Benchmarking in Manipulation Research: The YCB Object and Model Set and Benchmarking Protocols

Berk Calli, *Member, IEEE*, Aaron Walsman, *Student Member,* Arjun Singh, *Student Member, IEEE*, *IEEE*, Siddhartha Srinivasa, *Senior Member, IEEE,* Pieter Abbeel, *Senior Member, IEEE,* and Aaron M. Dollar, *Senior Member, IEEE*

***Abstract*—In this paper we present the Yale-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, weight and rigidity, as well as some widely used manipulation tests. The associated database provides high-resolution RGBD 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 existing benchmarks and object datasets is also presented and their scope and limitations are discussed. The 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 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.**

I. Introduction

Benchmarks are crucial for the progress of a research field, allowing performance to be quantified in order to give insight into the effectiveness of an approach compared to 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 for which researchers are working towards. 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 such that others can repeat them. Moreover, the objects used may also be insufficiently specified and/or are typically not available to other researchers (e.g. they have been custom-fabricated or are only available for purchase in limited 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 efforts have been put into providing mesh model databases of objects, generally for object recognition or grasp planning purposes (e.g. [1-4] with a thorough overview provided in section II). There have, however, been very 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.

In this paper, 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. These were selected based on a survey of the most common objects utilized in the robotics field as well as 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 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 in order 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, weight, rigidity and texture, 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

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

B. Calli and A. Dollar are with the Department of Mechanical Engineering and Materials Science, Yale University, New Haven, CT, USA (203-436-9122; e-mail: {berk.calli; aaron.dollar}@yale.edu).

A. Singh and P. Abbeel are with the Department of Computer Science, University of California at Berkeley, Berkeley, CA, USA (e-mail: arjun810@gmail.com; pabbeel@cs.berkeley.edu).

A. Walsman and S. Srinivasa are with the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA (e-mail: aaronwalsman@gmail.com; ss5@andrew.cmu.edu).

protocol. To facilitate the design of well-defined future protocols and benchmarks, guidelines are provided through a template. Protocols and benchmarks are intended to generally be platform-independent in order 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 as well as for 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.

We expect to continually expand on this work, both with our own efforts (via more objects properties and additional benchmarks) but more importantly, via our web portal: http://www.ycbbenchmarks.org/. By this web portal, the user community can engage in this effort, proposing changes to the object set and putting forth their own protocols and benchmarks, etc.

The remainder of this paper is organized as follows: First a comprehensive literature survey on object sets and benchmarks is presented in Section II. Following that our object set is presented and explained in Section III. In Section IV, guidelines are provided for designing protocols and benchmarks. In Section V, the example protocols and benchmarks are presented. The paper is concluded with discussions and future work in Section VI.

II. RELATED WORK

For benchmarking in manipulation, specifying an object set is useful for the standardization of experimental conditions. Table 1 summarizes object sets that have been proposed for manipulation tasks in the fields of robotics, prosthetics and rehabilitation. Even though there have been a large number of efforts that provide datasets of object mesh models which are useful for many simulation and planning applications, as well as for benchmarking in shape retrieval, these datasets have limited utility for manipulation benchmarking due to several reasons: First, 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. Second, none of these databases provide the objects' physical properties, which are necessary to conduct realistic simulations. Lastly and perhaps 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), the recently-announced Amazon Picking Challenge [7], which provides a shopping list to purchase objects meant for a narrow bin-picking task. In prosthetics and rehabilitation field, commercial kits are available for upper limp assessment tests [8-11]. While demonstrating the benefits of utilizing a standard set for manipulation assessment, the scope of these kits are limited for benchmarking in robotics as they are not representative of a wide range of manipulation tasks. The current 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.

In the following sections, 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.

A. 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 (e.g. [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 (e.g. [2, 16, 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 (CGD) [3] rearranges the object models of the Princeton Shape Benchmark (PSB) [18] for robotic manipulation and provides mesh models of 8000 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 multi-purpose object set which also targets manipulation is the 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 datasets, 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.

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. 3D 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 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,

Table I: Object Datasets in Literature (Sorted by Year)

	Dataset 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	8987 / 171	No	http://www.itl.nist.gov/iad/vug/sharp/contest/2014/Generic3D/
4	SHREC'12 [21]	2012	Mesh models	Object retrieval	1200 / 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/ObjectModelsWebUI/
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 Dataset [26]	2011	RGB-D dataset of room scenes	Object detection	N/A	No	http://kinectdata.com/
12	RGB-D Object Dataset [27]	2011	RGB-D dataset	Object detection and recognition	300 / 51	No	http://rgbd-dataset.cs.washington.edu/
13	The OpenGRASP benchmarking suite [20]	2011	Mesh with texture, stereo images	Grasping	Uses KIT database	No	http://opengrasp.sourceforge.net/benchmarks.html
14	SHREC'10 [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	~8000	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	2257	No	ftp://ftp.cg.cs.uni-bonn.de/pub/outgoing/ArchitectureBenchmark
18	Engineering Shape Benchmark [30]	2008	Mesh models	Object retrieval	867	No	https://engineering.purdue.edu/PRECISE/shrec08
19	[3D Object Retrieval Benchmark] [31]	2008	Mesh models	Object retrieval	800 / 40	No	http://www.itl.nist.gov/iad/vug/sharp/benchmark/
20	McGill 3D Shape Benchmark [32]	2008	Mesh models	Shape retrieval	N/A	No	http://www.cim.mcgill.ca/~shape/benchMark/

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	1180	No	http://shapes.aim-at-shape.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 Dataset] [34]	2004	Mesh models	Mesh transformation	N/A / 13	No	http://people.csail.mit.edu/sumner/research/deftransfer/data.html
26	NTU 3D 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	1200	No	http://digimation.com/3d-libraries/model-bank-library/
32	SketchUp [4]	N/A	Mesh with and w/o texture	General	N/A	No	https://3dwarehouse.sketchup.com/
33	Robocup @home [38]	Multi.	no data	Manipulation	N/A	No	http://www.robocupathome.org/

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 utilizing real stereo images of the target objects for grasp synthesis, and executing and evaluating 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.

B. *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 published on, 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 (ADLs).

The tests that are commercially available are Box and Blocks Test [41], 9-hole-peg test [42], Jebsen-Taylor Hand Function Test [11], Action Research Arm Test (ARAT) [10], Graded Redefined Assessment of Strength, Sensibility and Prehension (GRASSP) test [9] and the Southampton Hand Assessment Procedure (SHAP) [8]. The setups for Box and Blocks and 9-hole-peg tests are very specific, with evaluation based on timed movements of simple objects. The setup for Jebsen-Taylor Hand Function Test includes objects for manipulation actions such as card turning, moving small (paper clips, bottle caps), light (empty cans) and heavy objects (1lb weighted cans), but utilizes a small number of objects of limited shape and size variety. 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 a glass to another. The GRASSP measure has also been proposed for the assessment of upper limb impairment. This measure is based on a commercial kit available in [44]. Apart from a specialized manipulation setup, the kit also includes 9-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 etc. Considering manipulation benchmarking in robotics, Box and Blocks, 9-hole-peg, 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 \$1300 and \$3000, respectively).

Some well-known tests that do not provide a commercial setup are Grasp and Release Test [45], The Toronto Rehabilitation Institute Hand Function Test [33] and Activities Measure for Upper Limp 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 these 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 Rehabilitation Engineering Laboratory Hand Function Test [47]) evaluates palmer (power) and lateral (precision) grasp abilities of individuals by using an object set consist of a mug, a book, a paper, a soda can, dice, a pencil etc. Even though it is claimed that the objects used in this test are easily obtainable, maintaining the exact object definition is hard and one of the objects is an outdated cellular phone. 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 AMULA activities are not standardized.

In addition to these tests, some works in literature use their own setups for assessment. In [48], tasks such as "use a hammer and nail", "stir a bowl", "fold a bath towel", "use a key in a lock" are proposed for evaluating upper limb prosthesis. In [49], the performance of the 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 spoon and fork. In [50], for evaluating the outcomes of a protocol for stoke rehabilitation, blocks, Lego and pegs are used together with daily life activities like folding, buttoning, pouring and lifting. In [51], the outcomes of the neuroprosthesis are measured with 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, stirring a spoon in a bowl. In none of the abovementioned assessment procedures, the descriptions of the objects are provided, however.

In our object set, we have included the objects that are commonly used in these assessment procedures (i.e. a mug, a bowl, a pitcher, washers, bolts, kitchen items, pens, key-padlock etc.). We also included objects that will allow designing protocols which focus on activities of daily living. Moreover, widely used manipulation tests such as 9-hole peg, box and blocks and clothes peg allocation are also provided.

III. The Object and Data Set

The contents of the proposed object set can be seen in Figures 1-8 and listed in Table II. The objects in the set are divided into the following categories: food items, kitchen items, tool items, shape items, task items. In this section, we describe the object set and the reasoning behind the choices (section III.A), a description of the process and data involved in the scans of the objects (III.B), the models and integration into simulation and planning packages (III.C) and a brief functional demonstration of the integration (III.D).

A. *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

Fig. 1: Food items in the YCB Object Set: back row: chips can, coffee can, cracker box, box of sugar, tomato soup can; middle row: mustard container, tuna fish can, chocolate pudding box, gelatin box, potted meat can; front: plastic fruits (lemon, apple, pear, orange, banana, peach, strawberries, plum).

Fig. 2: Kitchen items in the YCB Object Set: back row: pitcher, bleach cleanser, glass cleaner; middle row: plastic wine glass, enamel-coated metal bowl, metal mug, abrasive sponge; front: cooking skillet with glass lid, metal plate, eating utensils (knife, spoon, fork), spatula, white table cloth.

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

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

• *Variety:* In order to cover as many aspects of robotic manipulation as possible, we included objects that have a wide variety of shape, size, transparency, deformability, and texture. Considering size, the necessary grasp aperture varies from 14 cm (diameter of the soccer ball) to 0.64 cm (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

Fig. 3: Tool items in the YCB Object Set: back row: power drill, wood block; middle row: scissors, padlock and keys, markers (two sizes), adjustable wrench, phillips and flat screwdrivers, wood screws, nails (two sizes), plastic bolt and nut, hammer; front: spring clamps (four sizes).

Fig. 4: Shape items in the YCB Object Set: back row: Mini soccer ball, softball, baseball, tennis ball, racquetball, golf ball, front: plastic chain, washers (seven sizes), foam brick, dice, marbles, rope, stacking blocks (set of 10), credit card blank.

have included a transparent plastic wine glass, a glass skillet lid, and a semi-transparent glass cleaner bottle. 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 food items or balls in shape items) and relatively easy for grasp synthesis and execution, while other objects have higher shape complexity (e.g. spring clamps in tool items, or spatula in kitchen items) and more challenging for grasp synthesis and execution. Considering these aspects, the proposed set has a superior variety comparing to commercially available sets [8, 11, 41, 42, 44] which are designed to address some particular manipulation aspects only.

Table II: Object Set Items and Properties

ID	Class	Object	Mass	Dims. (mm)
1	Food items	Chips Can	205g	75 x 250
2		Master Chef Can	414g	102 x 139
3		Cracker Box	411g	60 x 158 x 210
4		Sugar Box	514g	38 x 89 x 175
5		Tomato Soup Can	349g	66 x 101
6		Mustard Bottle	603g	58 x 95 x 190
7		Tuna fish can	171g	85 x 33
8		Pudding Box	187g	35 x 110 x 89
9		Gelatin Box	97g	28 x 85 x 73
10		Potted Meat Can	370g	50 x 97 x 82
11		Banana	66g	36 x 190
12		Strawberry	18g	43.8 x 55
13		Apple	68g	75
14		Lemon	29g	54 x 68
15		Peach	33g	59
16		Pear	49g	66.2 x 100
17		Orange	47g	73
18		Plum	25g	52
19	Kitchen Items	Pitcher Base	178g	108 x 235
20		Pitcher Lid	66g	123 x 48
21		Bleach Cleanser	1131g	250 x 98 x 65
22		Windex Bottle	1022g	80 x 105 x 270
23		Wine glass	133g	89 x 137
24		Bowl	147g	159 x 53
25		Mug	118g	80 x 82
26		Sponge	6.2g	72 x 114 x 14
27		Skillet	950g	270 x 25 x 30
28		Skillet Lid	652g	270 x 10 x 22
29		Plate	279g	258 x 24
30		Fork	34g	14 x 20 x 198
31		Spoon	30g	14 x 20 x 195
32		Knife	31g	14 x 20 x 215
33		Spatula	51.5g	35 x 83 x 350
34		Table Cloth	1315	2286 x 3352
35	Tool Items	Power Drill	895g	35 x 46 x 184
36		Wood Block	729g	85 x 85 x 200
37		Scissors	82g	87 x 200 x 14
38		Padlock	304g	24 x 47 x 65
39		Keys	10.1g	23 x 43 x 2.2
40		Large Marker	15.8g	18 x 121
41		Small Marker	8.2g	8 x 135

Table II (cont): Object Set Items and Properties

ID	Class	Object	Mass	Dims. (mm)
42	Tool Items	Adjustable Wrench	252g	5 x 55 x 205
43		Phillips Screwdriver	97g	31 x 215
44		Flat Screwdriver	98.4g	31 x 215
45		Nails	[2,2.7,4.8] g	[4x25, 3x53, 4x63]
46		Plastic Bolt	3.6g	43 x 15
47		Plastic Nut	1g	15 x 8
48		Hammer	665g	24 x 32 x 135
49		S Clamp	19.2g	85 x 65 x 10
50		M Clamp	59g	90 x 115 x 27
51		L Clamp	125g	125 x 165 x 32
52		XL Clamp	202g	165 x 213 x 37
53	Shape Items	Mini Soccer Ball	123g	140
54		Soft Ball	191g	96
55		Baseball	148	75
56		Tennis Ball	58g	64.7
57		Racquetball	41g	55.3
58		Golf Ball	46g	42.7
59		Chain	98g	1149
60		Washers	[0.1,0.7,1.1,3,5.3,19,48] g	[6.4, 10, 13.3, 18.8, 25.4, 37.3, 51]
61		Foam Brick	28g	50 x 75 x 50
62		Dice	5.2g	16.2
63		Marbles	N/A	N/A
64		Rope	18.3g	3000 x 4.7
65		Cups	[13,14,17,19, 21,26,28,31,35,38] g	[55x60, 60x62, 65x64, 70x66, 75x68, 80x70, 85x72, 90x74, 95x76, 100x78]
66		Blank Credit Card	5.2g	54 x 85 x 1
67		Rope	81g	3000
68	Task Items	Clear Box	302g	292 x 429 x 149
69		Box Lid	159g	292 x 429 x 20
70		Colored Wood Blocks	10.8g	26
71		9-Peg-Hole Test	1435g	1150 x 1200 x 1200
72		Toy Airplane	570g	171 x 266 x 280
73		Lego Duplo	523g	N/A
74		T-shirt	105g	736 x 736
75		Magazine	73g	265 x 200 x 1.6
76		Timer	102g	85 x 80 x 40

- *Use:* We included objects that are not only interesting for grasping, but also have a range of manipulation uses. For example, a pitcher and a cup; nails and a hammer; pegs, cloths and rope. We also included “assembly” items/tasks: a set of children’s stacking cups, a toy airplane (Fig. 6) that must be assembled and screwed together and LEGO Duplo (Fig. 7). Additionally, widely used standard manipulation tests in rehabilitation, such as an improvised box and blocks [41] and a 9-hole-peg test [42] are included. As above, these tasks are intended to span a wide range of difficulty, from relatively easy to very difficult. Furthermore, the ability to quantify task performance was also prioritized, including aspects such as level of difficulty, time-to-completion, and success rate, among others.

- *Durability:* We aimed for objects that can be useful long term, and therefore avoid objects that are fragile or perishable. Also, to increase the longevity of the object set, we chose the objects that are likely to remain in circulation and change relatively little in the near future.

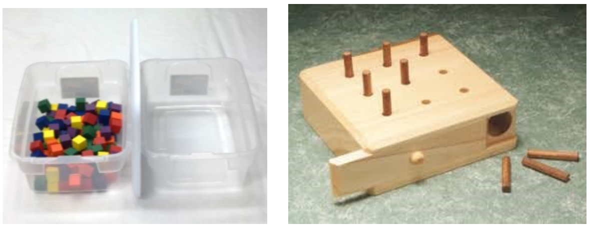


Fig. 5: (Left) Improvised box-and-blocks test objects: set of 100 wooden cubes, two containers and height obstacle (container lid) between them. (Right) 9-hole peg test: wooden pegs are placed in holes and stored in body.

Fig. 6: Assembly object: toy airplane disassembled (left), including toy power screwdriver, and fully assembled (right).

Fig. 7: Assembly object: Lego Duplo pieces.

Fig. 8: Task items: left: Black t-shirt, right: Timer for accurate timing and as a manipulation object with a keypad.

- *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. Current cost for the objects is approximately $350.

- *Portability:* We aimed to have an object set that fits in a large-sized suitcase and be below the normal airline weight limit (22kg) in order to allow easy shipping and storage.

After these considerations, the final objects were selected (Table II, Figs. 1-8). Objects from ID 1 to 18 are the *food items*, containing real boxed and canned items, as well as plastic fruits, which have complex shapes. The objects from ID 19 to 34 are *kitchen items*, containing objects for food preparation and serving, as well as glass cleaner and a sponge. The objects from 35 to 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 from ID 53 to 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 the objects with IDs 68 to 76, and include two widely used tasks in rehabilitation benchmarking (box-and-blocks [41] and 9-hole peg test [42]) as well as items for relatively simple and a 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 and a magazine. We also include a timer in the kit (Fig. 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 III (with in-depth discussion of tasks and protocols in Section IV).

B. *Object Scans*

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

We then used Poisson surface reconstruction to generate watertight meshes [54] (Fig. 10). Afterwards, 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:

- 600 RGBD images
- 600 high-resolution RGB images
- Segmentation masks for each image
- Calibration information for each image
- Texture-mapped 3D mesh models

The object scans can be found at [55].

Table III 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 pad 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, • 9-peg hole tests, • Lego assembly/disassembly. • Cloth folding.

C. 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 of specifying mass properties, and can link to alternate representations of the mesh for visualization and collision. Integration with the OpenRAVE [57] simulation package is similarly straight-forward 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 dataset, 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 CHOMP [58], require signed-distance fields to avoid collisions which can be computed from the included watertight meshes. In other cases such as CBiRRT [59] compute collisions directly using an optimized mesh collision checker.

Fig. 9: BigBIRD Object Scanning Rig: the box contains a computer-controlled turntable.

Fig. 10: Point cloud and textural data overlays on two YCB objects: mustard bottle and power drill.

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 dataset, but 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.

D. Functional Demonstration of Integration into Simulation Software:

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

Inside OpenRAVE, the HERB robot uses CBiRRT, the OMPL [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 provides a mechanism for incorporating feedback from perception systems, which similarly benefit

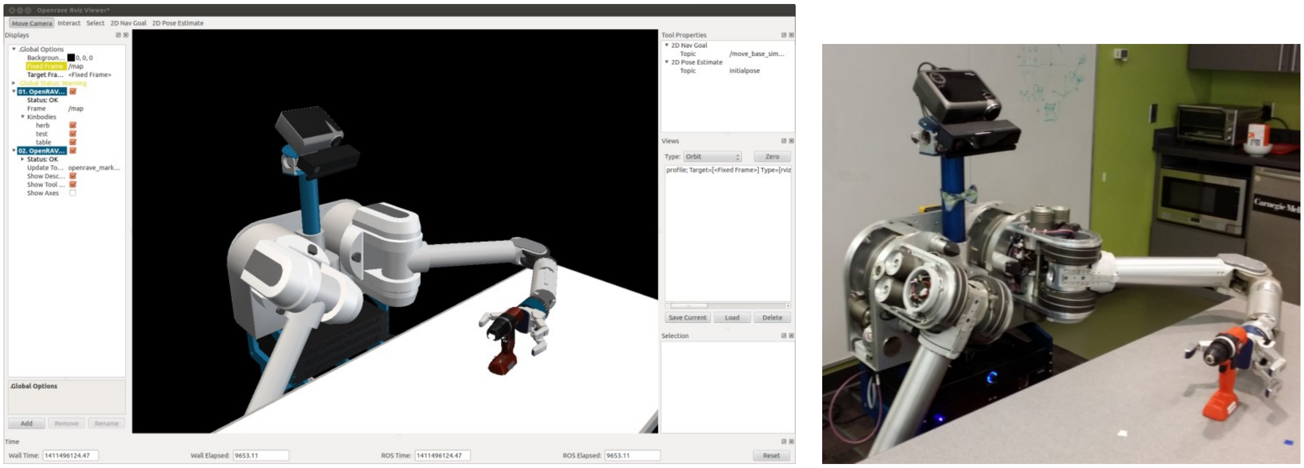


Fig. 11: (left) Screen-capture from Openrave simulation and planning environment showing the HERB robot [34] planning a grasp of the power drill object in the set. (right) actual grasp being executed by the robot on the physical object.

from this dataset. 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 dataset 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.

IV. Protocol Design For Manipulation

A standard set of objects and associated models are 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 in order 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 sub-communities. 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.

In order 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 identifies 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.

A. *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 won't speak to some quantifiable performance. On the other hand, supplying too many constraints may limit a protocol's applicability, and therefore narrows down its scope. For example, due to the variety of utilized hardware by different research groups in 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, namely (1) task description, (2) setup description, (3) robot/hardware/subject description, (4) procedure, and (5) execution constraints. These categories are explained below, and the protocol template provided in Appendix A:

1) *Task Description:* 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 wood," or "grasping an apple".

2) *Setup Description:* This category provides the list of objects used in the manipulation experiment and their initial poses with respect to each other. Also, if there are any other

objects used as obstacles or clutter in the manipulation scenario, their description and layout will be described. As discussed in the previous sections, the usage of non-standard objects introduces uncertainty to many manipulation experiments presented in 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 towards that purpose. Also, in the protocols proposed in this paper, 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.

3) 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 etc.) are specified in this category. Also, 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.

4) *Procedure:* In this category, actions that are needed to be taken by the person who conducts the experiment are explained step by step.

5) *Execution Constraints:* In this category, the constraints on how to execute the task are provided. For instance in 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 "fetching a mug", the robot may be required to grasp the mug from its handle. In Appendix A, we provide a template for easily designing manipulation protocols using the abovementioned categories.

The proposed template and categories have several advantages: First, the categorization helps researchers think about the protocol design in a structured way. Second, 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.

B. 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.

1) *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.

2) *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.

3) *Details of Setup:* In this field the user gives detailed information about setup description that is not specified by the protocol. These could be robot type, gripper type, grasping strategy, motion planning algorithm, grasp synthesis algorithm, etc.

4) *Results To Submit:* This filed specifies the results and scores that needs to be submitted by the user. Moreover, asking the user to submit 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]).

V. YCB Protocols and Benchmarks:

While this protocol structure definition (and template provided in Appendix A) 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.

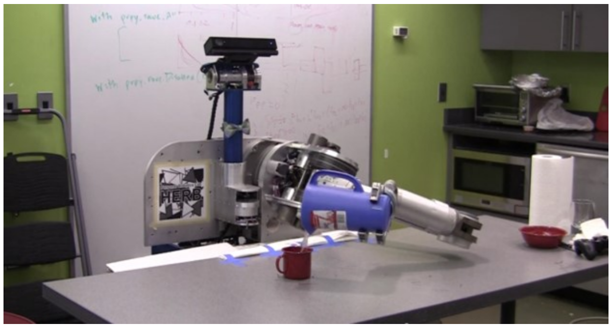
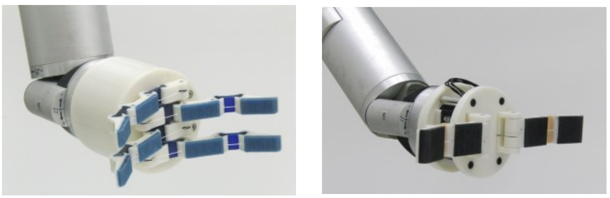


Fig. 12: HERB robot implementing Pitcher-Mug Benchmark



(a) (b)

Fig. 13: Grippers compared with Gripper Assessment Benchmark. (a) Model T, (b) Model T42.

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. Short descriptions for the protocol and benchmarks and summary of the benchmarking results are provided below.

A. 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 is aimed to maintain 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 Fig. 12. The reported results show that the task is successfully executed for 8 out of 10 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.

B. YCB Gripper Assessment Protocol and Benchmark:

The abilities of a robot's gripper affect its manipulation performance significantly. In literature and in commercial market, various gripper designs are available each of which have 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 can be seen in Fig. 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 object. 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 finger tips. However, not being able to interlace its fingers brings a disadvantage while grasping articulated objects. Using the same benchmark for evaluating different gripper designs did not only provide a basis of comparison, but also gave many clues about how to improve the designs.

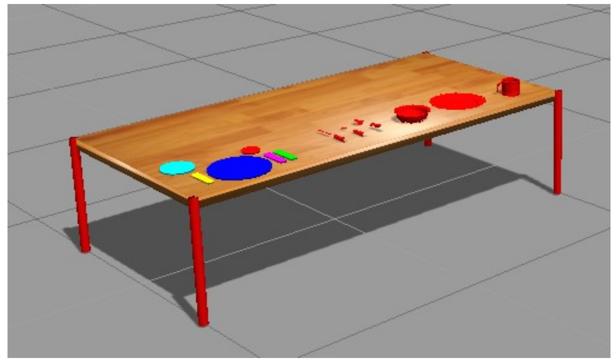


Fig. 14: The simulation environment for Table Setting Benchmark. This environment can be spawned by using the URDF provided at http://rll.eecs.berkeley.edu/ycb

C. 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 which 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 Fig. 14.

D. 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

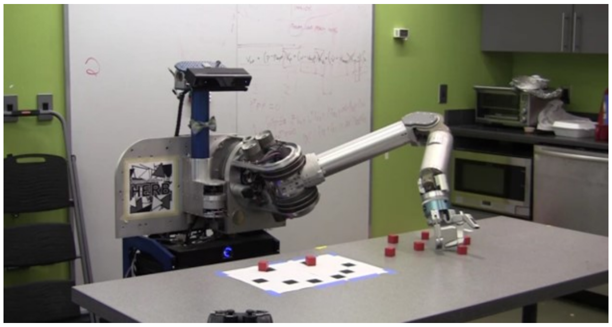


Fig. 15: HERB robot implementing Block Pick and Place Benchmark

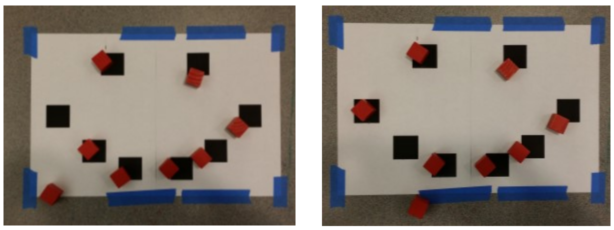


Fig. 16: The results of the Block Pick and Place Benchmark

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 Fig. 15. An image of the printed layout with the placed blocks after task completion can be seen in Fig. 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.

E. YCB Peg Insertion Learning Assessment Protocol and Benchmark:

The Peg Insertion Learning Assessment Benchmark is designed for allowing comparison between 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] (Fig. 17). The state of the controller consists of the joint angles and angular velocities of the robot, and

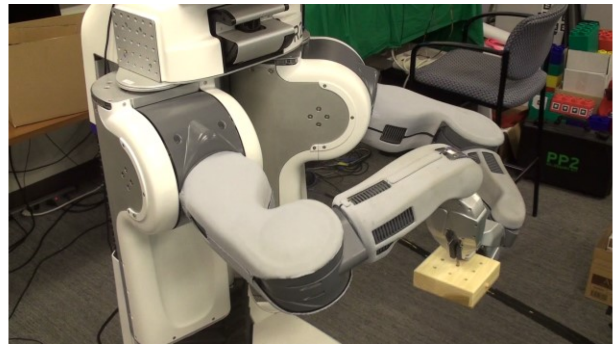


Fig. 17: PR2 executing the Peg Insertion Learning Assessment Benchmark.

the positions and velocities of three points in the space of the end effector (3 points in order 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, 4 success out of 10 trials, for the case of 5mm position perturbation to a random direction. This success rate can be achieved by executing the controller for only one second. However, the performance does not improve even if the controller is run for a longer period of time. In the case of 10mm 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.

F. Box and Blocks Test:

As mentioned previously in Section 2, 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 (Fig. 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. In order to establish a baseline performance for this test for robotic manipulators, we applied the Box and Blocks Test with a PR2 robot (Fig. 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 10 experiments of 2 minutes each, and report the results at [64]:

Fig. 18: PR2 executing the Box and Blocks Test.

VI. Conclusions and Future Work

This paper 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 RGBD scans of the object in the set were completed and 3D 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.

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 paper by proposing a structure for protocols and benchmarks, implemented in a template, as well as six example protocols. We aim that specification of protocols and benchmarks will become sub-community 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 towards having the majority of such protocols come from the user community rather than the authors of this paper. Additionally, 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 modelling 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 work will help to address the long-standing 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 paper.

References

[1] A. Singh, J. Sha, K. S. Narayan, T. Achim, and P. Abbeel, "BigBIRD: A Large-Scale 3D Database of Object Instances," in *International Conference on Robotics and Automation*, 2014.

[2] B. Li, Y. Lu, C. Li, A. Godil, T. Schreck, M. Aono, *et al.* (2014). *SHREC'14 Track: Large Scale Comprehensive 3D Shape Retrieval*. 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 *Robotics and Automation, 2009. ICRA '09. IEEE International Conference on*, 2009, pp. 1710-1716.

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

[5] S. Chitta, I. Sucan, and S. Cousins, "MoveIt!," *IEEE Robotics Automation Magazine,* vol. 19, pp. 18-19, March 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 *Rehabilitation Robotics, 2009. ICORR 2009. IEEE International Conference on*, 2009, pp. 510-517.

[7] http://amazonpickingchallenge.org/.
[8] J. Adams, K. Hodges, J. Kujawa, and C. Metcalf, "Test-retest Reliability of the Southampton Hand Assessment Procedure," *International Journal of Rehabilitation Research,* vol. 32, p. S18, 2009.
[9] S. Kalsi-Ryan, A. Curt, M. Verrier, and F. MG, "Development of the Graded Redefined Assessment of Strength, Sensibility and Prehension (GRASSP): reviewing measurement specific to the upper limb in tetraplegia," *Journal of Neurosurg Spine,* vol. 1, pp. 65-76, 2012.
[10] N. Yozbatiran, L. Der-Yeghiaian, and C. SC, "A standardized approach to performing the action research arm test," *Neurorehabil Neural Repair,* vol. 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 Am,* vol. 35, pp. 30-7, Jan 2010.
[12] A. P. del Pobil, "Benchmarks in Robotics Research," 2006.
[13] R. Madhavan, R. Lakaemper, and T. Kalmar-Nagy, "Benchmarking and standardization of intelligent robotic systems," in *Advanced Robotics, 2009. ICAR 2009. International Conference on*, 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, Springer Berlin Heidelberg, 2005, pp. 403-414.
[15] R. Detry, O. Kroemer, and D. Kragic, "International Workshop on Autonomous Grasping and Manipulation: An Open Challenge," 2014.
[16] (2006). *AIM@SHAPE Shape Repository v4.0*. Available: http://shapes.aim-at-shape.net/viewmodels.php
[17] A. Richtsfeld. (2012). *The Object Segmentation Database (OSD)*. Available: http://www.acin.tuwien.ac.at/?id=289
[18] P. Shilane, P. Min, M. Kazhdan, and T. Funkhouser, "The Princeton Shape Benchmark," in *Shape Modeling Applications, 2004. Proceedings*, 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," *The International Journal of Robotics Research,* 2012.
[20] S. Ulbrich, D. Kappler, T. Asfour, N. Vahrenkamp, A. Bierbaum, M. Przybylski*, et al.*, "The OpenGRASP benchmarking suite: An environment for the comparative analysis of grasping and dexterous manipulation," in *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, 2011, pp. 1761-1767.
[21] B. Li, A. Godil, M. Aono, X. Bai, T. Furuya, L. Li*, et al.* (2012). *SHREC'12 Track: Generic 3D Shape Retrieval*. 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, 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 *Signal Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012 Asia-Pacific*, 2012, pp. 1-10.
[24] R. B. Rusu. (2011). *NIST and Willow Garage: Solutions in Perception Challenge*. Available: http://www.nist.gov/el/isd/challenge-022511.cfm; http://www.willowgarage.com/blog/2011/02/28/nist-and-willow-garage-solutions-perception-challenge
[25] H. Dutagaci, A. Godil, P. Daras, A. Axenopoulos, G. Litos, S. Manolopoulou*, et al.* (2011). *SHREC'11 Track: Generic Shape Retrieval*. 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*, et al.*, "A category-level 3-D object dataset: Putting the Kinect to work," in *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, 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 *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, 2011, pp. 1817-1824.
[28] Z. Lian, A. Godil, T. Fabry, T. Furuya, J. Hermans, R. Ohbuchi*, et al.* (2010). *SHREC'10 Track: Non-rigid 3D Shape Retrieval*. Available: http://tosca.cs.technion.ac.il/book/shrec_robustness2010.html
[29] R. Wessel, m. Blumel, Ina, and R. Klein, "A 3D Shape Benchmark for Retrieval and Automatic Classification of Architectural Data," in *Eurographics 2009 Workshop on 3D Object Retrieval*, 2009, pp. 53-56.
[30] N. Iyer, S. Jayanti, and K. Ramani, "An Engineering Shape Benchmark for 3D Models," in *Proceedings of IDETC/CIE 2005 ASME 2005 International Design Engineering Technical Conferences \& Computers and Information in Engineering Conference*, 2005.
[31] R. Fang, A. Godil, X. Li, and A. Wagan, "A New Shape Benchmark for 3D Object Retrieval," in *Proceedings of the 4th International Symposium on Advances in Visual Computing*, Berlin, Heidelberg, 2008, pp. 381-392.
[32] (2005). *McGill 3D Shape Benchmark*. 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," in *Topics in Spinal Cord Injury Rehabilitation*, 2012.
[34] R. W. Sumner and J. Popovic, "Deformation Transfer for Triangle Meshes," *ACM Transactions on Graphics,* vol. 3, 2014.

[35] (2003). *NTU 3D Model Benchmark*. Available: http://3d.csie.ntu.edu.tw/

[36] (2014). *the ITI database*. Available: http://vcl.iti.gr/3d-object-retrieval/

[37] (2014). *Model Bank Library*. 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, Springer Berlin Heidelberg, 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 *Robotics and Automation, 2009. ICRA '09. IEEE International Conference on*, 2009, pp. 2493-2498.

[40] G. A. Kragten, C. Meijneke, and J. L. Herder, "A proposal for benchmark tests for underactuated or compliant hands," *Mechanical Sciences,* vol. 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," *The American journal of occupational therapy,* vol. 89, 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," *Occupational Therapy Journal of Research,* vol. 5, pp. 24-38, 1985.

[43] S. Rehab. *http://saliarehab.com/*.

[44] *GRASSP*. Available: grassptest.com

[45] *Grasp and Release Test*. Available: http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=1053

[46] L. Resnik, L. Adams, M. Borgia, J. Delikat, R. Disla, C. Ebner*, et al.*, "Development and Evaluation of the Activities Measure for Upper Limb Amputees," *Archives of Physical Medicine and Rehabilitation,* vol. e4, pp. 488-494, 2013.

[47] M. Popovic and C. Contway, "Rehabilitation Engineering Laboratory Hand Function Test For Functional Electrical Stimulation Assisted Grasping," 2003.

[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," *Journal of Rehabilitation M,* vol. 40, 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 Engineering in Medicine and Biology Magazine,* vol. 1, pp. 82-93, 2001.

[50] J. Harris, J. Eng, W. Miller, and D. AS, "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, 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 *Proceedings of the MyoElectric Controls/Powered Prosthetics Symposium*, 2005.

[52] K. Matheus and A. M. Dollar, "Benchmarking grasping and manipulation: Properties of the Objects of Daily Living," in *IROS*, 2010, pp. 5020-5027.

[53] *The Dash - "Disabilities of the Arm, Shoulder and Hand"*. 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 Symposium on Geometry Processing, 2006.

[55] http://rll.eecs.berkeley.edu/ycb/.

[56] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs*, et al.*, "ROS: an open-source Robot Operating System," *ICRA workshop on open source software,* vol. 3, p. 5, 2009.

[57] R. Diankov and J. Kuffner, "OpenRAVE: A Planning Architecture for Autonomous Robotics," 2008.

[58] N. Ratliff, M. Zucker, J. A. Bagnell, and S. Srinivasa, "CHOMP: Gradient Optimization Techniques for Efficient Motion Planning," *Icra: 2009 Ieee International Conference on Robotics and Automation, Vols 1-7,* pp. 4030-4035, 2009.

[59] D. Berenson, S. S. Srinivasa, D. Ferguson, and J. J. Kuffner, "Manipulation planning on constraint manifolds," in *Robotics and Automation, 2009. ICRA '09. IEEE International Conference on*, 2009, pp. 625-632.

[60] M. Garland and P. S. Heckbert, "Surface simplification using quadric error metrics," presented at the Proceedings of the 24th annual conference on Computer graphics and interactive techniques, 1997.

[61] S. S. Srinivasa, D. Berenson, M. Cakmak, A. Collet, M. R. Dogar, A. D. Dragan*, et al.*, "Herb 2.0: Lessons Learned From Developing a Mobile Manipulator for the Home," *Proceedings of the IEEE,* vol. 100, pp. 2410-2428, 2012.

[62] I. A. Sucan, M. Moll, and L. E. Kavraki, "The Open Motion Planning Library," *Ieee Robotics & Automation Magazine,* vol. 19, pp. 72-82, Dec 2012.

[63] http://www.ycbbenchmarks.org.

[64] http://protocols.ycbbenchmarks.org.

[65] R. R. MA, L. U. Odhner, and A. M. Dollar, "A Modular, Open-Source 3D Printed Underactuated Hands,," in *IEEE International Conference on Robotics and Automation (ICRA)*, Karlsruhe, Germany, 2013.

[66] W. Garage. (2009). *Overview of the PR2 robot*. Available: http://www.willowgarage.com/pages/pr2/overview

APPENDIX A. PROTOCOL AND BENCHMARK TEMPLATE FOR MANIPULATION RESEARCH:

,

MANIPULATION PROTOCOL TEMPLATE

Reference No / 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 No / Version	
Authors	
Institution	
Contact information	
Adopted Protocol	
Scoring	
Details of Setup	
Results to Submit	