtainable, and the set includes an outdated object, i.e., a videotape. The Toronto Rehabilitation Institute hand-function test (also known as the Rehabilitation Engineering Laboratory hand-function test [47]) evaluates the palmar (power) and lateral (precision) grasp abilities of individuals using an object set comprising a mug, a book, a piece of paper, a soda can, dice, a pencil, and so on. Even though it is claimed that the objects used in this test are easily obtainable, maintaining the exact object definitions is hard, and one of the objects is an outdated cellular phone. The AM-ULA defines several quality measures for assessing the manipulation tasks, and various daily activities are proposed for the assessment. The objects used in the AM-ULA activities are not standardized.

In addition to these tests, some works in the literature use their own setups for assessment. In [48], tasks such as using a hammer and nail, stirring a bowl, folding a bath towel, and using a key in a lock are proposed for evaluating an upper-limb prosthesis. In [49], the performance of a neuroprosthesis is evaluated by asking the patient to perform grasping and lifting tasks as well as phone dialing, pouring liquid from a pitcher, and using a spoon and fork. In [50], to evaluate the outcomes of a protocol for stoke rehabilitation, blocks, Lego bricks, and pegs are used together with daily life activities like folding, buttoning, pouring, and lifting. In [51], the outcomes of a neuroprosthesis are measured with the box-and-blocks test and clothes-pin relocation task together with the evaluation of actions of daily living, i.e., using a fork and a knife, opening a jar, and stirring a spoon in a bowl. But none of the above-mentioned assessment procedures provides descriptions of the objects used.

Access to the objects is crucial to performance benchmarking as many aspects of the manipulation process cannot be modeled.

In our object set, we have included objects that are commonly used in these assessment procedures (i.e., a mug, a bowl, a pitcher, washers, bolts, kitchen items, pens, a padlock, and so on). We also included objects that will allow designing protocols that focus on activities of daily living. Moreover, widely used manipulation tests such as the nine-hole peg and box-and-blocks tests are also provided.

Object and Data Set

The contents of the proposed object set are shown in Figures 1–8 and listed in Table 2. The objects in the set are divided into the following categories: 1) food items, 2) kitchen items, 3) tool items, 4) shape items, and 5) task items tests are also provided.

Object Choices

We aimed to choose objects that are frequently used in daily life and also went through the literature to take into account objects that are frequently used in simulations and

Figure 1. The food items in the YCB object set. Back row, from left: a can of chips, a coffee can, a cracker box, a box of sugar, and a can of tomato soup. Middle row, from left: a container of mustard, a can of tuna fish, a box of chocolate pudding, a box of gelatin, and a can of potted meat. Front row: plastic fruits (a lemon, an apple, a pear, an orange, a banana, a peach, strawberries, and a plum).

Figure 2. The kitchen items in the YCB object set. Back row, from left: a pitcher, a container of bleach cleanser, and a container of glass cleaner. Middle row, from left: a plastic wine glass, an enamel-coated metal bowl, a metal mug, and an abrasive sponge. Front row, from left: a cooking skillet with a glass lid, a metal plate, eating utensils (knife, spoon, and fork), a spatula, and a white table cloth.

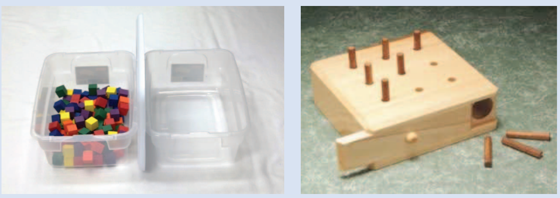
experiments. We also benefit from the studies on objects of daily living [52] and daily activities checklists such as [53].

In compiling the proposed object and task set, we needed to take a number of additional practical issues into consideration.

- *Variety*: To cover as many aspects of robotic manipulation as possible, we included objects that have a wide variety of shapes, sizes, transparencies, deformabilities, and textures. Considering size, the necessary grasp aperture varies from 14 cm (the diameter of the soccer ball) to 0.64 cm (the diameter of the smallest washer). Considering deformability, we have rigid objects together with foam bricks, a sponge, deformable balls, and articulated objects. Regarding transparency, we have included a transparent plastic wine glass, a glass skillet lid, and a semitransparent bottle of glass cleaner. The set includes objects with uniform plain textures, such as the pitcher and the stacking cups, and objects with irregular textures, like most of the groceries. Grasping and manipulation difficulty was also a criterion: for instance, some objects in the set are well approximated by simple geometric shapes (e.g., the box-shaped objects in the food category or the balls in the shape category) and relatively easy

Figure 3. The tool items in the YCB object set. Back row, from left: a power drill and wood block. Middle row, from left: scissors, a padlock and keys, markers (two sizes), an adjustable wrench, Phillips- and flat-head screwdrivers, wood screws, nails (two sizes), plastic bolts and nuts, and a hammer. Front row: spring clamps (four sizes).

Figure 4. The shape items in the YCB object set. Back row, from left: a mini soccer ball, a softball, a baseball, a tennis ball, a racquetball, and a golf ball. Front row, from left: a plastic chain, washers (seven sizes), a foam brick, dice, marbles, a rope, stacking cups (set of ten), and a blank credit card.



(a) (b)

Figure 5. (a) The improvised box-and-blocks test objects: a set of 100 wooden cubes, two containers, and a height obstacle (container lid) between them. (b) The nine-hole peg test: wooden pegs are placed in holes and stored in the body of the box.

(a) (b)

Figure 6. The assembly object: (a) the toy airplane disassembled, including a toy power screwdriver, and (b) the fully assembled airplane.

Figure 7. The assembly object: Lego Duplo pieces.

(a) (b)

Figure 8. The task items: (a) a black T-shirt and (b) a timer for accurate timing and as a manipulation object with a keypad.

Table 2. Object set items and properties.

Identification Number	Class	Object	Mass (g)	Dimensions (mm)
1	Food items	Chips can	205	75 × 250
2		Master chef can	414	102 × 139
3		Cracker box	411	60 × 158 × 210
4		Sugar box	514	38 × 89 × 175
5		Tomato soup can	349	66 × 101
6		Mustard bottle	603	58 × 95 × 190
7		Tuna fish can	171	85 × 33
8		Pudding box	187	35 × 110 × 89
9		Gelatin box	97	28 × 85 × 73
10		Potted meat can	370	50 × 97 × 82
11		Banana	66	36 × 190
12		Strawberry	18	43.8 × 55
13		Apple	68	75
14		Lemon	29	54 × 68
15		Peach	33g	59
16		Pear	49	66.2 × 100
17		Orange	47	73
18		Plum	25g	52
19	Kitchen items	Pitcher base	178	108 × 235
20		Pitcher lid	66	123 × 48
21		Bleach cleanser	1,131	250 × 98 × 65
22		Windex bottle	1,022	80 × 105 × 270
23		Winelass	133	89 × 137
24		Bowl	147	159 × 53
25		Mug	118	80 × 82
26		Sponge	6.2	72 × 114 × 14
27		Skillet	950	270 × 25 × 30
28		Skillet lid	652	270 × 10 × 22
29		Plate	279	258 × 24
30		Fork	34	14 × 20 × 198
31		Spoon	30	14 × 20 × 195
32		Knife	31	14 × 20 × 215
33		Spatula	51.5	35 × 83 × 350
34		Table cloth	1,315	2,286 × 3,352
35	Tool items	Power drill	895	35 × 46 × 184
36		Wood block	729	85 × 85 × 200
37		Scissors	82	87 × 200 × 14
38		Padlock	304	24 × 47 × 65
39		Keys	10.1	23 × 43 × 2.2
40		Large marker	15.8	18 × 121
41		Small marker	8.2	8 × 135
42		Adjustable wrench	252	5 × 55 × 205
43	Tool items	Phillips screwdriver	97	31 × 215
44		Flat screwdriver	98.4	31 × 215
45		Nails	[2, 2.7, 4.8]	[4 × 25, 3 × 53, 4 × 63]
46		Plastic bolt	3.6	43 × 15
47		Plastic nut	1	15 × 8
48		Hammer	665	24 × 32 × 135
49		Small clamp	19.2	85 × 65 × 10
50		Medium clamp	59	90 × 115 × 27
51		Large clamp	125	125 × 165 × 32
52		Extra-large clamp	202	165 × 213 × 37
53	Shape items	Mini soccer ball	123	140
54		Softball	191	96
55		Baseball	148	75
56		Tennis ball	58	64.7
57		Racquetball	41	55.3
58		Golf ball	46	42.7
59		Chain	98	1,149
60		Washers	[0.1, 0.7, 1.1, 3, 5.3, 19, 48]	[6.4, 10, 13.3, 18.8, 25.4, 37.3, 51]
61		Foam brick	28	50 × 75 × 50
62		Dice	5.2	16.2
63		Marbles	N/A	N/A
64		Rope	18.3	3,000 × 4.7
65		Cups	[13, 14, 17, 19, 21, 26, 28, 31, 35, 38]	[55 × 60, 60 × 62, 65 × 64, 70 × 66, 75 × 68, 80 × 70, 85 × 72, 90 × 74, 95 × 76, 100 × 78]
66		Blank credit card	5.2	54 × 85 × 1
67		Rope	81	3,000
68	Task items	Clear box	302	292 × 429 × 149
69		Box lid	159	292 × 429 × 20
70		Colored wood blocks	10.8	26
71		Nine-hole peg test	1,435	1,150 × 1,200 × 1,200
72		Toy airplane	570	171 × 266 × 280
73		Lego Duplo	523	N/A
74		T-shirt	105	736 × 736
75		Magazine	73	265 × 200 × 1.6
76		Timer	102	85 × 80 × 40
77		Rubik's Cube	94	57 × 57 × 57

for grasp synthesis and execution, while other objects have higher shape complexity (e.g., the spring clamps in the tool category, or the spatula in the kitchen-items category) and are more challenging for grasp synthesis and execution. Considering these aspects, the proposed set has a superior variety compared with the commercially available sets [8], [11], [41], [42], [44], which are designed to address some particular manipulation aspects only.

- *Use*: We included objects that are not only interesting for grasping but that also have a range of manipulation uses. For example, a pitcher and a cup; nails and a hammer; and pegs, cloths, and rope. We also included assembly items/tasks: a set of children's stacking cups, a toy airplane (Figure 6) that must be assembled and screwed together, and Lego Duplo bricks (Figure 7). In addition, widely used standard manipulation tests in rehabilitation, such as an improvised box-and-blocks [41] and a nine-hole peg test [42], are included. As mentioned above, these tasks are intended to span a wide range of difficulty, from relatively easy to very difficult. Furthermore, the ability to quantify the task performance was also prioritized, including aspects such as level of difficulty, time to completion, and success rate, among others.

Our goal is to do as much as possible to facilitate the widespread usage of a common set of objects and tasks to allow easy comparison of results between research groups worldwide.

Table 3. The suggestions for manipulation tasks.

Object Category	Suggested Tasks
Food items	• Packing/unpacking the groceries
Kitchen items	• Table setting • Wipe down table with sponge and Windex • Cooking scenarios
Tool items	• Nailing • Drilling • Unlocking the padlock using the key • Placing the pegs on the rope • Unscrewing a bolt using the wrench • Cutting a paper with the scissors • Writing on a paper • Screwing the nut on the bolt
Shape items	• Sorting marbles into the plastic blocks • Unstacking/stacking the cups • Placing the washers onto the bolt
Task items	• Box-and-blocks test • Toy-plane assembly/disassembly • Nine-hole peg tests • Lego assembly/disassembly • Cloth folding

- *Durability*: We aimed for objects that can be useful in the long term, and, therefore, avoid objects that are fragile or perishable. In addition, to increase the longevity of the object set, we chose objects that are likely to remain in circulation and change relatively little in the near future.
- *Cost*: We aimed to keep the cost of the object set as low as possible to broaden accessibility. We, therefore, selected standard consumer products, rather than, for instance, custom-fabricated objects, and tests. The current cost for the objects is approximately US$350.
- *Portability*: We aimed to have an object set that fits in a large-sized suitcase and be below the normal airline weight limit (22 kg) to allow easy shipping and storage.

After these considerations, the final objects were selected (Table 2 and Figures 1–8). Objects 1–18 are the food items, including real boxed and canned items as well as plastic fruits, which have complex shapes. Objects 19–34 are kitchen items, including objects for food preparation and serving as well as glass cleaner and a sponge. Objects 35–52 form the tool items category, containing not only common tools but also items—such as nails, screws, and wood—with which to utilize the tools. The shape items are objects 53–67, which span a range of sizes (spheres, cups, and washers), as well as compliant objects such as foam bricks, rope, and chain. The task items are objects 68–77, and they include two widely used tasks in rehabilitation benchmarking (box-and-blocks [41] and nine-hole peg test [42]) as well as items for relatively simple and complex assembly tasks (a Lego Duplo set and children's airplane toy, respectively). Furthermore, the set includes a black T-shirt for tasks like cloth folding as well as a magazine and a Rubik's cube. We include a timer in the kit (Figure 8), which not only provides accurate timing of the task but also serves as a manipulation object with a keypad. While there are an unlimited number of manipulation tasks that might be able to be done with these objects, we provide some examples for each category in Table 3 (with an in-depth discussion of tasks and protocols in the "Conclusions and Future Work" section).

Object Scans

To ease adoption across various manipulation research approaches, we collected visual data that are commonly required for grasping algorithms and generate 3-D models for use in simulation. We used the same scanning rig used to collect the BigBIRD data set [1]. The rig, shown in Figure 9, has five RGB-D sensors and five high-resolution RGB cameras arranged in a quarter-circular arc. Each object was placed on a computer-controlled turntable, which was rotated by 3° at a time, yielding 120 turntable orientations. Together, this yields 600 RGB-D images and 600 high-resolution RGB images. The process is completely automated, and the total collection time for each object is under 5 min.

We then used Poisson surface reconstruction to generate watertight meshes [54] (Figure 10). Afterward, we projected the meshes onto each image to generate segmentation masks. Note that Poisson reconstruction fails on certain objects with missing depth data; specifically, transparent or reflective regions of objects usually do not register depth data. We will later provide better models for these objects using algorithms that take advantage of the high-resolution RGB images for building models.

In total, for each object, we provide the following:

- 600 RGB-D images
- 600 high-resolution RGB images
- segmentation masks for each image
- calibration information for each image
- texture-mapped 3-D mesh models.

The object scans can be found online at [55].

Models

Based on the scans of the objects, there are several ways in which object models can be easily integrated into a variety of robot simulation packages. For example, in the MoveIt [5] simulation package, the mesh can be used as a collision object directly. Furthermore, a unified robot description format (URDF) file can be automatically constructed to integrate with ROS [56]. This provides a way to specify mass properties and can link to alternate representations of the mesh for visualization and collision. Integration with the OpenRAVE [57] simulation package is similarly straightforward where we link to the display and collision meshes from a KinBody XML file. Using the scans, we have created URDF and KinBody files for all of the objects in the data set, provided alongside the scans at [55].

Once in a simulation environment, a variety of motion planners and optimizers can use these models either as collision or manipulation objects. Some algorithms, such as Covariant Hamiltonian Optimization for Motion Planning [58], require signed-distance fields to avoid collisions, which can be computed from the included watertight meshes. Other cases, such as Constrained Bi-directional Rapidly-Exploring Random Tree [59], compute collisions directly using an optimized mesh collision checker.

In many cases, collision checking is a computational bottleneck for motion planning. Execution time can be reduced using a simplified mesh produced either by hand or with automatic decimation methods [60]. We have not yet provided simplified meshes in this data set, but we view this as an opportunity in future work to further explore mesh approximation algorithms and their impact on motion-planning problems using the standardized benchmarks.

Functional Demonstration of Integration into Simulation Software

The entire pipeline is shown in Figure 11. Here, we see the HERB robot [61] preparing to grasp the virtual drill object. This demonstration uses an integration of ROS and OpenRAVE. The ROS is used to provide communication between the various hardware and software components of the robot, while OpenRave handles planning and collision checking.

A variety of motion planners and optimizers can use these models either as collision or manipulation objects.

Inside OpenRAVE, the HERB robot uses CBiRRT, the Open Motion Planning Library [62] library, and CHOMP to plan and optimize motion trajectories. Using these tools, chains of several actions can be executed in sequence. The simulation environment also

Figure 9. The BigBIRD object-scanning rig: the box contains a computer-controlled turntable.

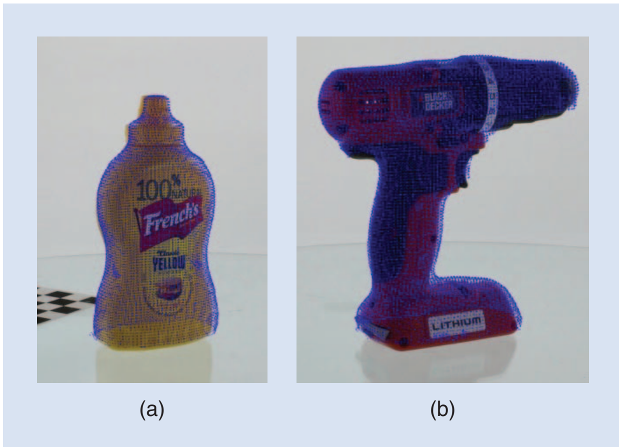


Figure 10. The point-cloud and textural-data overlays on two YCB objects: (a) the mustard bottle and (b) the power drill.

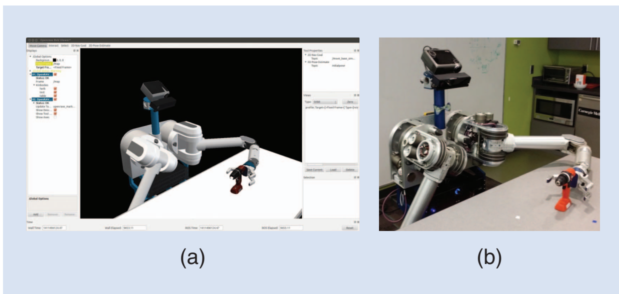


Figure 11. (a) The screen-capture from the OpenRAVE simulation and planning environment showing the HERB robot [34] planning a grasp of the power drill object in the set. (b) The actual grasp being executed by the robot on the physical object.

provides a mechanism for incorporating feedback from perception systems, which similarly benefit from this data set. The provided images, meshes, and physical objects can all be used as training data for various object-detection and pose-estimation algorithms, which can then be incorporated into the manipulation pipeline.

Access to both the physical object and a corresponding model for simulation is important for developing and testing new planning and manipulation algorithms. This data set vastly reduced the time required to set up this example by providing access to object models and meshes that have already been prepared for this purpose. This has removed the burden of scanning or modeling new objects and provides benchmark environments that streamline experimental design.

The objects in the set are designed to cover a wide range of aspects of the manipulation problem.

Protocol and Benchmark Template for Manipulation Research

Manipulation Protocol Template

Reference number/ version		
Authors		
Institution		
Contact information		
Purpose		
Task description		
Setup description	Description of the manipulation environment:	
	List of objects and their descriptions:	
	Initial poses of the objects:	
Robot/hardware/ subject description	Targeted robots/hardware/ subjects:	
	Initial state of the robot/ hardware/subject with respect to the setup:	
	Prior information provided to the robot/hardware/subject:	
Procedure		
Execution constraints		

Manipulation Benchmark Template

Reference number/ version	
Authors	
Institution	
Contact information	
Adopted protocol	
Scoring	
Details of setup	
Results to submit	

Protocol Design for Manipulation

A standard set of objects and associated models is a great starting point for common replicable research and benchmarking in manipulation, but there must be a sufficient amount of specification about what should be done with the objects to directly compare approaches and results. Given the wide range of technical interests, research approaches and applications being examined in the manipulation research community, along with how quickly the field moves, we cannot possibly provide sufficient task descriptions that will span the range of interests and remain relevant in the long term. Instead, we seek to lay the groundwork for those to be driven by the research community and subcommunities. We, therefore, focus on two efforts: developing a framework for task protocols, setting, formatting, and content guidelines to facilitate effective community-driven specification of standard tasks; and a preliminary set of example protocols that we believe are relevant for our respective communities and approaches, along with experimental implementation of those, including reporting the performance outcomes.

To enable effective community-driven evolution of protocols and benchmarks, the web portal associated with this effort [63] will serve as a jumping-off point. Protocols proposed by the community will be hosted at this portal, allowing them to be easily posted, shared, and cited, as well as easily updated as researchers give feedback and identify shortcomings. The portal will provide a forum for discussions on individual protocols and will provide links to matured protocols that meet the standards laid out in the template.

Protocol Guidelines

While developing protocols and benchmarks, one challenging aspect is to decide on the level of detail. Providing only high-level descriptions of the experiment (in other words, setting too few constraints) makes the repeatability of a benchmark, as well as its ability to assess the performance, questionable. Variations caused by incomplete descriptions of test setups and execution processes induce discrepancy in measurements and would not speak to some quantifiable performance. On the other hand, supplying too many constraints may limit a protocol's applicability and, therefore, narrow down its scope. For example, due to the variety of utilized hardware by different research groups in the robotics field, satisfying constrained hardware descriptions is not usually possible or preferred.

The aim of this section is to provide guidelines that help to maintain both reliable and widely applicable benchmarks for manipulation. For this purpose, five categories of information are introduced for defining manipulation protocols, i.e., 1) task description, 2) setup description, 3) robot/

hardware/subject description, 4) procedure, and 5) execution constraints. These categories are explained below, and, for the template, see "Protocol and Benchmark Template for Manipulation Research."

- *Task Description*: The task description is the highest level of information about the protocol. It describes the main action(s) of a task and (most of the time implicitly) the expected outcome(s). In this level, no constraints are given on the setup layout or how the task should be executed. Some task description examples are pouring liquid from a pitcher to a glass, hammering a nail on a piece of wood, or grasping an apple.
- *Setup Description*: This category provides the list of objects used in the manipulation experiment and their initial poses with respect to each other. In addition, if there are any other objects used as obstacles or clutter in the manipulation scenario, their description and layout will be described. As discussed above, the usage of nonstandard objects introduces uncertainty to many manipulation experiments presented in the literature. We believe that removing uncertainties in this category of information is crucial to maintain well-defined benchmarks. Providing the YCB object and model set is a step toward that purpose. In addition, in the protocols proposed in this article, the initial poses of the objects are accurately provided. Naturally, a task description can have various setup descriptions designed to assess the manipulation performance in different conditions.
- *Robot/Hardware/Subject Description*: This category provides information about the task executor. If the protocol is designed for a robotic system, the initial state of the robot with respect to the target object(s) and a priori information provided to the robot about the manipulation operation (e.g., the semantic information about the task, whether or not object shape models are provided.) are specified in this category. In addition, if the protocol is designed for a specific hardware setup (including sensory suite), the description is given. If the task executor is a human subject, how the subject is positioned with respect to the manipulation setup and a priori information given to the subject about the task at hand are described here.
- *Procedure*: In this category, actions that are needed to be taken by the person who conducts the experiment are explained step by step.
- *Execution Constraints*: In this category, the constraints on how to execute the task are provided. For instance in the box-and-blocks test, the subject is expected to use his/her dominant hand and needs to transfer one block at a time, or, if the task is to fetch a mug, the robot may be required to grasp the mug from its handle. In "Protocol and Benchmark Template for Manipulation Research," we provide a template for easily designing manipulation protocols using the aforementioned categories.

The proposed template and categories have several advantages as follows.

- The categorization helps researchers think about the protocol design in a structured way.
- It separates high-level task description from setup and robot/hardware/subject description so that protocols can be designed for analyzing different scenarios of the same manipulation problem.

Furthermore, describing setup and robot/hardware/subject separately allows platform-independent benchmark designs. Especially in the robotics field, the researchers usually have limited access to hardware. The designer may prefer to impose few constraints on the robot/hardware/subject description category to increase the applicability of the protocol. The amount and specifics of the detail in a given protocol will naturally vary based on the particular problem being examined, and therefore the insight of the authors about the intended application will be crucial in crafting an effective set of task descriptions and constraints. Related to this point, we anticipate protocols to be regularly improved and updated with feedback from the research community.

Benchmark Guidelines

After the task description, the second major part of each protocol is the specification of the associated benchmark, which details the metrics for scoring performance for the given protocol. Benchmarks allow the researchers to specify the performance of their system or approach and enable direct comparison with other approaches. The following categories of information are introduced for defining manipulation benchmarks.

- *Adopted protocol*: A well-defined description can be obtained for a manipulation benchmark by adopting a protocol that is defined considering the above-mentioned aspects.
- *Scoring*: Providing descriptive assessment measures is crucial for a benchmark. The output of the benchmark should give reasonable insight of the performance of a system. While designing the scoring criteria, it is usually a good practice to avoid binary (success/fail) measures; if possible, the scoring should include the intermediate steps of the task, giving partial points for a reasonable partial execution.
- *Details of setup*: In this field, the user gives detailed information about setup description that is not specified by the protocol. This could include the robot type, gripper type, grasping strategy, motion-planning algorithm, grasp synthesis algorithm, and so on.
- *Results to submit*: This field specifies the results and scores that need to be submitted by the user. Moreover, asking the user to submit the detailed reasoning for the failed attempts and the factors that bring success would help researchers who analyze the results. Therefore, having explicit fields for result analysis would be a good practice (see example benchmarks in [64]).

YCB Protocols and Benchmarks

While this protocol structure definition (and the template provided in "Protocol and Benchmark Template for Manipulation Research") helps to guide the development of effective task specification for various manipulation benchmarks, we have developed a number of example protocols to both provide more concrete samples of the types of task definitions that can be put forward as well as specific and useful benchmarks for actually quantifying performance. We have defined five protocols to date:

- pitcher–mug protocol
- gripper-assessment protocol
- table-setting protocol
- block pick-and-place protocol
- peg-insertion learning-assessment protocol.

From each protocol, a benchmark of reported performance is derived with the same name. We have implemented each of the protocols experimentally and reported the benchmark performance of our implementations for each. All these protocols and benchmarks and the results discussed in this section can be found at [64]. We have also implemented the box-and-blocks test for maintaining a baseline performance of this test for robotic manipulation.

A URDF file that spawns the scenario for Gazebo simulation environment is given.

YCB Pitcher–Mug Protocol and Benchmark

One of the popular tasks among robotics researchers is pouring a liquid from a container. This task is interesting as it necessitates semantic interpretation and smooth and precise manipulation of the target object. A protocol is designed for executing this manipulation task. The protocol uses the pitcher and the mug of YCB object and model set and provides scenarios by specifying ten initial configurations of the pitcher and the mug. By standardizing the objects and providing detailed initial state information, it aims at maintaining a common basis of comparison between different research groups. The benchmark derived from this protocol uses a scoring scheme that penalizes the amount of liquid that remains in the pitcher or spilled on the table. This benchmark was applied using the HERB robot platform [61], which can be seen in Figure 12. The reported results show that the task is successfully executed for eight of ten pitcher–mug configurations. For the two failed cases, the robot is able to grasp the pitcher but cannot generate a suitable path for pouring the liquid. This shows the importance of planning the manipulation task as a whole rather than in segments.

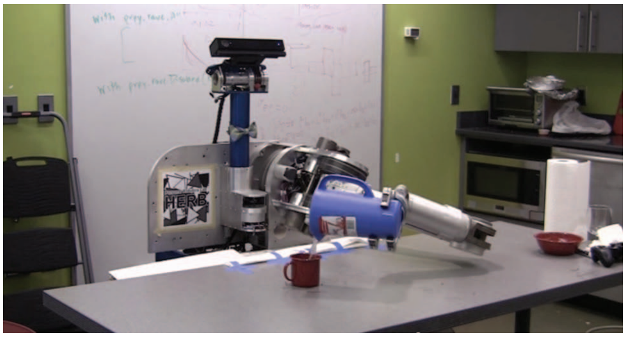


Figure 12. The HERB robot implementing the pitcher–mug benchmark.

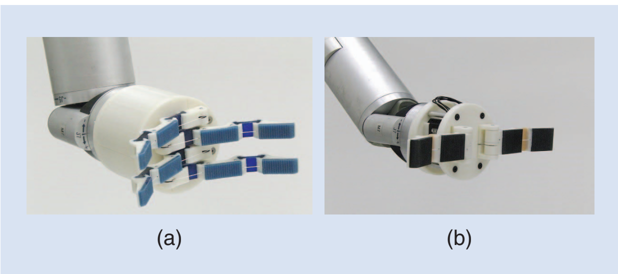


Figure 13. The grippers compared with gripper assessment benchmark: (a) Model T and (b) Model T42.

YCB Gripper-Assessment Protocol and Benchmark

The abilities of a robot's gripper affect its manipulation performance significantly. In the literature and in the commercial market, various gripper designs are available, each of which has different manipulation capabilities. The protocol defines a test procedure for assessing the performance of grippers for grasping objects of various shapes and sizes. This protocol utilizes objects from the shape and tool categories of the YCB object and model set. Using this protocol, a benchmark is defined based on a scoring table. We applied this benchmark to two grippers designed in Yale GRAB Lab, the Model T and Model T42 [65], which are shown in Figure 13. The results show that the Model T can provide successful grasp for only a limited range of object sizes. This gripper is not suitable for grasping small and flat objects. However, the ability to interlace its fingers increases the contact surface with the object and brings an advantage, especially for grasping concave and articulated objects. The Model T42 is able to provide stable power grasps for large objects and precision grasps for small objects. This model is also successful in grasping flat objects thanks to its nail-like fingertips. However, not being able to interlace its fingers brings a disadvantage while grasping articulated objects. Using the same benchmark for evaluating different gripper designs not only provided a basis of comparison but also gave many clues about how to improve the designs.

YCB Protocol and Benchmark for Table Setting

Pick-and-place is an essential ability for service robots. The benchmark assesses this ability by the daily task of table setting. The protocol uses the mug, fork, knife, spoon, bowl, and plate of the YCB object and model set. These objects are placed to predefined initial locations, and the robot is expected to replace them to specific final configurations. The benchmark scores the performance of the robot by the accuracy of the final object poses. This benchmark can also be applied in a simulation environment, since the models of the objects are provided by the YCB object and model set. A URDF file that spawns the scenario for Gazebo simulation environment is given at http://rll.eecs.berkeley.edu/ycb/. A snapshot of this setting can be seen in Figure 14.

YCB Block Pick-and-Place Protocol and Benchmark

Manual dexterity and the manipulation of small objects are critical skills for robots in several contexts. The block pick-and-place protocol is designed to test a robot's ability to grasp small objects and transfer them to a specified location. This task is an important test of both arm and gripper hardware and motion planning software, as both contribute to overall dexterity. Points are awarded based on completion and precision of the manipulation. We executed this test on the HERB robot [61], as seen in Figure 15. An image of the printed layout with the placed blocks after task completion can be seen in Figure 16. The results show that the robot is not able to succeed in precise pick-and-place task. The main reason is the utilized open-loop grasping approach. The robot executes a robust push grasp strategy, which allows it to grasp the blocks successfully. However, the pose of the block with respect to the gripper is not known precisely after the grasp. This prevents placing the blocks accurately to the target locations.

YCB Peg-Insertion Learning-Assessment Protocol and Benchmark

The peg-insertion learning-assessment benchmark is designed to allow comparison among various learning techniques. The benchmark measures the performance of a learned peg-insertion action under various positioning perturbations. The perturbations are applied by moving the peg board to a random direction for certain amount of distance. We applied this benchmark to assess the performance of a learned linear-Gaussian controller using a PR2 robot [66] (Figure 17). The state of the controller consists of the joint angles and angular velocities of the robot as well as the positions and velocities of three points in the space of the end effector (three points to fully define a rigid body configuration). No information is available to the controller at run time except for this state information. The results show that the learned controller shows reasonable performance, with four successes out of ten trials, for the case of 5-mm position perturbation to a random direction. This success rate can be achieved by executing the controller for only 1 s. However, the performance does not improve, even if the controller is run for a longer period of time. In the case of 10-mm position perturbation, the controller fails completely. We are planning to learn the same task with different learning techniques and compare their performances using the benchmark.

The Model T42 is able to provide stable power grasps for large objects and precision grasps for small objects.

Figure 14. The simulation environment for the table-setting benchmark. This environment can be spawned using the URDF provided at http://rll.eecs.berkeley.edu/ycb.

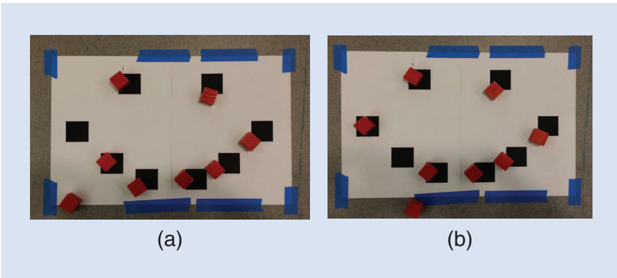


Figure 16. (a) and (b) The results of the block pick-and-place benchmark.

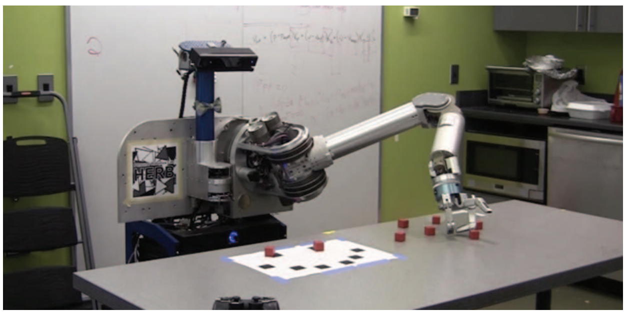


Figure 15. The HERB robot implementing the peg-insertion learning-assessment benchmark.

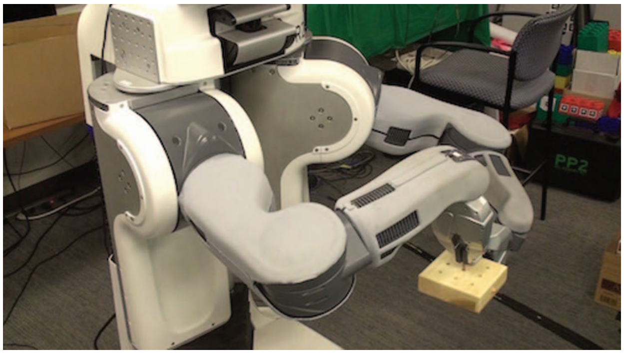


Figure 17. The PR2 executing the peg-insertion learning-assessment benchmark.

Box-and-Blocks Test

As mentioned in the "Related Work" section, the box-and-blocks test [41] is a widely used assessment technique that is utilized in prosthetics and rehabilitation fields. The test evaluates how many blocks can be grasped and moved from one side of the box (Figure 18) to the other in a fixed amount of time. We believe that the application of this test can also be quite useful for assessing the manipulation capabilities of robots. To establish a baseline performance for this test for robotic manipulators, we applied the box-and-blocks test with a PR2 robot (Figure 18) by implementing a very simple heuristic rules. The robot picks a location from a uniform distribution over the box and attempts to pick up a block. The gripper's pose aligns with the length of the box. The gripper is then closed and checked if it is fully closed. If the gripper closes fully, this means no blocks have been grasped and, therefore, the robot chooses a new location to attempt another pick. The robot repeats this heuristic until the gripper is not fully closed. When a grasp is detected, the robot moves to the destination box and releases the block. By using this heuristic, we run ten experiments of 2 min each and report the results at [64].

The OpenGRASP benchmarking suite [20] presents a simulation framework for robotic manipulation.

Conclusions and Future Work

This article proposes a set of objects and related tasks as well as high-resolution scans and models of those objects, intended to serve as a widely distributed and widely utilized set of standard objects to facilitate the implementation of standard performance benchmarks for robotic grasping and manipulation research. The objects were chosen based on an in-depth literature review of other object sets and tasks previously proposed and utilized in robotics research, with additional consideration to efforts in prosthetics and rehabilitation. Furthermore, a number of practical constraints were considered, including a reasonable total size and mass of the set for portability, low cost, durability, and the likelihood that the objects would remain mostly unchanged in years to come. High-resolution RGB-D scans of the object in the set were completed, and 3-D models have been constructed to allow easy portability into simulation and planning environments. All of these data are freely available in the associated repository [55]. Over the course of 2015, 50 objects sets will be freely distributed to a large number of research groups through workshops/tutorials associated with this effort. Additional object sets will be made available to purchase otherwise.

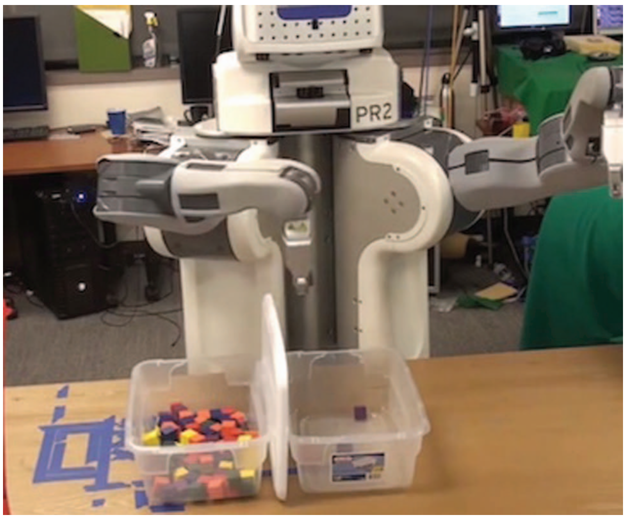


Figure 18. The PR2 executing the box-and-blocks test.

While a common set of widely available objects is a much-needed contribution to the manipulation research community, the objects themselves form only part of the contribution of the YCB set. The generation of appropriately detailed tasks and protocols involving the objects is ultimately what will allow for replicable research and performance comparison. We make inroads into that problem in this article by proposing a structure for protocols and benchmarks, implemented in a template as well as six example protocols. We hope that specification of protocols and benchmarks will become subcommunity driven and continually evolving. Specific aspects of manipulation and other specific research interests will naturally require different task particulars (i.e., specified and free parameters). We, therefore, plan to involve the research community in this effort via our web portal [63]. We will work toward having the majority of such protocols come from the user community rather than the authors of this article. In addition, we plan to have on this portal a records-keeping functionality to keep track of the current world records for the different tasks and protocols, along with video and detailed descriptions of the approaches utilized, generating excitement, buzz, motivation, and inspiration for the manipulation community to compare approaches and push forward the state of the art.

Other efforts that we plan to undertake include more detail about the objects proposed, including information about the inertia of the objects, as well as frictional properties between the objects and common surfaces. Additionally, we will expand our treatment of the modeling of the objects, including addressing the tradeoffs between number of triangles in a mesh and the reliable representation of the object geometry. Furthermore, before final publication and distribution of the object set, we will seek additional input from the research community on the specific objects in the set.

It is our hope that this article will help to address the longstanding need for common performance comparisons and benchmarks in the research community and will provide a starting point for further focused discussion and iterations on the topic.

Acknowledgment

The authors would like to thank Michael Leddy for his efforts in measuring the physical properties of the objects in the set. We would also like to thank Peter Allen and Jeff Trinkle for their feedback on the objects in the set and the overall approach taken in this article.

Funding for this work was provided in part by the U.S. National Science Foundation under grants IIS-0953856, IIS-1139078, and IIS-1317976.

References

[1] A. Singh, J. Sha, K. S. Narayan, T. Achim, and P. Abbeel, "BigBIRD: A large-scale 3D database of object instances," in *Proc. Int. Conf. Robotics Automation*, 2014, pp. 509–516.

[2] B. Li, Y. Lu, C. Li, A. Godil, T. Schreck, M. Aono, Q. Chen, N. K. Chowdhury, B. Fang, T. Furuya, H. Johan, R. Kosaka, H. Koyanagi, R. Ohbuchi, and A. Tatsuma. (2014). SHREC'14 track: Large scale comprehensive 3D shape retrieval. [Online]. Available: http://www.itl.nist.gov/iad/vug/sharp/contest/2014/Generic3D/

[3] C. Goldfeder, M. Ciocarlie, H. Dang, and P. K. Allen, "The Columbia grasp database," in *Proc. IEEE Int. Conf. Robotics Automation*, 2009, pp. 1710–1716.

[4] (2014). Sketchup. Available: [Online]. Available: http://www.sketchup.com/

[5] S. Chitta, I. Sucan, and S. Cousins, "MoveIt!," *IEEE Robot. Automat. Mag.*, vol. 19, pp. 18–19, Mar. 2012.

[6] Y. S. Choi, T. Deyle, T. Chen, J. D. Glass, and C. C. Kemp, "A list of household objects for robotic retrieval prioritized by people with ALS," in *Proc. IEEE Int. Conf. Rehabilitation Robotics*, 2009, pp. 510–517.

[7] Amazon Picking Challenge. [Online]. Available: http://amazonpickingchallenge.org/

[8] J. Adams, K. Hodges, J. Kujawa, and C. Metcalf, "Test-retest reliability of the Southampton hand assessment procedure," *Int. J. Rehab. Res.*, vol. 32, p. S18, Aug. 2009.

[9] S. Kalsi-Ryan, A. Curt, M. Verrier, and M. G. Fehlings, "Development of the graded redefined assessment of strength, sensibility and prehension (GRASSP): Reviewing measurement specific to the upper limb in tetraplegia," *J. Neurosurg Spine*, vol. 1, pp. 65–76, Sept. 2012.

[10] N. Yozbatiran, L. Der-Yeghiaian, and S. C. Cramer, "A standardized approach to performing the action research arm test," *Neurorehabil Neural Repair*, vol. 22, no. 1, pp. 78–90, 2008.

[11] E. D. Sears and K. C. Chung, "Validity and responsiveness of the Jebsen-Taylor hand function test," *J. Hand Surg. Amer.*, vol. 35, pp. 30–37, Jan. 2010.

[12] A. P. del Pobil, "Benchmarks in robotics research," in *Proc. IEEE Int. Conf. Robotics System*, 2006.

[13] R. Madhavan, R. Lakaemper, and T. Kalmar-Nagy, "Benchmarking and standardization of intelligent robotic systems," in *Proc. Int. Conf. Advanced Robotics*, 2009, pp. 1–7.

[14] I. Iossifidis, G. Lawitzky, S. Knoop, and R. Zöllner, "Towards benchmarking of domestic robotic assistants," in *Advances in Human-Robot Interaction*, vol. 14. Berlin, Heidelberg, Germany: Springer, 2005, pp. 403–414.

[15] R. Detry, O. Kroemer, and D. Kragic, "International workshop on autonomous grasping and manipulation: An open challenge," in *Proc. ICRA*, 2014.

[16] (2006). AIM@SHAPE Shape Repository v4.0. [Online]. Available: http://shapes.aim-atshape.net/viewmodels.php

[17] A. Richtsfeld. (2012). The object segmentation database (OSD). [Online]. Available: http://www.acin.tuwien.ac.at/?id=289

[18] P. Shilane, P. Min, M. Kazhdan, and T. Funkhouser, "The Princeton shape benchmark," in *Proc. Shape Modeling Applications*, 2004, pp. 167–178.

[19] 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," *Int. J. Robot. Res.*, vol. 31, no. 8, pp. 927–934, 2012.

[20] 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 *Proc. IEEE/RSJ Int. Conf. Intelligent Robots Systems*, 2011, pp. 1761–1767.

[21] B. Li, A. Godil, M. Aono, X. Bai, T. Furuya, L. Li, R. López-Sastre, H. Johan, R. Ohbuchi, C. Redondo-Cabrera, A. Tatsuma, T. Yanagimachi, and S. Zhang, (2012). SHREC'12 track: Generic 3D shape retrieval. [Online]. Available: http://www.itl.nist.gov/iad/vug/sharp/contest/2012/Generic3D/

[22] G. Kootstra, M. Popovic, J. A. Jorgensen, D. Kragic, H. G. Petersen, and N. Kruger, "VisGraB: A benchmark for vision-based grasping," *Paladyn*, vol. 3, no. 2, pp. 54–62, 2012.

[23] A. Tatsuma, H. Koyanagi, and M. Aono, "A large-scale Shape Benchmark for 3D object retrieval: Toyohashi shape benchmark," in *Proc Asia-Pacific Signal Information Processing Association Annu. Summit Conf.*, 2012, pp. 1–10.

[24] R. B. Rusu. (2011). NIST and willow garage: Solutions in perception challenge. [Online]. Available: http://www.nist.gov/el/isd/challenge-022511.cfm and http://www.willowgarage.com/blog/2011/02/28/nistand-willow-garage-solutions-perception-challenge

[25] H. Dutagaci, A. Godil, P. Daras, A. Axenopoulos, G. Litos, S. Manolopoulou, K. Goto, T. Yanagimachi, Y. Kurita, S. Kawamura, T. Furuya, and R. Ohbuchi. (2011). SHREC'11 Track: Generic Shape Retrieval. [Online]. Available: http://www.itl.nist.gov/iad/vug/sharp/contest/2011/NonRigid/

[26] A. Janoch, S. Karayev, Y. Jia, J. T. Barron, M. Fritz, K. Saenko, and T. Darrell, "A category-level 3-D object dataset: Putting the Kinect to work," in *Proc. IEEE Int. Conf. Computer Vision Workshops*, 2011, pp. 1168–1174.

[27] K. Lai, L. Bo, X. Ren, and D. Fox, "A large-scale hierarchical multi-view RGB-D object dataset," in *Proc. IEEE Int. Conf. Robotics Automation*, 2011, pp. 1817–1824.

[28] Z. Lian, A. Godil, T. Fabry, T. Furuya, J. Hermans, R. Ohbuchi, C. Shu, D. Smeets, P. Suetens, D. Vandermeulen, and S. Wuhrer. (2010). SHREC'10 track: Non-rigid 3D shape retrieval. [Online]. Available: http://tosca.cs.technion.ac.il/book/shrec_robustness2010.html

[29] R. Wessel, M. B. Ina, and R. Klein, "A 3D shape benchmark for retrieval and automatic classification of architectural data," in *Proc. Eurographics Workshop 3D Object Retrieval*, 2009, pp. 53–56.

[30] N. Iyer, S. Jayanti, and K. Ramani, "An engineering shape benchmark for 3D models," in *Proc. IDETC/CIE ASME Int. Design Engineering Tech. Conf. Computers Information Engineering Conf.*, 2005, pp. 501–509.

[31] R. Fang, A. Godil, X. Li, and A. Wagan, "A new shape benchmark for 3D object retrieval," in *Proc. 4th Int. Symp. Advances Visual Computing*, Berlin, Heidelberg, Germany, 2008, pp. 381–392.

[32] (2005). McGill 3D Shape Benchmark. [Online]. Available: http://www.cim.mcgill.ca/~shape/benchMark/

[33] N. Kapadia, V. Zivanovic, M. Verrier, and M. R. Popovic, "Toronto rehabilitation institute-hand function test: Assessment of gross motor function in individuals with spinal cord injury," *Top. Spinal Cord Inj. Rehabil.*, vol. 18, no. 2, pp. 167–186, 2012.

[34] R. W. Sumner and J. Popovic, "Deformation transfer for triangle meshes," *ACM Trans. Graphics*, vol. 3, no. 3, pp. 399–405, Aug. 2014.

[35] (2003). NTU 3D Model Benchmark. [Online]. Available: http://3d.csie.ntu.edu.tw/
[36] (2014). The ITI database. [Online]. Available: http://vcl.iti.gr/3dobject-retrieval/
[37] (2014). Model Bank Library. [Online]. Available: http://digimation.com/3d-libraries/model-bank-library/
[38] T. Wisspeintner, T. van der Zan, L. Iocchi, and S. Schiffer, "RoboCup@Home: Results in benchmarking domestic service robots," in *RoboCup 2009: Robot Soccer World Cup XIII*, vol. 5949. Berlin, Heidelberg, Germany: Springer, 2010, pp. 390–401.
[39] G. Kragten, A. C. Kool, and J. Herder, "Ability to hold grasped objects by underactuated hands: Performance prediction and experiments," in *Proc. IEEE Int. Conf. Robotics Automation*, 2009, pp. 2493–2498.
[40] G. A. Kragten, C. Meijneke, and J. L. Herder, "A proposal for benchmark tests for underactuated or compliant hands," *Mech. Sci.*, vol. 1, no. 1, pp. 13–18, 2010.
[41] V. Mathiowetz, G. Volland, N. Kashman, and K. Wever, "Adult norms for the box and blocks test of manual dexterity," *Amer. J. Occup. Ther.*, vol. 39, no. 6, pp. 386–391, 1985.
[42] V. Mathiowetz, K. Weber, N. Kashman, and G. Volland, "Adult norms for the nine hole peg test of finger dexterity," *Occup. Ther. J. Res.*, vol. 5, no. 1, pp. 24–38, 1985.
[43] Salia Rehab. (2015, Aug. 7). Action Research Arm Test (ARAT). [Online]. Available: http://saliarehab.com/
[44] (2015, Aug. 7). GRASSP. [Online]. Available: http://grassptest.com
[45] (2015, Aug. 7). Grasp and Release Test. [Online]. Available: http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=1053
[46] L. Resnik, L. Adams, M. Borgia, J. Delikat, R. Disla, C. Ebner, and L. S. Walters, "Development and evaluation of the activities measure for upper limb amputees," *Arch. Phy. Med. Rehab.*, vol. 94, no. 3, pp. 488–494, 2013.
[47] M. Popovic and C. Contway, "Rehabilitation engineering laboratory hand function test for functional electrical stimulation assisted grasping," in *Proc. 8th Annu. Conf. Int. Functional Electrical Stimulation Society*, 2003, pp. 231–234.
[48] H. Burger, F. Franchignoni, A. W. Heinemann, S. Kotnik, and A. Giordano, "Validation of the orthotics and prosthetics user survey upper extremity functional status module in people with unilateral upper limb amputation," *J. Rehab. Med.*, vol. 40, no. 5, pp. 393–399, 2008.
[49] M. R. Popovic, T. Keller, I. P. I. Papas, V. Dietz, and M. Morari, "Surface-stimulation technology for grasping and walking neuroprostheses," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 1, pp. 82–93, 2001.
[50] J. Harris, J. Eng, W. Miller, and A. S. Dawson, "A self-administered graded repetitive arm supplementary program (GRASP) improves arm function during inpatient stroke rehabilitation: A multi-site randomized controlled trial," *Stroke*, vol. 40, no. 6, pp. 2123–2131, 2009.
[51] K. A. Stubblefield, L. A. Milller, R. D. Lipschutz, M. E. Phillips, C. W. Heckathorne, and T. A. Kuiken, "Occupational therapy outcomes with targeted hyper-reinnervation nerve transfer surgery: Two case studies," in *Proc. MyoElectric Controls/Powered Prosthetics Symp.*, 2005, pp. 169–177.
[52] K. Matheus and A. M. Dollar, "Benchmarking grasping and manipulation: Properties of the objects of daily living," in *Proc. IEEE/RSJ Int. Conf.Intelligent Robots Systems*, 2010, pp. 5020–5027.
[53] (2015, Aug. 7). The Dash—"Disabilities of the Arm, Shoulder and Hand". [Online]. Available: http://dash.iwh.on.ca/system/files/dash_questionnaire_2010.pdf
[54] M. Kazhdan, M. Bolitho, and H. Hoppe, "Poisson Surface Reconstruction," presented at the Eurographics Symp. Geometry Processing, 2006.
[55] (2015, Aug. 7). [Online]. Available: http://rll.eecs.berkeley.edu/ycb/
[56] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, E. Berger, R. Wheeler, and A. Y. Ng, "ROS: An open-source robot operating system," in *Proc. ICRA Workshop Open Source Software*, 2009, vol. 3, p. 5.
[57] R. Diankov and J. Kuffner, "OpenRAVE: A planning architecture for autonomous robotics," Robotics Institute, Carnegie Mellon Univ., Pittsburgh, PA, Tech. Rep. CMU-RI-TR-08-34, 2008.
[58] N. Ratliff, M. Zucker, J. A. Bagnell, and S. Srinivasa, "CHOMP: Gradient optimization techniques for efficient motion planning," in *Proc. IEEE Int. Conf. Robotics Automation*, 2009, vols. 1–7, pp. 4030–4035.
[59] D. Berenson, S. S. Srinivasa, D. Ferguson, and J. J. Kuffner, "Manipulation planning on constraint manifolds," in *Proc. IEEE Int. Conf. Robotics Automation*, 2009, pp. 625–632.
[60] M. Garland and P. S. Heckbert, "Surface simplification using quadric error metrics," presented at the Proc. 24th Annu. Conf. Computer Graphics and Interactive Techniques, 1997, pp. 209–216.
[61] S. S. Srinivasa, D. Berenson, M. Cakmak, A. Collet, M. R. Dogar, A. D. Dragan, R. A. Knepper, T. D. Niemueller, K. Strabala, and J. M. Vandeweghe, "Herb 2.0: Lessons learned from developing a mobile manipulator for the home," *Proc. IEEE*, vol. 100, no. 8, pp. 2410–2428, 2012.
[62] I. A. Sucan, M. Moll, and L. E. Kavraki, "The open motion planning library," *IEEE Robot. Automat. Mag.*, vol. 19, pp. 72–82, Dec. 2012.
[63] (2015, Aug. 7). YCB Protocols and Benchmarks. [Online]. Available: http://www.ycbbenchmarks.org
[64] (2015, Aug. 7). YCB Protocols. [Online]. Available: http://protocols.ycb-benchmarks.org
[65] R. R. MA, L. U. Odhner, and A. M. Dollar, "A modular, open-source 3D printed underactuated hands," in *Proc. IEEE Int. Conf. Robotics Automation*, Karlsruhe, Germany, 2013, pp. 2737–2743.
[66] W. Garage. (2009). Overview of the PR2 robot. [Online]. Available: http://www.willowgarage.com/pages/pr2/overview

Berk Calli, Yale University, New Haven, Connecticut, USA. E-mail: berkcc@gmail.com.

Aaron Walsman, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA. E-mail awalsman@andrew.cmu.edu.

Arjun Singh, University of California, Berkeley, California, USA. E-mail: arjunsingh@berkeley.edu.

Siddhartha Srinivasa, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA. E-mail: siddh@cs.cmu.edu.

Pieter Abbeel, University of California, Berkeley, California, USA. E-mail: pabbeel@cs.berkeley.edu.

Aaron M. Dollar, Yale University, New Haven, Connecticut, USA. E-mail: aaron.dollar@yale.edu.