Evaluating Human Gaze Patterns During Grasping Tasks: Robot versus Human Hand

Sai Krishna Allani[1*] Brendan John Javier Ruiz Saurabh Dixit[1] Jackson Carter[1] Cindy Grimm[1]
Ravi balasubramanian[1]
[1]Oregon State University, [2]Rochester Institute of Technology, [3]University of California, Santa Cruz

Abstract

Perception and gaze are an integral part of determining where and how to grasp an object. In this study we analyze how gaze patterns *differ* when participants are asked to manipulate a robotic hand to perform a grasping task when compared with using their own. We have three findings. First, while gaze patterns for the object are similar in both conditions, participants spent substantially more time gazing at the robotic hand then their own, particularly the wrist and finger positions. Second, We provide evidence that for complex objects (eg, a toy airplane) participants essentially treated the object as a collection of sub-objects. Third, we performed a follow-up study that shows that choosing camera angles that clearly display the features participants spend time gazing at are more effective for determining the effectiveness of a grasp from images. Our findings are relevant both for automated algorithms (where visual cues are important for analyzing objects for potential grasps) and for designing tele-operation interfaces (how best to present the visual data to the remote operator).

Keywords: robot grasp, eye-gaze, camera control

Concepts: •**Human-centered computing** → *Empirical studies in interaction design;*

1 Introduction

In this paper, we examine how human eye gaze differs between a human performing a grasp with their own hand versus positioning a robot hand in a similar grasp. Differences between eye gaze patterns can help distinguish extrinsic (visual) cues from intrinsic (touch, proprioception) ones, and more importantly, which visual cues humans use as substitutes for missing intrinsic cues. We perform a follow-up study that demonstrates that the visual cues identified from the eye-gaze patterns are important for effective grasp evaluation from images.

As robotics advances, it is expected that humans will guide robots in performing increasingly complex manipulation tasks in a variety of different environments. A wide spectrum of possible human-robot tele-operation interaction paradigms exist to enable this (see Fig. 1). At one extreme, the robot and the human are physically separate from each other — the robot provides just video information about the environment, and the human sends remote commands to the robot (such as a robot used for disaster rescue [Murphy 2004]).

*e-mail:allanis@oregonstate.edu

SAP '16,, July 22-23, 2016, Anaheim, CA, USA
ISBN: 978-1-4503-4383-1/16/07
DOI: http://dx.doi.org/10.1145/2931002.2931007

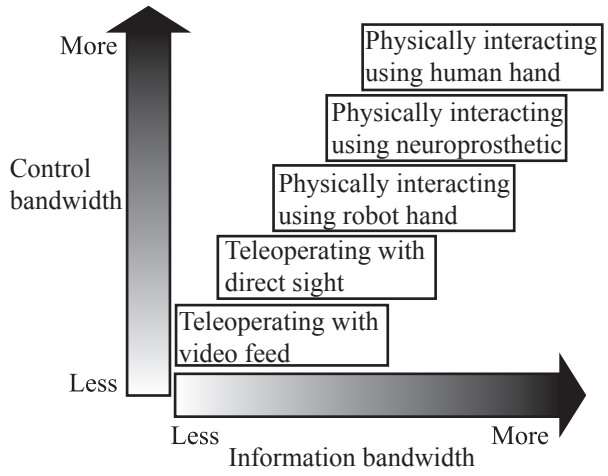

Figure 1: *Spectrum of human-guided robot manipulation in terms of available control (vertical axis) and information returned from the robotic manipulator (horizontal axis).*

The other extreme is when the human uses a neuroprosthetic robotic hand that provides touch and force information directly to the human's neural system and in turn receives commands through the neural system [Taylor et al. 2002]. In either case, the robotic manipulation can be thought of as an extended embodiment of the human. Across this spectrum, it is important to understand how humans process information when making decisions during physical interaction tasks in order to provide an effective interface. In this paper, we use eye gaze information to explore how visual information is processed in the context of grasping and manipulating objects, and specifically, how it changes when a human positions a robotic hand to perform the grasp versus using his/her own hand.

The spectrum of information available to the human when teleoperating a robot to perform a grasping task is diverse. It includes direct 3D views, 3D point clouds, 2D video or images, and contact and tactile information. Prior work has partially explored how to present the information to operators to get the quickest response time as well as the best decision from the operator [Drury et al. 2003; Murphy 2004; Burke and Murphy 2004; Steinfeld et al. 2006]. However, there is little work in understanding how the visual information provided to the operator is processed when performing physical interaction tasks such as grasping, and how humans might compensate for missing tactile cues using visual ones.

In this paper, we compare eye gaze between two different points on the manipulation spectrum. First, we use the human hand as an example of an "ideal robotic tool", where the human has the best information about the object and the manipulator and optimal control over the manipulator. Second, we use physically positioning a robotic hand as an example of tele-operation, where the operator has full, natural control and complete visuals. We analyze the eye gaze difference between the two conditions in three different stages of the manipulation task: pre-grasp, during manipulation, and post-

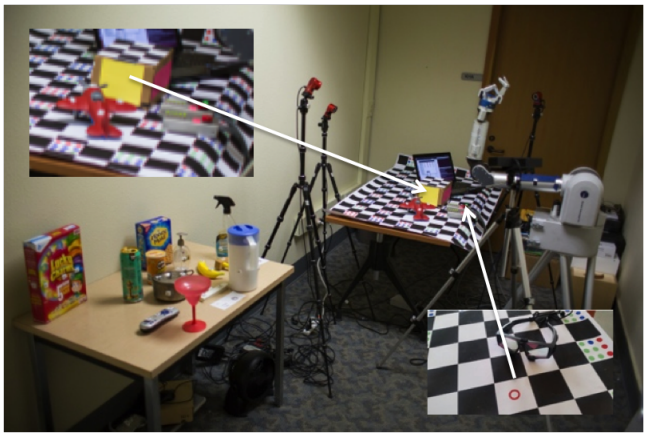

Figure 2: *Study set up. The table included a checkerboard pattern for further calibration. The red circle (inset, lower right) was used to calibrate the eye tracker. The box on the table (inset upper left) was used in the object placement tasks. Participants were seated at the table. The robotic arm can be moved by simply manipulating it; closing the fingers required sliders (in a box on the table).*

manipulation grasp evaluation. We use this information to design a simple camera viewpoint algorithm (based on the work of [Gooch et al. 2001; Christie and Olivier 2009]), which uses our identified features to determine the best view for grasp evaluation from images.

Understanding operator gaze is important in a grasping task because it provides information about how the human perceives the object and the task environment, and what visual cues are important to completing that task. Specifically, eye gaze provides information about how important different features of the object — such as the object's silhouette, surface, center of mass, and center line — are when performing the task with either a robot hand or a human hand. More subtly, *changes* in visual cues between using their own hand and the robotic one provide information about which features of the robotic hand's position (eg finger location, wrist orientation) are important. In particular, humans rarely look at their own hand when grasping. This information can be used to improve efficacy of remote and simulator interfaces by choosing camera angles that show these visual cues. Long-term, we can improve grasp planning algorithms by structuring the perceptual space based on these cues.

Contributions: We analyze visual cue differences for human-hand versus robot-hand manipulation tasks. We confirm visual cues found in previous human-hand studies [Lawrence et al. 2011; Desanghere and Marotta 2011; Prime and Marotta 2013] and additionally identify a more complex fixation pattern for more complex objects. For the robot-hand manipulation task, we identify the visual cues used by the human to compensate for missing proprioceptive cues. We use these cues to create better camera viewpoints for image-based grasp evaluation. We validate the viewpoints using an online survey.

2 Related work

The domain of robotic grasping and manipulation has seen significant progress both in terms of hardware [Dollar and Howe 2010; Birglen et al. 2008; Brown et al. 2010] and software development [Saxena et al. 2008; Lopez-Damian et al. 2005; León et al. 2010; Chitta et al. 2012]. However, there is a strong need to improve the ability of robots to robustly physically interact with the environment. Specifically, prior work has shown that even in a

Figure 3: *Objects used for the study.*

laboratory environment with almost perfect information for grasp planning, robotic grasping performance only succeeds about 75% of the time; that is, one in four grasps fail [Balasubramanian et al. 2012]. The primary reason for this poor performance is that even small differences in object shape or object position cause the object to, say, slip out during the grasping process. There has been significant effort to address these issues using physics-based heuristics and brute-force search algorithms to find more robust grasps with mixed success [Goins et al. 2015; Bohg et al. 2013; Balasubramanian et al. 2012; Miller and Allen 2004].

Prior work has also explored "learning from demonstration", where humans teach robots [Ekvall and Kragic 2004; Argall et al. 2009] to advance robot performance. However, most previous approaches for gathering data are time-intensive [Balasubramanian et al. 2010]. In prior work, we used crowd-sourcing where we have employed images or video of the grasps to receive human input [Unrath et al. 2014]. That work showed that humans are likely to over-estimate how successful the grasp will be. Despite this over-estimate, humans are still more accurate than learning approaches that use standard grasp metrics (for instance, center of grasp, center of mass) for certain subsets of grasp types. In on-going work we have also shown that view point and rendering can influence accuracy in this context; part of the goal of this work is to automate viewpoint selection for on-line or simulation applications.

Other work in the context of learning from demonstration also revealed a novel heuristic that humans use for improving grasp quality, namely, "skewness" where the human aligns the robot's wrist to the object's principal axis [Balasubramanian et al. 2012]. We also studied gaze patterns for evaluating static images of grasps [Sundburg et al. 2016], which showed that participants use many of the same cues as they do for grasping, and that participants are similarly likely to overestimate the effectiveness of grasps that look "human". However, no prior work has studied human eye gaze in 3D when controlling a robot arm in a physical interaction task.

There is a growing body of prior work on where humans look when performing grasps using their own hands [Lawrence et al. 2011; Desanghere and Marotta 2011; Prime and Marotta 2013]. This work showed that that people's gaze patterns are a mix of tracking the object's center of mass, looking at the top of the object, and looking at where the forefinger will make contact with the object (which in their case was the top of the object). Varying the task [Desanghere and Marotta 2011] or asking the participants to do the grasp from memory [Prime and Marotta 2013] changed the ratios of which region was gazed at, and in what order, but did not substantially change the types of regions. In this study (and our previous study of gaze patterns in images [Sundburg et al. 2016]) we see that these same patterns hold for the robotic grasping task, but that participants (not too surprisingly) also spend substantial time looking at the fingers, wrist, and other contact points.

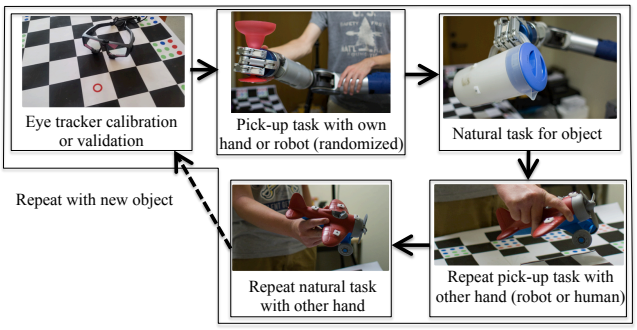

Figure 4: *Flow chart of study procedure.*

There are several existing techniques for camera viewpoint selection [Gooch et al. 2001; Christie and Olivier 2009] which evaluate a large number of visually salient features and artistic guidelines. Our contribution is a method for picking these features (and weighting them) for the more specific task of grasp evaluation.

3 Physical study (PS): Method

In this section we describe our physical-interaction user study where the participants performed specific object manipulation tasks, both with their own hand and by physically manipulating a robotic arm to do the task (see Figure 2). The participants performed two tasks per object. The participant's eye-gaze was tracked throughout, and they were also asked to "think aloud" — verbally evaluating their choice of grasp.

We first describe the study protocol, including objects, tasks, and participant pool. We then describe our eye-gaze analysis approach (Section 4), then provide results on eye-gaze differences between the two conditions (Section 5).

The visual cues from this physical study was used to develop camera viewpoints for grasp evaluation. These viewpoints were then validated in an online survey, which is described in section 6. Overall, this work was part of a larger study; we only describe the study elements relevant to this paper. The study was approved by Oregon State University's Institutional Review Board.

3.1 PS: Objects and tasks

A photo of the objects is in Figure 3. For each object the participants were asked to perform two tasks. The first was to pick up the object from the table and place it on a box on the table. The second task was object-specific (see Table 1). These object-specific tasks are tasks or actions that are associated naturally with each object, such as throwing a ball, squeezing a trigger, or handing the object to someone.

The participants performed the task with the robotic hand by grabbing the hand and moving it to the desired location and orientation. Unfortunately, changing the finger spread and closing the fingers has to be performed through software; we placed a box with physical sliders and a knob on the table to enable this (one slider for each finger plus a master slider to close all of them at once, the knob for finger spread). Once the object was secure in the grasp the participants moved the robot arm and hand as necessary to perform the task.

Table 1: *Object-specific tasks and number of grasps captured (including pick-up task).*

Object	**Natural Task**	**Total Grasps**
Water Pitcher	Pour water out of pitcher	11
Spray Bottle	Pull trigger to spray	14
Margarita Glass	Drink out of glass	14
Cereal Box	Pour cereal out of box	12
Cracker Box	Pour crackers out of box	15
Television Remote	Press power button on remote	11
Toy Plane	Pretend to fly plane around	13
Food Clip	Open clip as if using it to close bag	10
Soap Dispenser	Press down on nozzle to dispense soap	10
Foam Cylinder	Throw object overhand	16
Bison Plush Toy	Hand toy to someone	5
Plush Ball	Throw ball underhand	19
Sock Doll	Hand doll to someone	16
Decorative Cord	Hang cord by its metal ring	5
Tape Roll	Support tape roll so that another hand can be used to rip tape off	11
Total Grasps/object		182 12.1

3.2 PS: Capture and training phases

Our study protocol is designed to capture both human grasping and human-planned robotic grasping. To do this, the study features a training phase and two distinct (randomly ordered) capture phases: in the first capture phase the participants use their own hands to grab an object, while in the second, the participant physically positions the robotic arm and hand to grasp the object. To prevent learning effects, the order of the two phases was randomized for each object.

In the training phase (which happened before any data capture) participants were asked to familiarize themselves with the hand by moving it around and adjusting the fingers through the sliders. Although there was a gravity compensation mode for the arm, it did not adjust well when the hand was opened and closed, so participants were also given instructions to ask for help in supporting the hand if needed.

For the human-hand grasping phase, participants were asked to use just their thumb and first two fingers to mimic the three fingers of the robotic hand.

3.3 PS: Prompts and think-aloud

The subjects were asked to think out loud as they performed the study to provide insight into what they were thinking of while performing the grasping tasks.

For the move-the-object task, participants were asked to actually move the object using the robotic hand and arm. For the other tasks, they were not required to perform the task, but simply needed to position the hand. They were given explicit permission to pick up the object, position it how they wanted, and to use their other hand if they needed two hands.

Figure 5: *Example frames from the eye-tracking video showing the different eye gaze locations we identified (center, top, side, finger, wrist).*

At the end of each grasp participants were asked: "Is this grasp exactly what you wanted? Or are the finger placements slightly different that what you were intending? (How so?)". This prompt is aimed at determining how much the robotic hand limitations affected the participant's grasp choice.

3.4 PS: Data capture equipment and procedure

The equipment used for this study included a pair of SMI Eye Tracking Glasses 2.0 to collect eye-gaze data and a Barrett WAM Arm with BH280 BarrettHand to perform the robotic grasping. We also instrumented the working space with spatial calibration patterns (see Fig. 2), a Kinect sensor to track objects, and an audio cue to ensure calibration between data sources (eye-gaze, Kinect sensor, and BarrettHand).

Eye tracking: The SMI glasses record both where the user is looking and what they are looking at. The data is recorded as a 960x720 video stream at 30 Hz, plus an eye gaze location for each video frame (as x, y image coordinates). The eye gaze data also includes other information such as pupil diameter, fixations, and saccades. The eye tracker has to be fit to the person's head (similar to goggles) using two nose pieces and calibrated to their eyes. To perform the calibration the participant was asked to sit down in front of the table and fixate on a red dot on the table (see Figure 2). This one point calibration was performed using the SMI software. We checked the calibration at the end of each grasp trial by having the participant focus on the red dot again. We did not detect any significant calibration drift during the study.

Arm and hand tracking: We used a Barrett WAM and BarrettHand (BH-280) in the study. The arm is backdrivable and gravity compensated; that is, the arm location can be physically adjusted with ease. However, the BarrettHand's fingers cannot be physically adjusted from external forces (only through its motors). We used a physical set of three sliders to control how much each finger was closed, and a knob to control the spread of the fingers. Note that the two joints of the finger are controlled with one actuator.

Audio and temporal alignment: The eye-tracker records audio with the video. In addition to recording what the participant said we also used this information to temporally align the eye-tracker to the arm data streams using a generated beep. All other data alignment was through the Robot Operating System (ROS) toolkit.

3.5 PS: Protocol management and flow

The study is designed to be run by two researchers. One researcher handled the Ubuntu Linux PC running ROS and the arm, the other handled the SMI eye tracking laptop. Both researchers were involved in explaining the study and talking to the participant.

Table 2: *Annotation codes*

Step	Codes		
Phase	Pre-grasp	During grasp	Post-grasp
Regions (object)	Centerline	Top	Edges
Regions (hand)	Wrist	Finger tips	

The average time for a data collection session was an hour and a half, covering two grasps each for three or four objects. The maximum time was capped at two hours due to eye strain generated by the eye tracking glasses, as well as general fatigue from performing the experiment. New participants went through a training session to familiarize themselves with the robot arm and hand before starting data collection.

The general flow of the study can be seen in Figure 4 and is also outlined in the list below.

1. Participant enters room and signs consent form.
2. Brief training session with a test object (not in capture set).
3. Eye tracking calibration performed.
4. Study trials explained to participant.
   (a) Object placed on table, and participant told to use robot hand (group 1) or their own hand (group 2) to perform pick up task.
      i. Pick up task performed.
      ii. Repeat until no new grasps.
   (b) Natural task explained.
      i. Natural task performed.
      ii. Repeat until no new grasps.
   (c) Object-tasks (a-b) repeated with human hand (group 1) or robot hand (group 2)
   (d) Eye tracking recording stopped, and re-calibration if needed.
5. Repeat a-d with as many objects as possible
6. Eye tracking recording stopped, all other data collection ended.

3.6 PS: Participants

We recruited 15 participants, ranging in age from 16 [1] to late 50's, all with normal or corrected to normal with contacts vision (it is not possible to wear regular eye glasses with the Eye-gaze ones). On average participants specified 4.5 (maximum 8) grasps per object across the two tasks.

4 PS: Analysis of eye-tracking data

We perform two types of analysis on the eye-gaze data. The first analysis focuses on labeling what the participant was looking at before, during, and after grasping the object to perform the manipulation task. The second analysis focuses on differences in fixations

[1] High-school students involved in a summer STEM program.

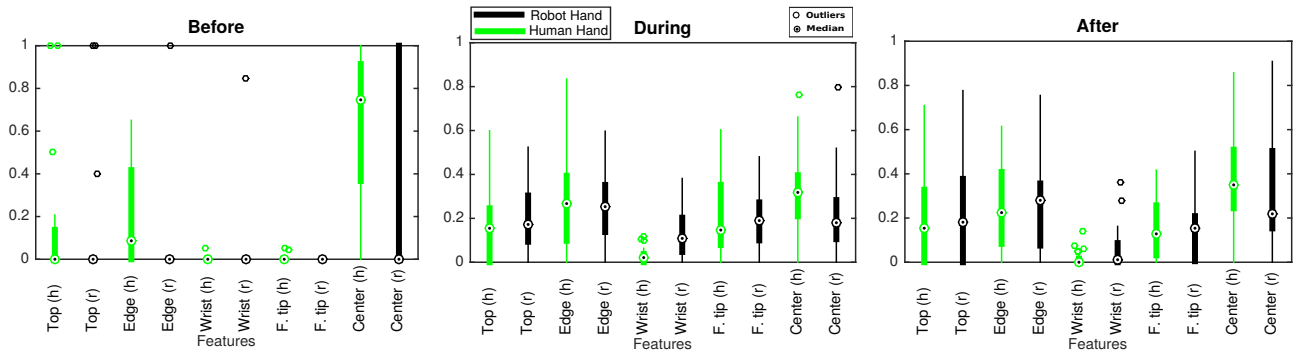

Figure 6: *Gaze time differences between the human hand and the robot hand. From left to right: Before, during, and after phases. Gaze time is normalized across* all *three phases. Circles with dots are medians; circles without dots are outliers.*

for different objects. Example frames from the video are shown in Figure 5.

4.1 PS: What did they look at?

We analyzed the eye-gaze data using a two-step process. In the first step, we identify three phases: Pre-grasp (when the participant is looking at the object but has yet to touch it), during grasp (when the participant closes the hand around the object) and post-grasp evaluation (when the participant evaluates the grasp in the think-aloud protocol). In the second step we label what the participant was looking at. The view is split into five regions, three of which focus on the object and two on the hand (see Table 2). Recall that we used eye tracking glasses so both the view point and the object were free to move. Reliably tracking the human hand and object's location with the Kinect proved to be difficult because of occlusions; for these reasons we chose to manually label the video.

To annotate the video data and produce the statistics we used MaxQDA [max 1989-2015]. We marked all time segments where the gaze was fixed on the object, hand, or robotic arm and hand for longer than thirty frames. To verify inter-coder reliability, we had a second coder repeat the coding for one participant; the code alignment was over 95%. For this analysis, the gaze points outside of the object and hand were ignored.

Not all participants had all codes, most notably, very few participants had a pre-grasp gaze for the robotic hand, and there were also 2 participants who had no pre-grasp gaze for the human hand. We hypothesize two reasons for this: The first is that peripheral vision was sufficient in some cases for the participant to categorize the object. The second is that if the participants were doing the robotic hand second they had no need to look at the object again (10 of 11 with no pre-grasp gaze).

4.2 PS: Fixations

We implemented the EyeMMV fixation detection algorithm [Krassanakis et al. 2014], which filters the coordinate sequences by applying a threshold of dispersion to the points. We used standard settings [Salvucci and Goldberg 2000] for the algorithm: a 90 ms minimum fixation duration and a maximum fixation dispersion of 0.5 degree of visual angle (DVA), with a preliminary filter of 5 pixels greater than $1/2$ DVA. We measured visual angle based on gaze frames where the participant was focused on the object. The algorithm produces a sequence of fixations, with each fixation centered at the average of the coordinates and lasting a given duration. Using fixations, over the raw coordinate data, both reduces processing time and removes saccades, where the viewer is essentially blind.

We overlay the fixation counts with the annotations to find the number of fixations on the object during each phase.

5 Physical study: Results

We first summarize the difference between the two conditions (human hand versus robot hand), then the differences in fixations between objects, and finally the think-alound results. Statistical significance is inferred at the $p = 0.05$ level.

5.1 PS: Gaze difference

Figure 6 shows the normalized gaze times for both using the robot hand and the human hand in the three phases of grasping (before, during, and after). Several patterns are clear. In the "before" phase when using the robot hand, the human subjects almost never look at the objects' top and edges or the robot's wrist or fingertips. They only focus on the object's centerline. This is in contrast to using their own hand, where the focus is primarily on the top and edges of the object, as found in previous studies [Lawrence et al. 2011].

In the "during" and "evaluation" phases, the two gaze patterns were more similar. Primary differences are that the participants spent more time observing the robot's wrist (versus looking at their own) and less time looking at the object's centerline.

Overall, during the robotic grasp task participants spent significantly less time looking at the edges and top of the object before beginning the manipulation (some participants barely glanced at the object before starting — see Figure 7). In human grasping studies, gaze on the edges and top corresponds to participants determining potential contact points for their fingers. We hypothesize that the lack of these pre-grasp glances for placing the robotic hand implies that visualizing contact points is part of the *control* strategy for guiding the hand to the desired grasp, but not for *planning* the grasp in the first place. The placement of the fingers relative to the edges or top of the object does, however, play a significant role in evaluating the grasp for both conditions.

Robotic hand control strategy: Participants varied on exactly how they moved the robot hand, but in general they positioned the hand roughly where they wanted it and with the desired finger spread, then iterated a few times between adjusting the fingers and repositioning the hand.

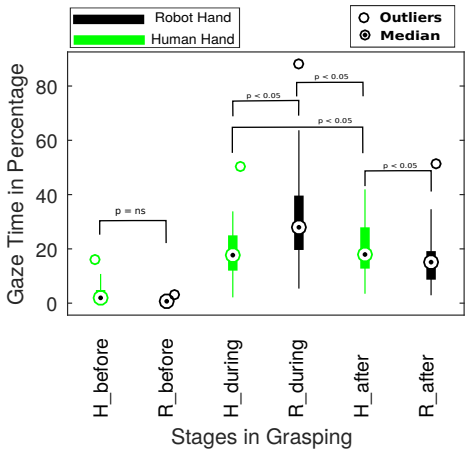

Figure 7: *Percentage time spent in each phase (before, during, and after grasping) for the human hand (H_, green) versus the robot hand (R_, black).*

5.2 PS: Fixation differences for objects

Objects with more complex geometry saw more fixations than simple objects in the "before" phase (see Fig. 8). We define complexity by the number of components produced by an automated shape analysis approach such as [Kaick et al. 2014].Their is a positive correlation between complexity of objects and number of fixation points, r = 0.83. From an informal observation of the gaze patterns, participants appeared to be moving between the center line and top of each convex regions of the object (eg, the wings to the plane body). A more formal evaluation of what regions they were looking at would require tracking the object in the video.

5.3 PS: Think-aloud evaluations

We did not formally analyze the participants comments; we summarize here overall statements. Around half of the participants said at least one grasp was not quite what they wanted, particularly for more complex objects such as the plane. The major refrain was that the participants didn't like that the joints in the fingers couldn't be controlled individually (the Barrett hand only supports bending the two finger joints simultaneously, not controlling each joint independently). This was most noticeable in cases where the finger locks up due to collision with the object — one part of the finger comes in contact and stops, while the remaining part of the finger stops where it is and doesn't close all the way around the object. Other issues were the fingers being too thick, the hand too big, or the controls being too fidgety to achieve some of the more precise grasps the participants had intended to perform.

6 On-line study (OS): Method

Physical studies provide very high-quality data, but are very time-consuming and do not scale. Previous work [Unrath et al. 2014] shows that we can leverage on-line surveys to quickly label and classify grasps by asking participants to evaluate images of them; however, on-going work also shows that view point selection plays a key role both in how effective participants are in labeling grasps and how confident they are. Our goal is to use the eye-tracking data to guide an automatic view-point selection algorithm for this

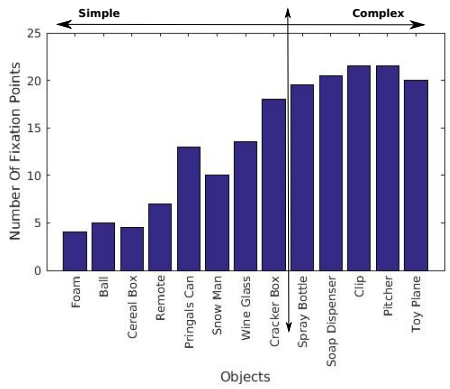

Figure 8: *Fixation counts on the objects, organized from simpler shapes to more complex ones, as measured by (approximately) the number of components a shape segmentation algorithm would produce.*

use case. Specifically, we use the relative percentage viewing time of the features during the robotic grasping hand stage to create an optimization function for specifying camera viewing angles. We used an on-line survey to evaluate if the algorithm selected views that are both effective and useful.

6.1 OS: Viewpoint optimization algorithm

Because gravity and object orientation are important components in grasp evaluation, we limit our viewpoint search to azimuth and elevation (effectively searching a hemisphere of viewpoints). The camera is pointed at the center of the object and the **up** vector aligned with gravity. Our optimization function for a given viewpoint is simply the sum of the percentage of visible pixels for each feature (normalized by the maximum number of pixels for that feature seen from any viewpoint). Each term is weighted by the percentage of time participants spent viewing that feature, averaged across all participants(top=0.17,edges=0.24,fingertips=0.27,wrist=0.024, center line=0.29). For the contact point features, we added a sphere roughly half the size of the finger tip, centered on the contact point. The best view is the one with highest score; the second best is the one with the next highest score that is at least 30 degrees from the best view.

6.2 OS: Survey format

Our survey was designed to measure both how effective the views were for evaluation (did this grasp work, yes or no?) and perceived usefulness. We used four viewpoints: best and second best views (good pair), and worst and second worst views (bad pair). The survey had four questions (5-point Likert scale) (complete survey in Supplemental materials). Each image was shown in all positions roughly equally.

- Q1 GBPairs work: (All good and bad pairs): Would the grasp work, yes or no, and how confident are you in your answer.
- Q2 GBPairs useful: (Good pair and bad pair): Rank the usefulness of the first pair with respect to the second pair.
- Q3 Second view useful: (Best and second best/Worst and second worst:) Rank the usefulness of the *second* view.
- Q4 Rank views: (Good pair and bad view): Rank the usefulness of the three views.

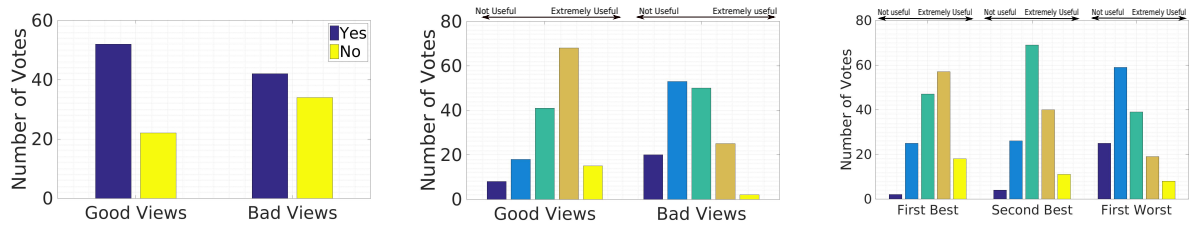

Figure 9: *Results of on-line survey. Left: Grasp works y/n (Q1 GBPairs work). Middle: Usefulness of view pairs (Q2 GBPairs useful). Right: Usefulness of first and second best view to worst (Q4 Rank views). (Useful is to the right.)*

For simplicity, we evaluated our algorithm with three objects (spray bottle, ball, glass), two grasps each. These were grasps given by participants in the physical study and subsequently verified as being effective using a shake test. Image order was randomized. Each of the 30 participants saw 20 of the 24 total questions, again randomized. Participants were recruited through Mechanical Turk; we verified that the participants spent enough time on each question to have seen the images and didn't click the same response for every question.

We determined that there was an order bias: Images on the left tended to be preferred over images shown on the middle or right (Question 1 mean 0.68 versus 0.35, Question 4 mean 3.45 versus 3.07 and 2.31, $p < 0.0007, 0.0005$ respectively). In a previous study [Sundburg et al. 2016] we also saw a distinct pattern of looking at the left image then the right, with only far more salient views on the right drawing the gaze first.

7 On-line study: Results

Refer to Figure 9. For Q1 (GBPairs work) participants were not only more likely to rank the good grasp pairs as effective (means 0.68,0.35, $p < 0.0007$) but were also more confident in their answer (means $2.3, 1.4$, $p < 0.05$). Grasp views were ranked as expected for usefulness both as pairs and individual images (Q2 GBPairs useful and Q4 Rank views). Interestingly, the second best (and first worst) views were both rated approximately as useful (mean $3.07, 2.31$, $p < 0.00039, 0.00019$) relative to their corresponding first view (Q4 Rank views). The answers to Q3 (Second view useful) did not yield data with statistical significance.

8 Discussion

We have presented an analysis of the difference in eye gaze when participants used their own hand versus manipulating a robotic one. This begins to provide an understanding of how humans substitute visual cues for tactile feedback when performing a physical interaction task. We have demonstrated that these visual cues lead to more effective view-point selection for grasp evaluation from images.

Future work will focus on view-point selection and visual feedback for virtual physical manipulation tasks, and on evaluating how gaze patterns differ from the physical robotic hand to the virtual one. We will also focus on generating optimal view points in simulation software (such as Rviz) to improve operator's effectiveness.

Acknowledgements

This research was supported in part by NSF grants CNS 1359480 (REU site: Robots in the real world) and IIS 1302142 and by a TORC Robotics grant AFR 03-101-OS-02. We gratefully acknowledge Dr. Reynold Bailey (Rochester Institute of Technology), who kindly loaned us the SMI eye-tracking unit for the summer.

References

ARGALL, B. D., CHERNOVA, S., VELOSO, M., AND BROWNING, B. 2009. A survey of robot learning from demonstration. *Robotics and Autonomous Systems 57*, 5, 469 – 483.

BALASUBRAMANIAN, R., XU, L., BROOK, P., SMITH, J., AND MATSUOKA, Y. 2010. Human-guided grasp measures improve grasp robustness on physical robot. In *Robotics and Automation (ICRA)*, 2294–2301.

BALASUBRAMANIAN, R., XU, L., BROOK, P., SMITH, J. R., AND MATSUOKA., Y. 2012. Physical human interactive guidance: A simple method to study human grasping. *IEEE Transactions on Robotics*. DOI: 10.1109/TRO.2012.2189498. (In press).

BIRGLEN, L., LALIBERTÉ, T., AND GOSSELIN, C. 2008. *Underactuated Robotic Hands*. Springer.

BOHG, J., MORALES, A., ASFOUR, T., AND KRAGIC, D. 2013. Data-driven grasp synthesis—a survey.

BROWN, E., RODENBERG, N., AMEND, J., MOZEIKA, A., STELTZ, E., ZAKIN, M. R., LIPSON, H., AND JAEGER, H. M. 2010. Universal robotic gripper based on the jamming of granular material. *Proceedings of the National Academy of Sciences 107*, 44, 18809–18814.

BURKE, J., AND MURPHY, R. 2004. Human-robot interaction in usar technical search: Two heads are better than one. In *Robot and Human Interactive Communication, 2004. ROMAN 2004. 13th IEEE International Workshop on*, IEEE, 307–312.

CHITTA, S., SUCAN, I., AND COUSINS, S. 2012. Moveit! *IEEE Robotics Automation Magazine 19*, 1, 18–19.

CHRISTIE, M., AND OLIVIER, P. 2009. Camera control in computer graphics: Models, techniques and applications. In *ACM SIGGRAPH ASIA 2009 Courses*, ACM, New York, NY, USA, SIGGRAPH ASIA '09, 3:1–3:197.

DESANGHERE, L., AND MAROTTA, J. 2011. graspability of objects affects gaze patterns during perception and action tasks. *Experimental Brain Research 212*, 2, 177–187.

DOLLAR, A. M., AND HOWE, R. D. 2010. The highly adaptive SDM Hand: Design and performance evaluation,. *Internat. J. Robotics Res 29*, 5, 585–597.

DRURY, J. L., SCHOLTZ, J., YANCO, H., ET AL. 2003. Awareness in human-robot interactions. In *Systems, Man and Cybernetics, 2003. IEEE International Conference on*, vol. 1, IEEE, 912–918.

EKVALL, S., AND KRAGIC, D. 2004. Interactive grasp learning based on human demonstration. In *Robotics and Automation (ICRA)*, vol. 4, 3519–3524 Vol.4.

GOINS, A., CARPENTER, R., WONG, W.-K., AND BALASUBRAMANIAN, R. 2015. Implementation of a gaussian process-based machine learning grasp predictor. *Autonomous Robots*, 1–13.

GOOCH, B., REINHARD, E., MOULDING, C., AND SHIRLEY, P. 2001. Artistic composition for image creation. In *Proceedings of the 12th Eurographics Workshop on Rendering Techniques*, Springer-Verlag, London, UK, UK, 83–88.

KAICK, O. V., FISH, N., KLEIMAN, Y., ASAFI, S., AND COHEN-OR, D. 2014. Shape segmentation by approximate convexity analysis. *ACM Trans. Graph. 34*, 1 (Dec.), 4:1–4:11.

KRASSANAKIS, V., FILIPPAKOPOULOU, V., AND NAKOS, B. 2014. Eyemmv toolbox: An eye movement post-analysis tool based on a two-step spatial dispersion threshold for fixation identification. *Journal of Eye Movement Research 7*, 1, 1–10.

LAWRENCE, J., ABHARI, K., PRIME, S., MEEK, B., DESANGHERE, L., BAUGH, L., AND MAROTTA, J. 2011. A novel integrative method for analyzing eye and hand behaviour during reaching and grasping in an mri environment. *Behavior Research Methods 43*, 2, 399–408.

LEÓN, B., ULBRICH, S., DIANKOV, R., PUCHE, G., PRZYBYLSKI, M., MORALES, A., ASFOUR, T., MOISIO, S., BOHG, J., KUFFNER, J., ET AL. 2010. Opengrasp: a toolkit for robot grasping simulation. In *Simulation, Modeling, and Programming for Autonomous Robots*. Springer, 109–120.

LOPEZ-DAMIAN, E., SIDOBRE, D., AND ALAMI, R. 2005. A grasp planner based on inertial properties. In *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*, IEEE, 754–759.

1989-2015. Maxqda, software for qualitative data analysis.

MILLER, A., AND ALLEN, P. K. 2004. Graspit!: a versatile simulator for robotic grasping. *IEEE Robotics and Automation Magazine*.

MURPHY, R. R. 2004. Human-robot interaction in rescue robotics. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on 34*, 2, 138–153.

PRIME, S., AND MAROTTA, J. 2013. Gaze strategies during visually-guided versus memory-guided grasping. *Experimental Brain Research 225*, 2, 291–305.

SALVUCCI, D. D., AND GOLDBERG, J. H. 2000. Identifying fixations and saccades in eye-tracking protocols. In *ETRA 2000*, ACM, New York, NY, USA, 71–78.

SAXENA, A., DRIEMEYER, J., AND NG, A. Y. 2008. Robotic grasping of novel objects using vision. *Int. J. Robotics Res. 27*, 2, 157–173.

STEINFELD, A., FONG, T., KABER, D., LEWIS, M., SCHOLTZ, J., SCHULTZ, A., AND GOODRICH, M. 2006. Common metrics for human-robot interaction. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, ACM, 33–40.

SUNDBURG, M., GRIMM, C. M., AND BALASUBRAMANIAN, R. 2016. Visual cues used to evaluate grasps from images. In *Proceedings of the IEEE International Conference on Robotics and Automation ICRA*.

TAYLOR, D. M., TILLERY, S. I. H., AND SCHWARTZ, A. B. 2002. Direct cortical control of 3d neuroprosthetic devices. *Science*.

UNRATH, M., ZHANG, Z., GOINS, A., CARPENTER, R., WONG, W.-K., AND BALASUBRAMANIAN, R. 2014. Using crowdsourcing to generate surrogate training data for robotic grasp prediction. In *Proceedings of the Second AAAI Conference on Human Computation and Crowdsourcing (HCOMP-14)*, 60–61.