An experiment to evaluate robotic grasping of occluded objects

Arnon Hurwitz, Marshal Childers, Andrew Dornbush, Dhruv Mauria Saxena, Maxim Likhachev, et al.

Arnon Hurwitz, Marshal Childers, Andrew Dornbush, Dhruv Mauria Saxena, Maxim Likhachev, Craig Lennon, "An experiment to evaluate robotic grasping of occluded objects," Proc. SPIE 10640, Unmanned Systems Technology XX, 106400E (3 May 2018); doi: 10.1117/12.2305498

Event: SPIE Defense + Security, 2018, Orlando, Florida, United States

An Experiment to Evaluate Robotic Grasping of Occluded Objects

Arnon Hurwitz[a], Marshal Childers[a], Andrew Dornbush[b], Dhruv Mauria Saxena[b], Maxim Likhachev[b], Craig Lennon[a]

[a]US Army Research Laboratory, 2800 Powder Mill Rd. Adelphi, MD 20783

[b]Carnegie Mellon University Robotics Institute, 5000 Forbes Ave, Pittsburgh, PA 15213

ABSTRACT

In December of 2017, members of the Army Research Laboratory's Robotics Collaborative Technology Alliance (RCTA) conducted an experiment to evaluate the progress of research on robotic grasping of occluded objects. This experiment used the Robotic Manipulator (RoMan) platform equipped with an Asus Xtion to identify an object on a table cluttered with other objects, and to grasp and pick up the target object. The identification and grasping was conducted with varying input factor assignments following a formal design of experiments; these factors comprised different sizes of target, varied target orientation, variation in the number and positions of objects which occluded the target object from view, and different levels of lighting. The grasping was successful in 18 out of 23 runs (78% success rate). The grasping action was conducted within constraints placed on the position and orientation of the RoMan with respect to the table of target objects. The Statistical approach of a 'deterministic' design and the use of odds ratio analysis were applied to the task at hand.

Keywords: Robotics Collaborative Technology Alliance, RCTA, robotic manipulation, robotic perception, deterministic design, odds ratios.

1. INTRODUCTION

The Army Research Laboratory's Robotics Collaborative Technology Alliance (RCTA) is conducting research to develop robotic capabilities which will allow robots to be better teammates. One such capability is the ability of a robot to do work in a real world environment, which may require manipulating objects that are only partially visible to the robot. Previously, the RCTA conducted experiments in which a robotic system attempted to search for, identify, and grasp a designated object [1, 2]. In these experiments, the activity of the robot was limited both by the perception of the system, and by the ability of the system to make full use of the manipulators on the platform. As a step toward improving robotic capabilities in general, and this system's capabilities in particular, RCTA researchers from Carnegie Mellon University developed perception algorithms to support robotic grasping of occluded objects.

The platform used in both the previous[2] and present assessment was the General Dynamics Land Systems Robotic Manipulator platform (RoMan). This platforms combines a Talon base, for mobility on rough terrain, with RoboSimian arms. In the previous assessment[2], the platform searched a room for a designated target object. The room was devoid of clutter, except for a few obstacles deliberately placed to obscure the target object from certain vantage points. When the robot reached a position from which it could obtain a clear view of the target object, it would approach and attempt to grasp the target object, maintaining a clear view of the target object during the entire process. The target was always the same, and was positioned in such a way as to be fully visible, approachable, and graspable from at least one avenue of approach. In general, the robot was able to grasp the target object when it could identify it.

In the present experiment, the focus was on the robot's ability to identify and grasp an object which might be occluded by other objects. The robot was placed in close proximity to the target object area, which was a table with a variety of objects on it, and attempted to grasp and lift the target object. The table on which the target object was located was cluttered with other objects, presenting a challenge to perception. So, while the system was expected to perceive and grasp an object in a complicated environment, it was not expected to move to a position from which to grasp, or to avoid contact with other objects subsequent to the grasping.

Section 2 describes the platform and the planning algorithms guiding the grasping. Section 3 describes the experiment and its results, while section 4 presents our conclusions.

Unmanned Systems Technology XX, edited by Robert E. Karlsen, Douglas W. Gage,
Charles M. Shoemaker, Hoa G. Nguyen, Proc. of SPIE Vol. 10640, 106400E
 · doi: 10.1117/12.2305498

2. PLATFORM AND PLANNING ALGORITHMS

The RoMan platform had a configuration similar to the platform used in previous experiments[2], and is described in detail in section 2.2. What was new in the present experiment was the perception algorithms[3, 4]. These algorithms improved on those examined before[2] by allowing for the identification and grasping of known but occluded objects in a scene, and are described below in section 2.1. The algorithms have important restrictions, which were taken into account in the experiment. First, 3D models of potential target objects must be known ahead of time. Second, for optimal performance, the number and type of objects in the scene needs to be known ahead of time. Third, objects are allowed to vary only in their position on table and in their yaw. For example, suppose that a model of an upright cup was a known model. Then a cup could be set upright at any position on the table and rotated so that the handle was facing in any direction, and would still be recognized as a cup. However, a cup on its side on the table would not be considered the same as the known cup, and would not be recognized unless a separate object, (cup-on-its-side) was introduced as a new model.

2.1 Planning Algorithms

The Perception via Search (PERCH) algorithm[3] approaches the problem of identifying occluded objects among a group of other, known, objects as an optimization problem. In this formulation, there is a set of hypothesized (potentially correct) scenes, and finding the scene which most closely matches the observed scene is a combinatorial search problem, in which there is a cost reflecting the difference between the hypothesized and perceived scenes. The hypothesized scenes are generated via a branching process in which new (hypothesized) objects are added one at a time. The scene generation is performed with the restriction that no added object may occlude another object (foreground to background) to reduce the size of the search. The cost function is designed to be monotonic decreasing with the number of (accurately) hypothesized objects, allowing the optimization to be conducted as a tree search, in which a path through the search space is guided by multiple heuristics, most importantly the number of objects represented and how many points of the observed point clouds can be assigned to the chosen hypothesized objects. The search is conducted using the Focal-MHA algorithm[5]. An example search tree is shown in Figure 1.

Note: The PERCH, algorithm requires that all objects in the scene be known in advance. To remove this restriction, a version of PERCH which allows clutter (C-PERCH) has been developed[4]. The C-PERCH algorithm allows points in the observed point cloud to be labeled a clutter, with a penalty equal to a parameter multiplied by the number of points so labeled. Thus scenes in which there are a higher proportion of unknown objects could be handled by lowering the value of the penalty parameter. There is, however, a cost to this in performance, as dealing with more unknown objects (more clutter) by lowering this parameter results in poorer recognition of known objects in complicated scenes[4]. The C-PERCH algorithm was not used in this experiment.

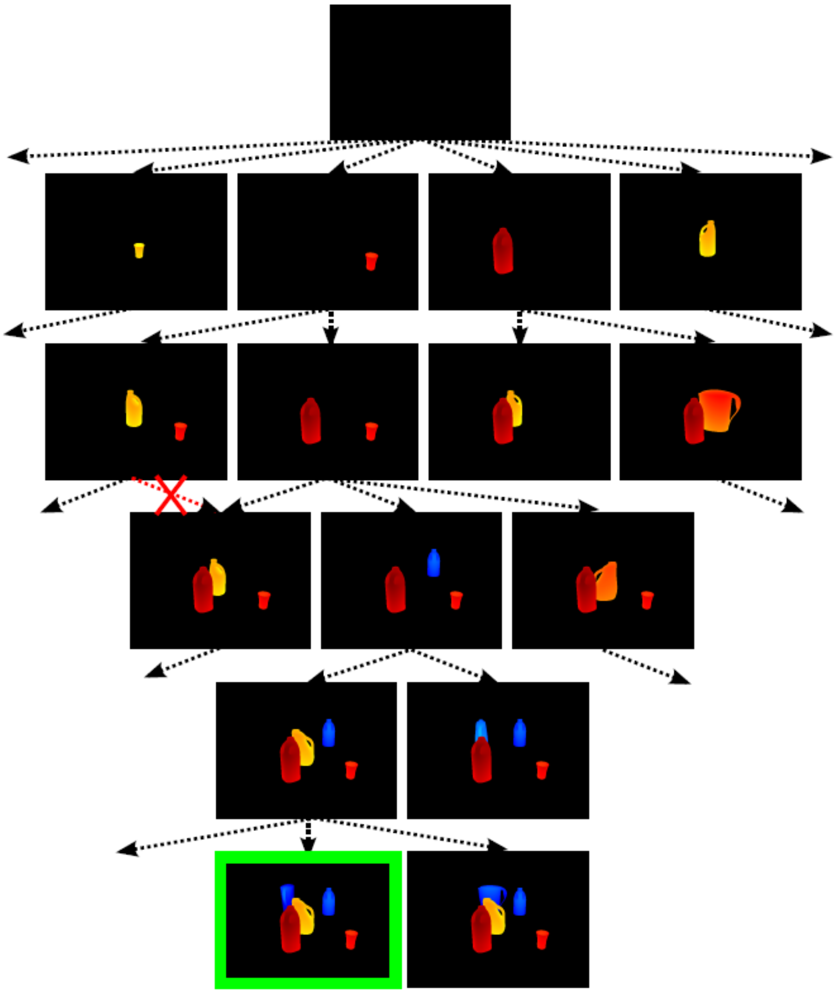


Figure 1: Example search tree for the PERCH algorithm[3]. The algorithm proposes new objects to be added to the background, preferring configurations where the proposed objects most closely match the observed point cloud. The red X shows that adding objects to the foreground is prohibited. The green box indicates the final scene.

2.2 Description of the RoMan

A RoMan platform, as configured for the experiment, is shown in Figure 2. The RoMan has two RoboSimian arms, only one of which was used for grasping. That arm, and the 3-Finger Robotiq Gripper used as the end effector, are labeled as (A) in Figure 2. The sensor is an Asus Xtion camera with depth sensing, labeled as (B) in Figure 2.

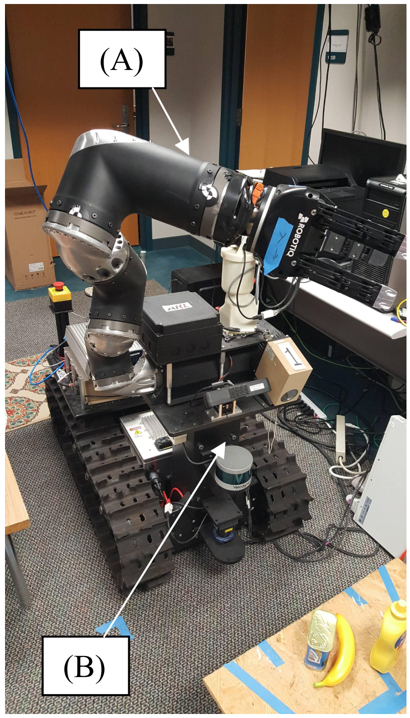


Figure 2: RoMan platform. This platform is equipped with a seven-degree-of-freedom arm, gripper, sensor suite, and capable of autonomously detecting and grasping objects

3. EXPERIMENTAL DESIGN AND RESULTS

3.1 Experimental design: Input factors & levels

Four input factors of greatest interest were chosen by Engineering judgement, (viz. Target size; Number of occluding objects; Target orientation; Placement of occluding objects). A fifth factor (Spotlight shining onto area of experiment) was assigned as a blocking factor to check for any 'lighting' effect on performance. These factors are listed in Table 1. A 'Low', 'Medium', and 'High' setting was chosen for each input factor, and the Spotlight factor was given 'On' or 'Off' as its two settings. These level settings are laid out in Table 1.

Table 1: Input factors and level settings for designed experiment.

Inputs (abbreviation)	Low (-1)	Medium (0)	High (1)
Size of target (TSIZE)	Small	Medium	Large
Number of Occluding Objects (ONUM)	2	3	4
Target Orientation (ORIENT)	Left	Center	Right
Occluding Object Placement (OPLACE)	Clustered	Even	Scatter
Spotlight (SPOT)	On/Off (1/0)		

3.2 Experimental design: Outputs

The outputs measured or observed are shown in Table 2. Comments were also recorded during the experiment. Time to localize was the time in seconds taken from the start of a run for the robot to determine the position of the target object. Time to drop was the total time (including Time to localize) from the start of the run to the moment the grasping arm dropped the grasped object. 'Grasp Success' (aka 'Grasp OK?') – that is, an initial successful grasp following PERCH Localization—was of primary interest, and is the main focus of this analysis.

Table 2: Outputs for designed experiment.

Variable Name	Value
Grasp Success	1 for success / 0 for failure
Wrong grasp	1 for grasped wrong object / 0 for correct object
Grasp Fail	1 for any type of grasp failure / 0 for grasp success
Time to localize	Seconds from start until target object identified
Time to drop	Seconds from start until target object is dropped (after pick up)

3.3 Objects and scenario description

The objects used as targets and occluding objects (see Figure 3) consisted of common household items including a cracker box, mustard bottle, soup can, sugar box, cup, clamp, cordless drill, canned meat, pitcher, and banana. The grasping algorithm considered three sizes of objects: large objects consisted of cracker box and pitcher, medium consisted of sugar box, cordless drill, clamp, and mustard bottle, and small consisted of the cup, canned meat, soup can, and banana.

Figure 3(a). Objects loosely arrayed with mustard bottle as target.

Figure 3(b). Objects tightly arrayed with cup as target and cracker box and canned meat occluding

The robot had a configuration similar to that of the previous experiment[1]. In this experiment, the chassis of the robot remained still with the torso and arm in motion as required in order to execute the grasp. Figures 4(a) and (b) show the orientation of the robot while performing a grasp and the setup for the runs that included the spotlight. The kinematics of the arm constrained the grasping area to a space of approximately 16" X 10".

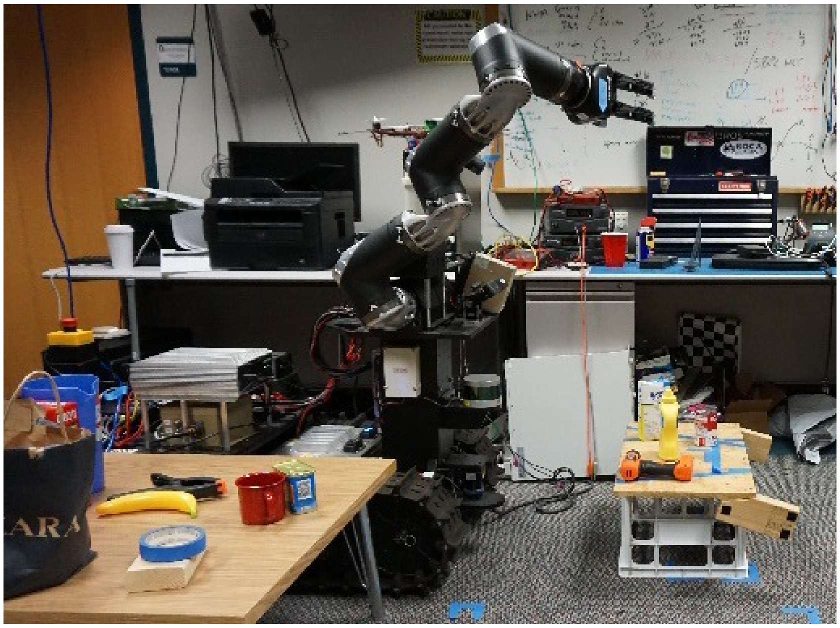


Figure 4(a). Robot executing grasp of mustard bottle.

Figure 4(b). Setup for the runs that included the spotlight.

3.4 Experimental design matrix

The design was a standard 2^4 full factorial[6] consisting of sixteen 'experimental' runs with five 'control' runs inserted (i.e. runs where all the input factor settings are taken at the '0' or nominal level). The potential outputs were deemed to be deterministic[9], so no replication was called for, and thus none was attempted except in the case of a run in which there was mechanical trouble, or in which the robot failed to successfully complete the initial identification ('Localization') and pick up. Our experimental plan of twenty-one runs is shown in Table 3.

Table 3: Experimental design matrix.

Run	DOE #	TSIZE	ONUM	ORIENT	OPLACE	SPOT (blocking)
1	0	0	0	0	0	0
2	1	-1	-1	-1	1	0
3	2	1	-1	-1	-1	0
4	3	-1	1	-1	-1	0
5	4	1	1	-1	1	0
6	0	0	0	0	0	0
7	5	-1	-1	1	-1	0
8	6	1	-1	1	1	0
9	7	-1	1	1	1	0
10	8	1	1	1	-1	0
11	0	0	0	0	0	0
12	9	-1	-1	-1	-1	1
13	10	1	-1	-1	1	1
14	11	-1	1	-1	1	1
15	12	1	1	-1	-1	1
16	0	0	0	0	0	1
17	13	-1	-1	1	1	1
18	14	1	-1	1	-1	1
19	15	-1	1	1	-1	1
20	16	1	1	1	1	1
21	0	0	0	0	0	1

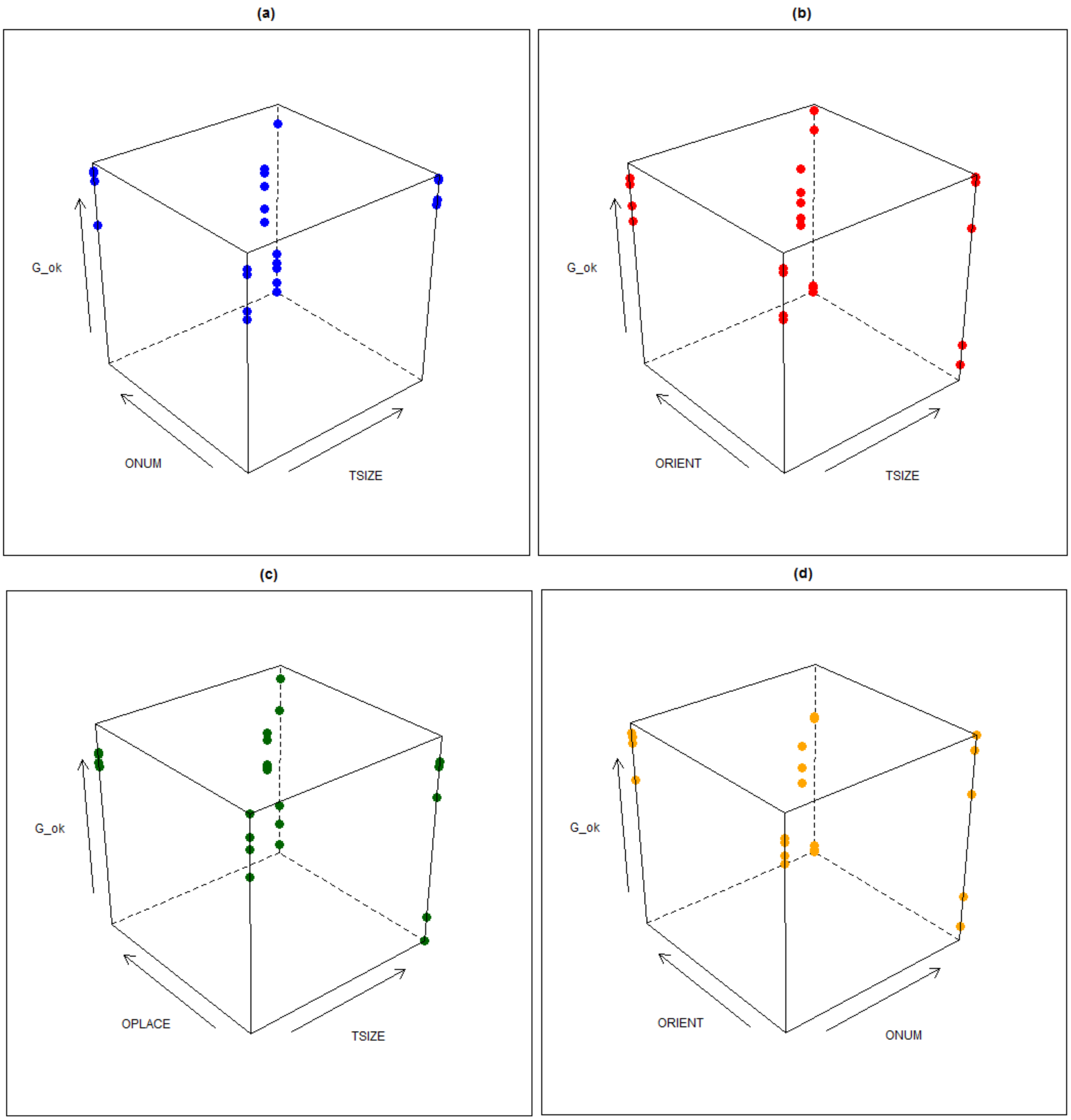
3.5 Outcomes of the experiment

The outcomes for each run are shown below in Table 4. Three runs had to be repeated, and those repetitions are denoted with a ‘B’ (for example, run 5 was repeated as 5B). Comments are entered as appropriate, with items of note presented in bold. We are primarily interested in the ‘Grasp OK?’ outcome, for which 1 denotes a successful grasp.

Table 4 is immediately followed by several 3D plots of the ‘Grasp OK?’ (G_ok) outcomes. Note that G_ok, which can take the values {0, 1}, is jittered in the vertical direction so that values which otherwise would overlay each other can be seen. These plots are labelled Figures 5(a) through 5(f).

Table 4: Observed outputs for the experiment.

	DOE	Input TSIZE	Input ONUM	Input ORIENT	Input OPLACE	Block SPOT	Grasp OK?	Wrong object?	Grasp fail?	localize Time (s)	drop Time (s)	COMMENTS
1	0	0	0	0	0	0	1	0	0	16.1	90.32	
2	1	-1	-1	-1	1	0	1	0	0	13.01	56.36	
3	2	1	-1	-1	-1	0	1	0	0	54.03	69.53	
4	3	-1	1	-1	-1	0	1	0	1	22.02	56.99	Cup (metal) fell from grasp
5		**1**	**1**	-1	1	0	**0**	1	1			**Failed to localize**
5 B	4	**1**	**1**	-1	1	0	**0**	1	1			**Failed to localize**
6	0	0	0	0	0	0	1	0	0	22.02	58.84	
7	5	-1	-1	1	-1	0	1	0	0	40.01	50.48	Grasped (metal) cup by lip
8	6	1	-1	1	1	0	1	0	0	56.05	79.61	
9	7	-1	1	1	1	0	1	0	0	15.01	48.15	
10	8	**1**	**1**	1	-1	**0**	0	1	1			**Failed to localize (sugar box)**
10 B	8	**1**	**1**	1	-1	**0**	0	1	1			**Failed to localize (CheezeIt box)**
11	0	0	0	0	0	0	1	0	0	16.01	88.13	
12	9	-1	-1	-1	-1	1	1	0	0	21.02	55.82	
13	10	1	-1	-1	1	1	1	0	1	19.01	68.17	Drill (plastic) fell from grasp
14	11	-1	1	-1	1	1	1	0	0	15.01	77.14	Grasped (metal) cup by lip
15	12	**1**	**1**	-1	-1	1	1	0	0	34.03	51.45	**knocked over can**
16	0	0	0	0	0	1	0	1	1			Torso failed to rotate
16 B	0	0	0	0	0	1	1	0	0	16.01	89.44	Repeated run: worked 2nd time
17	13	-1	-1	1	1	1	1	0	0	11.00	46.61	knocked over mustard
18	14	1	-1	1	-1	1	1	0	0	26.02	95.21	knocked over mustard
19	15	-1	1	1	-1	1	1	1	0	28.02	73.55	Grasped both cup & Spam can
20	16	**1**	**1**	1	1	1	**0**	0	1	54.05	98.41	**Pre-grasp failed (CheezeIt)**
21	0	0	0	0	0	1	1	0	0	22.01	63.52	knocked over mustard


(a)
G_ok
ONUM
TSIZE
(b)
G_ok
ORIENT
TSIZE
(c)
G_ok
OPLACE
TSIZE
(d)
G_ok
ORIENT
ONUM

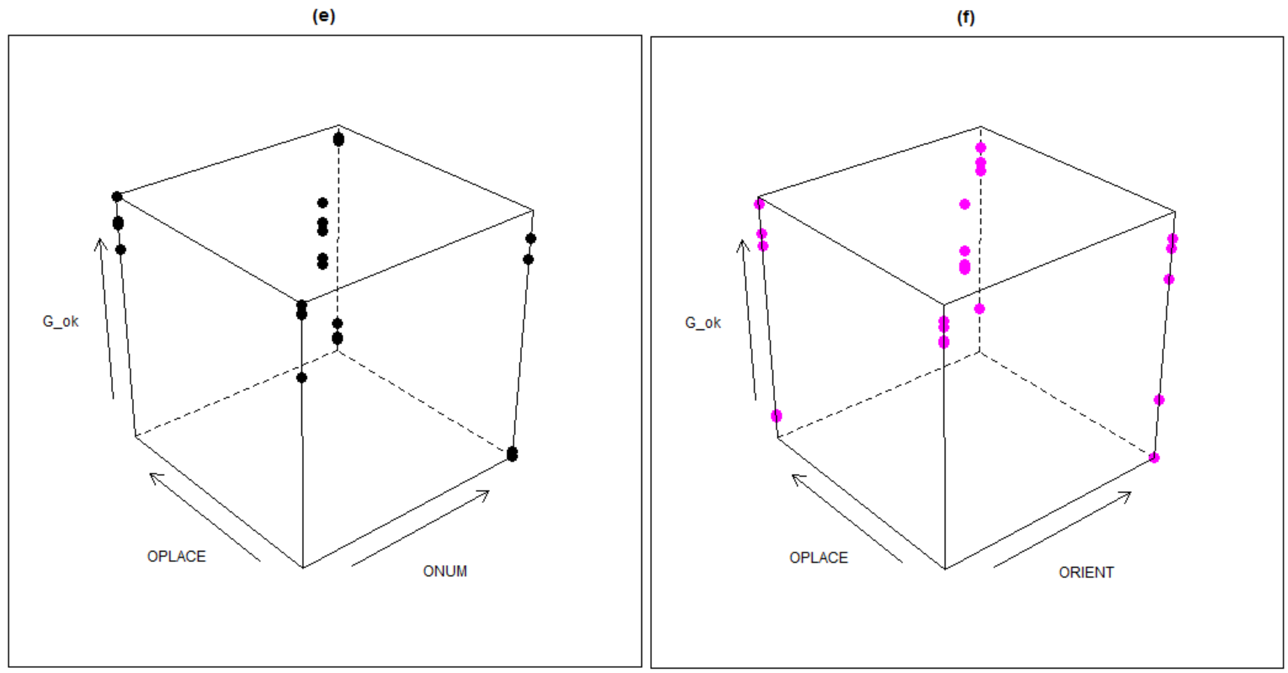


Figure 5(a) through (f): 3D plots of two-input combinations vs the (jittered) G_ok output.
[Note that the input (x, y) axes go from -1 through 0 to +1, and the vertical (z) axis from 0 to +1].

3.5.1 No observed effects from direct indoor lighting

Given the difficulties of IR depth sensing in sunlight, we wanted to examine whether there could be an effect from a strong, direct, indoor light source on the Asus Xtion sensor. Consequently, we added the spotlight as an experimental condition to determine whether a 'strong lighting' effect would alter the performance of the PERCH process, and to this end the 'Spotlight' factor was applied to light the area of activity in half the runs (see Figure 4(b). An independent-means, different-variances (i.e. a Welch[10]), t-test on the Time to Localize output was done to compare the effect of the 'Spotlight' on versus off input on this time. This test found no difference in the means. The probability of the observed means under the (null) hypothesis of no difference is 0.61. Our conclusion is that Spotlighting vs no Spotlighting makes no difference for the mean time to localize; time to drop produced a similar conclusion (prob. = 0.5). This may be due to truly no difference, as expected, or a large amount of noise in the measured times that obscures some (smaller) actual mean differences. Based on this, we expect that even strong and direct (indoor) lighting will not affect localization.

3.5.2 Perception effects are observed based on vantage point

In Table 4, note that all 'failed localization' events (apart from Run #16) occurred when TSIZE = 'Target Size' and ONUM= 'Number of Occlusion Objects' were set at their 'High=1' levels. Several other 'grasping' problems occurred when these same two inputs were set at their 'High' level (though run 15 was not a grasping issue *per se*). These items are highlighted in bold font in Table 4. This can also be seen in Figure 5(a).
Omitting the run (#16) where the torso failed to rotate, we constructed a 2x2 contingency table of the form [7, pg. 72]:

	Success	Failure
Treatment present	A	B
Treatment absent	C	D

Figure 6: A 2x2 'contingency' table.

In the case of Figure 5(a) that we examine here, 'Treatment absent' means: 'TSIZE and ONUM both = 1', and 'Treatment present' means its negation. 'Success' here means: 'G_ok = 1' (that is, the PERCH algorithm succeeded in localizing and grasping the target object). 'Fail' means: 'G_ok = 0'. The data we have from Table 4, (omitting Run #16 as a mechanical anomaly), gives Table 5(a). Following[7, pg. 76], because we have a 'zero' entry in the table, we add 0.5 to each entry of Table 5(a) to arrive at Table 5(b).

Table 5(a): 2x2 contingency table for G_ok 'Success' following from Figure 5(a).

	Success	Failure
TSIZE & ONUM not both = 1	**12**	**0**
TSIZE & ONUM both = 1	**1**	**5**

Table 5(b): 2x2 contingency table following from Table 5(a), with 0.5 added to all entries.

	Success	Failure
TSIZE & ONUM not both = 1	**12.5**	**0.5**
TSIZE & ONUM both = 1	**1.5**	**5.5**

Adding the 0.5 allows us to compute an approximate odds ratio (OR) of 'Success' to 'failure' for Table 5(b). This evaluates as an OR of (12.5x5.5/1.5x0.5) = 91.7. The odds of getting a G_ok = 1 is approximately 92 times more likely, on average, if the target size and number of occluding objects are not both at their 'High' settings simultaneously.

In order to get an approximate 95% confidence interval, based on these observations, we follow Woolf's method[8] and note that the log(odds ratio) (LOR) is approximately Normally distributed, with mean = log(OR) and standard error = sqrt(1/A+1/B+1/C+1/D). That is, a 95% confidence interval for LOR is ln(92) ± 1.96*1.71 = [1.17, 7.87]. As we are not interested in the LOR, but in the OR, we take exponents of the confidence interval's endpoints to arrive at a confidence interval for the OR of [3.21, 2633.05]. In other words, at least 95% of the time we expect an odds of success for G_ok that is between 3.21 and 2633 times greater if TSIZE and ONUM are not both at their 'High' settings vs. when they are both 'High.'

This result is worth noting, because it affects how a platform should approach an area in which it may identify a target object. In the experiment, the workspace was relatively small. This, coupled with the Xtion's field of view, sometimes makes it hard to see all of a large object. Given that the localization algorithm would receive partial point clouds in such cases, it is then challenging to successfully localize large objects. Consequently, during a search activity, the system may need to scan a scene from more than one distance in order to optimally identify objects, once from a further vantage point to ensure everything is in view, and one from a closer vantage point, at a distance which is optimal for grasping.

3.5.3 No conclusions drawn from object placement and orientation

In Table 4, note that no 'failed localization' events occurred when OPLACE = 'Placement of Occluding Objects' and ORIENT= 'Orientation of Target' were set at their 'Low = -1' levels... This can also be seen in Figure 5(f).

Following the same method for a confidence interval for the OR in the case of 'Treatment present' =='OPLACE & ORIENT both = -1', we have an adjusted 2x2 contingency table given as Table 6:

Table 6: Adjusted 2x2 contingency table for G_ok 'Success' following from Figure 5(f).

	Success	Failure
OPLACE & ORIENT both = -1	**4.5**	**0.5**
OPLACE & ORIENT not both = -1	**9.5**	**5.5**

This gives an approximate OR of 5.21, and a 95% confidence interval for the OR of [0.233, 116.2]. As this confidence interval contains 1 (i.e. equal odds ratios) there is no reason to expect that setting OPLACE and ORIENT both at their 'Low' settings together will give a better chance of Grasp_ok 'Success' than otherwise.

4. CONCLUSIONS

The PERCH grasping algorithm and RoMan grasping mechanism was largely successful in 18 out of our 23 counted runs (78% of the time) in this demonstration in isolating and grasping the target object from among several scenarios of occluding objects. The limitations of a tight workspace and a robot base placed in an immobile, close proximity to the workspace should be noted; however, the RCTA plans to integrate this improved grasping capability with existing mobility and search planners to enable a larger (room/building) search and grasp area.

REFERENCES

[1] Camden, R., Bodt, B., Childers, M. "An Assessment of an Autonomous Indoor Search and Grasp Capability," Proc. AUVSI 2015, 67-84 (2015).

[2] Lennon, C., Childers, M., Camden, R., Sapronov, L., Martin, M., Dornbush, A., Weller, E., Fields, M., Lebiere, C., "Assessment of RCTA Research," Proc. SPIE 10195, Unmanned Systems Technology XIX, 1019502 (2017).

[3] Narayanan, V., Likhachev, M., "PERCH: Perception via Search for Multi-Object Recognition and Localization," Robotics and Automation (ICRA), 2017 IEEE International Conference, 5052– 5059. IEEE, (2017).

[4] Narayanan, V., Likhachev, M., "Deliberative Object Pose Estimation in Clutter," Robotics and Automation (ICRA), 2017 IEEE International Conference, 3125–3130. IEEE, (2017).

[5] Narayanan, V., Likhachev, M., "Improved Multi-Heuristic A* for Searching with Uncalibrated Heuristics," Proceedings of Robotics: Science and Systems, (2014).

[6] Montgomery, D. C. "Design and Analysis of Experiments," John Wiley & Sons, Inc., (2005).

[7] Motulsky, H. 1995. Intuitive Biostatistics. Oxford University Press (New York). (1995).

[8] Woolf, B. On estimating the relation between blood group and disease. Ann. Human Genet. (London) 19, 251-253. (1955).

[9] Kuhn, D.R., Kacker, R.N. and Lei, Y. "Introduction to Combinatorial Testing," CRC Press (2013).

[10] Welch, B. L. "The generalization of "Student's" problem when several different population variances are involved" Biometrika **34** (1–2): 28–35 (1947).