Bandit-Based System Adaptation for Grasping

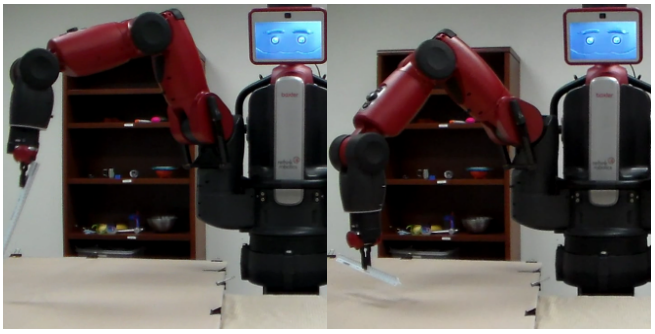
Abstract

A key aim of current research is to create robots that can reliably manipulate objects. However, in many instances, information needed to infer a successful grasp is latent in the environment and arises from complicated physical dynamics; for example, a heavy object might need to be grasped in the middle or else it will twist out of the robot's gripper. The contribution of this paper is a formalization of object picking as an N-armed bandit problem, where each potential grasp point corresponds to an arm with unknown pick success rate. This formalization enables us to apply bandit-based exploration algorithms to enable a robot to identify the best arm by attempting to pick up the object and tracking its successes and failures. Because the number of grasp points is very large, we define a new algorithm for best arm identification in budgeted bandits that computes confidence bounds while incorporating prior information, enabling the robot to quickly find a near optimal arm without pulling all the arms as in UCB-based approaches. We demonstrate that our adaptation step significantly improves accuracy over a non-adaptive system, enabling a robot to improve grasping models through experience.

1 Introduction

Robotics will assist us at childcare, help us cook, and provide service to doctors, nurses, and patients in hospitals. Many of these tasks require a robot to robustly perceive and manipulate objects in its environment, yet robust object manipulation remains a challenging problem. Systems for general-purpose manipulation are computationally expensive and do not enjoy high accuracy on novel objects [27]. A common source of error is the presence of latent dynamics that emerge from interactions between the object and the robot's gripper. For example, a heavy object might fall out of the robot's gripper unless it grabs it close to the center. Transparent or reflective surfaces that are not visible in IR or RGB make it difficult to infer grasp points [17].

To address these limitations, we propose an approach for enabling a robot to learn about an object through exploration and adapt its grasping model accordingly. We frame the problem of model adaptation as identifying the best arm for an N-armed bandit problem [31] where the robot aims to minimize simple regret after a finite exploration period [5]. Our robot can obtain a high-quality reward signal (although sometimes at a higher cost in time and sensing) by actively collecting additional information from the environment, and use this reward signal to adaptively identify grasp points that are likely to succeed.



(a) Before learning. (b) After learning.

Figure 1: Before learning, the robot grasps the ruler near the end, and it twists out of the gripper and falls onto the table; after learning, the robot grasps near the ruler's center of mass.

Identifying the best grasp point corresponds to best arm identification with pure exploration in bandit problems. Existing algorithms for best arm identification require pulling all the arms as an initialization step [19, 2, 6], a prohibitive expense when each arm pull takes on the order of 90 seconds and there are more than 1000 arms. To address this problem, we present a new algorithm, Prior Confidence Bound, based on Hoeffding races [20]. In our approach, the robot pulls arms in an order determined by a prior, which allows it to try the most promising arms first. It can then autonomously decide when to stop by bounding the confidence in the result. Figure 1 shows the robot's performance before and after training on a ruler; after training it grasps the object in the center, improving the success rate.

We evaluated Prior Confidence Bound on a Baxter robot, demonstrating that our adaptation step improves the overall

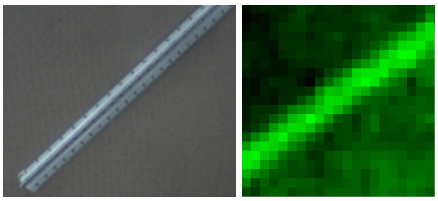


Figure 2: Automatically acquired RGB image for one of the objects in our training set, combined with the IR raster scan.

pick success rate from 55% to 75% on our test set of 30 household objects, shown in Figure 3. Moreover, our approach also enables the robot to learn success probabilities for each object it encounters; when the robot fails to infer a successful grasp for an object, it knows this fact, enabling it to take active steps to recover such as asking for help [30].

We first give an overview of our object detection and localization pipeline. Next we formalize our grasping framework as a bandit problem, where each arm corresponds to a grasp point on the object. Section 4 describes our evaluation in simulation and on the real robot with 30 household objects, Section 5 covers related work, and Section 6 concludes.

2 Object Detection and Localization

Our object detection and pose estimation pipeline uses conventional computer vision algorithms in a simple software architecture to achieve a frame rate of about 2Hz for object detection and pose estimation. Object classes consist of object instances rather than pure object categories. Using instance recognition means we cannot reliably detect categories, such as "mugs," but the system will be much better able to detect, localize, and grasp the specific instances for which it does have models.

Our detection pipeline runs on stock Baxter with one additional computer, using the arm cameras and IR sensor to localize objects in the environment. Our recognition pipeline takes video from the robot's wrist cameras, proposes a small number of candidate object bounding boxes in each frame, and classifies each candidate bounding box as belonging to a previously encountered object class.

When the robot moves to attempt a pick, it uses detected bounding boxes and visual servoing to move the arm to a position approximately above the target object. Next it uses image gradients to servo the arm to a known position and orientation above the object. Because we can know the arm's position relative to the object, we can reliably collect statistics about the success rate of grasps at specific points on the object.

To infer grasp points, we create a depth map of the object using the robot's IR sensor. (This process would be much faster with an RGB-D sensor such as the Kinect, but the advantage of our approach is that it can be used with a stock Baxter with no additional sensing.) We collect many measurements for the pointcloud by performing an overhead raster scan of the object. The IR sensor reports at about 30Hz but gives fairly good localization. In order to use the good resolution despite the sparsity of measurement induced by the frequency of the sensor and the speed of the robot's movement, we employ a Parzen kernel density estimator to record z measurements at 1mm (x, y) resolution. We then downsample to a 21×21 grid at 1cm resolution, which yields a clean depth map and a tractable state space for grasp inference. We consider 4 gripper orientations at each (x, y) location in the scan for a total of 1764 possible grasps. Our prior for grasp success is a map $\pi : (x, y, \theta) \rightarrow \mathbb{R}$ generated by convolving the 1cm depth map with 4 linear filters which correspond to the 4 orientations we consider. The responses are scaled linearly to be between 0 and 1 inclusive.

This pipeline works well on some objects, but often fails for a variety of reasons. For example, our grasp model is not able to reliably infer grasps for objects that are not visible in IR, such as transparent or very dark objects. Other objects need to be grasped in particular locations to avoid getting stuck on the gripper due to the compliance of the object and the dynamics of the robot's gripper. The next section defines how we adapt this grasping pipeline by learning to identify good grasps from experience.

3 Bandit-based Adaptation

The formalization we contribute treats grasp learning as an N-armed bandit problem. Because the problem domain is rooted in the physical world, we can obtain high-quality supervision at the cost of time and additional sensing by trying grasps with the robot. Formally, the agent is given an N-armed bandit, where each arm pays out 1 with probability μ_i and 0 otherwise. The agent's goal is to identify a good arm (with payout $>= k$) with probability c (e.g., 95% confidence that this arm is good) as quickly as possible. As soon as it has done this, it should terminate. The agent is also given a prior π on the arms so that it may make informed decision about which grasps to explore.

Our contributed algorithm, Prior Confidence Bound, iteratively chooses the arm with the highest observed (or prior) success rate but whose probability of being below k is less than c. It then tries that arm, records the results, and checks to see if its probability of success is above k or within $[k - \epsilon, k + \epsilon]$ with probability c (in which case it terminates). If the latter condition is not included, an arm with success probability equal to k will continue to be pulled indefinitely. Pseudo-code appears in Algorithm 1. To compute this probability we need to estimate the probability that the true payout probability, μ is greater than the threshold, c given the observed number of successes and failures:

$$\Pr(\mu_i > c|S, F) \tag{1}$$

We can compute this probability using the law of total probability:

$$\Pr(\mu_i > c|S, F) = \int_k^1 \Pr(\mu_i = \mu|S, F)d\mu \tag{2}$$

We assume a beta distribution on μ:

$$= \int_k^1 \mu^S(1-\mu)^F d\mu \tag{3}$$

This integral is the CDF of the beta distribution, and is called the regularized incomplete beta function [24]. Note that if $\mu = k$ the run time is unbounded, so we fix an $\epsilon > 0$ and accept a grasp if $\mu \in [k - \epsilon, k + \epsilon]$ with probability c.

`PriorConfidenceBound`(π, k, δ_{accept}, δ_{reject}, $maxTries$)
Initialize $S_0 \dots S_n$ to $\pi(0) \dots \pi(n)$
Initialize $F_0 \dots F_n$ to 0
$totalTries \leftarrow 0$
while $true$ **do**
  $totalTries \leftarrow totalTries + 1$
  Set $M_0 \dots M_n$ to $\frac{S_0}{S_0+F_0} \dots \frac{S_n}{S_n+F_n}$
  $j \leftarrow bestValidArm$; `// set j to the arm with` $p_{below} < \delta_{reject}$ `that has the highest marginal value`
  $r \leftarrow sample(arm_j)$
  if $r = 1$ **then**
    $S_j \leftarrow S_j + 1$
  else
    $F_j \leftarrow F_j + 1$
  end
  $p_{below} \leftarrow \int_0^k \Pr(\mu_j = \mu | S_j, F_j) d\mu$
  $p_{above} \leftarrow \int_k^1 \Pr(\mu_j = \mu | S_j, F_j) d\mu$
  $p_{threshold} \leftarrow \int_{k-\epsilon}^{k+\epsilon} \Pr(\mu_j = \mu | S_j, F_j) d\mu$
  if $p_{above} \geq \delta_{accept}$ **then**
    return j ; `// accept this arm`
  else if $p_{threshold} \geq \delta_{accept}$ **then**
    return j ; `// accept this arm`
  else if $totalTries > maxTries$ **then**
    return $maxI$; `// return the arm with the best marginal value out of those that were tried`
  else
    pass ; `// keep trying`
  end
end

Algorithm 1: Prior Confidence Bound for Best Arm Identification

4 Evaluation

The aim of our evaluation is to assess the ability of the system to acquire visual models of objects which are effective for grasping and object detection. We first assess our approach in simulation to allow comparison to baselines, which would be prohibitively time-consuming to run on the real robot. Next we describe our robotic evaluation, which assesses our system's ability to adaptively improve its ability to grasp objects, end-to-end.

4.1 Simulation

We simulate picking performance by creating a sequence of 50 bandits, where each arm pays out at a rate uniformly sampled between 0 and 1. For algorithms that incorporate prior knowledge, we sample a vector of estimates for each μ_i from a beta distribution with $\alpha = \beta = 1 + e * \mu_i$ where e controls the entropy of the sampling distribution.

Figure 3: The objects used in our evaluation.

To compare to a well-known baseline, we assess the performance of Thompson Sampling [31] in the fixed budget setting, although this algorithm minimizes total regret, including regret during training, rather than simple regret. Second, we compare to a Uniform baseline that pulls every arm equally until the budget is exceeded. This baseline corresponds to the initialization step in UCB or the confidence bound algorithms in Chen et al. [6]. If we implemented the state-of-the-art CLUCB algorithm from Chen et al. [6], it would not have enough pulls to finish this initialization step in our setting. Finally, we show the performance of three versions of Prior Confidence Bound, one with an uninformed prior ($e = 0$, corresponding to Hoeffding races [20]), one quite noisy with $e = 1$(but still informative), the other less noisy $e = 5$).

We report for each point after 100 trials, and report 95% confidence intervals around the algorithm's simple regret. For Thompson Sampling and Uniform, which always use all trials in their budget, we report performance at each budget level; for Prior Confidence Bound, we report the mean number of trials the algorithm took before halting, also at 95% confidence intervals.

Results appear in Figure 4. Thompson Sampling always uses all trials in its budget and improves performance over time. The Uniform method fails to find the optimal arm because there is not enough information when pulling each arm once. All variants of Prior Confidence Bound outperform these baselines, but as more prior information is incorporated, regret decreases. Even with a completely uninformed prior taking a confidence-based approach improves performance over Thompson sampling or a uniform baseline, but the approach realizes significant further improvement with more prior knowledge.

4.2 Robotic Evaluation

We have implemented our approach on the Baxter robot. The robot acquired visual and RGB-D models for 30 objects using our autonomous learning system. The objects used in our evaluation appear in Figure 3. We manually verified that the scans were accurate, and set the following parameters: height above the object for the IR scan (to approximately 2cm); this height could be acquired automatically by doing a first coarse IR scan following by a second IR scan 2cm above the tallest height, but we set it manually to save time. Additionally we set the height of the arm for the initial servo to acquire the

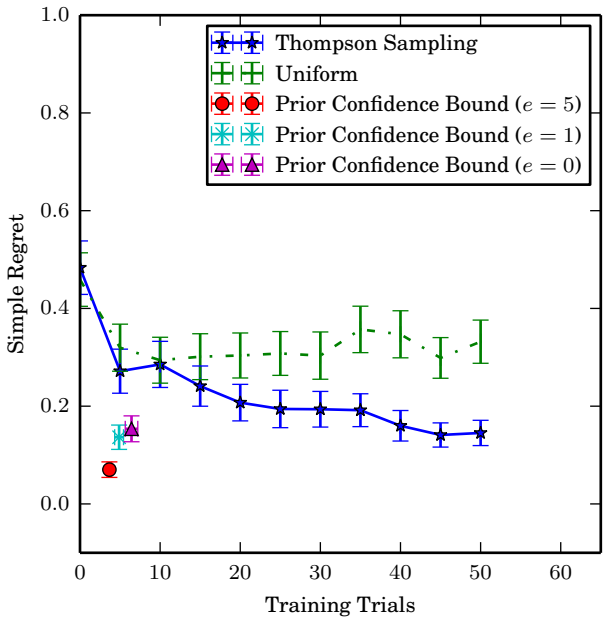


Figure 4: Results comparing our approach to various baselines in simulation.

object. After acquiring visual and IR models for the object at different poses of the arm, the robot performed the bandit-based adaptation step using Algorithm 1.

We report the performance of the robot at picking using the learned height for servoing, but without grasp learning, then the number of trials used for grasp learning by our algorithm, and finally the performance at picking using the learned grasp location.

After the robot detects an initially successful grab, it shakes the object vigorously to ensure that it would not fall out during transport. After releasing the object and moving away, the robot checks to make sure the object is not stuck in its gripper. If the object falls out during shaking or does not release properly, the grasp is recorded as a failure. If the object is stuck, the robot pauses and requests assistance before proceeding.

Most objects have more than one pose in which they can stand upright on the table. If the robot knocks over an object, the model taken in the reference pose is no longer meaningful. Thus, during training, we monitored the object and returned it to the reference pose whenever the robot knocked it over. In the future, we aim to incorporate multiple components in the models which will allow the robot to cope with objects whose pose can change during training.

Our algorithm used an accept threshold of 0.7, reject confidence of 0.95 and epsilon of 0.2. These parameters result in a policy that rejects a grasp after one failed try, and accepts if the first three picks are successful. Different observations of success and failure will cause the algorithm to try the grasp more to determine the true probability of success.

	Prior	Training	Marginal
$\Delta = 0; training = 3$			
Brush	10/10	3/3	10/10
Packing Tape	9/10	3/3	9/10
Purple Marker	9/10	3/3	9/10
Red Bowl	10/10	3/3	10/10
Red Bucket	5/10	3/3	5/10
Shoe	10/10	3/3	10/10
Stamp	8/10	3/3	8/10
Whiteout	10/10	3/3	10/10
Wooden Spoon	7/10	3/3	7/10
$\Delta >= 2$			
Big Syringe	1/10	13/50	4/10
Blue Salt Shaker	6/10	5/10	8/10
Bottle Top	0/10	5/17	7/10
Garlic Press	0/10	8/50	2/10
Gyro Bowl	0/10	5/15	3/10
Metal Pitcher	6/10	7/12	10/10
Mug	3/10	3/4	10/10
Round Salt Shaker	1/10	4/16	9/10
Sippy Cup	0/10	6/50	4/10
Triangle Block	0/10	3/13	7/10
Vanilla	5/10	4/5	9/10
Wooden Train	4/10	11/24	8/10
$\Delta <= 1$			
Clear Pitcher	4/10	3/4	4/10
Dragon	8/10	5/6	7/10
Epipen	8/10	4/5	8/10
Helicopter	2/10	8/39	3/10
Icosahedron	7/10	7/21	8/10
Ruler	6/10	5/12	7/10
Syringe	9/10	6/9	10/10
Toy Egg	8/10	4/5	9/10
Yellow Boat	9/10	5/6	9/10
Total	165/300	148/400	224/300
Rate	0.55	0.37	0.75

Table 1: Results from the robotic evaluation of Prior Confidence Bound. We tested on 30 objects. We use Δ to denote the difference in success rate between the grasp recommended by the prior (Prior column) and the grasp learned by the system (Marginal column). The top block contains objects which succeeded on the initial grasp estimate and therefor did not learn a new grasp. Since the grasp did not change we report the results from one round of 10 picks as both the prior and marginal success rate. The middle block contains objects which spent some time learning and improved their performance notably. The bottom block contains objects for which some learning occurred but little or more improvement was seen. A performance drop after learning was seen on one class: Dragon.

4.3 Discussion

After evaluating, we sorted objects into three blocks, shown in Table 1. We define the quantity Δ to be the difference in grasp successes before and after training. The first block shows objects where the learning process kept the same grasp. For these objects, since the prior and marginal grasps are the same, we ran that grasp 10 times and reported it as both the prior and marginal performance, to save time. Note that when ten trials are performed, three consecutive successes do not assure even near perfect performance. Using a low accept threshold and confidence means that sometimes the algorithm will accept a grasp that is actually below the threshold. By increasing the confidence, we could obtain higher certainty that the best grasp was found (including obtaining PAC [32] bounds on performance as in Maron and Moore [20]), but at the cost of significantly more trails.

The second block in the table shows objects that improved performance by more than two pick successes. These objects typically had some feature that prevented our grasping model from working, for example, being difficult to view in IR. For example, the triangular block failed with the prior grasp because the gripper slid over the sloped edges and pinched the block out of its grippers. The robot tried grasps until it found one that targeted the sides that were parallel to the grippers, resulting in a flush grasp, significantly improving accuracy. The round salt shaker , the robot first attempted to grab the round plastic dome, which is infeasible. It tried grasps until it found one on the handle.

Objects such as the round salt shaker and the bottle top are on the edge of tractability for thorough policies such as Thompson sampling. Prior Confidence Bound, on the other hand, rejects arms quickly so as to make these two objects train in relatively short order while bringing even more difficult objects such as the sippy cup and big syringe into the realm of possibility. It would have taken substantially more time and picks for Thompson sampling to reject the long list of bad grasps before finding the good ones.

The garlic press is a geometrically simple object but quite heavy compared to the others. The robot found a few grasps which might have been good for a lighter object, but it frequently shook the press out of its grippers when confirming grasp quality. The big syringe has some good grasps which are detected well by the prior, but due to its poor contrast and transparent tip, orientation servoing was imprecise and the robot was unable to learn well due to poor signal. What improvement did occur was due to finding a grasp which consistently deformed the bulb into a grippable shape regardless of the perceived orientation of the syringe. Similar problems were observed with the clear pitcher and icosahedron.

Some of the objects had several grasps of similar, mediocre quality, which caused the robot to try multiple grasps several times, eventually accepting one of the mediocre grasps by chance. For example, the helicopter, wooden train, and icosahedron.

Objects that failed to improve, shown in the next segment, fall into several categories. For some, performance was already high, so there was not much room to move. A common failure mode for poorly performing objects was failure to accurately determing the position and orientation through visual servoing. If the grasp map cannot be localized accurately, significant noise is introduced because the map does not correspond to the same physical location on the object at each trial. For example, there is only about a $5mm$ difference between the width of the dragon and the width of the gripper; objects such as these would benefit from additional servo iterations to increase localization precision. If we double the number of iterations during fine grained servoing we can more reliably pick it, but this would either introduce another parameter in the system (iterations) or excessively slow down other objects which are more tolerant to error.

The red bucket exhibited interesting behavior. Its gradient profile is rectangular, but at the chosen servo height only two opposing sides are visible in the camera. Due to symmetry, the robot found a good grasp despite the degeneracy in the model. Most of this object's failures were because the object is tall and the robot did not raise its arm up enough before checking if its gripper was empty, which triggered false negatives for grasp release.

5 Related Work

Pick-and-place has been studied since the early days of robotics [4, 16]. Forerunning systems relied on the user to provide models of object and end effector pose for the algorithm, and simply planned a motion for the arm to grasp. Bohg et al. [3] survey data-driven approaches to grasping.

Our approach can be thought of as a pipeline for automatically building an experience database consisting of object models and known good grasps. The system initially uses analytic approaches to generate a grasp hypothesis space which allows it to pick previously unencountered objects. It then uses our bandit-based method to try new grasps and learn instance-based distributions for the grasp experience database. In this way our system achieves the best of both approaches: models for grasping unknown objects can be applied; when they do fail, the system can attempt to recover by trying grasps and adapting itself based on that specific object.

Modern approaches use object recognition systems to estimate pose and object type, then libraries of grasps either annotated or learned from data [27, 9, 22, 7]. These approaches attempt to create systems that can grasp arbitrary objects based on learned visual features or the object's known 3d configuration. Collecting these training sets is an expensive process and is not accessible to the average user in a non-robotics setting. If the system does not work for the user's particular application, there is no easy way for it to adapt or relearn. Our approach, instead, enables the robot to autonomously acquire more information to increase robustness at detecting and manipulating the specific object that is important to the user at the current moment. Other approaches focus on object discovery and manipulation [18, 15, 8, 28]. By formalizing grasp identification as a bandit problem, we are able to leverege existing strategies for inferring the best arm.

Many existing approaches use a simulator to assess planned grasp quality, because assessing grasp quality in simulation is much faster than trying it on the real robot Miller and Allen [21]. In our approach it is more expensive to evalu-

ate grasp rates because it requires actually attempting to pick up the object. However, by assessing grasp quality in situ with the end-to-end system, our approach potentially obtains more accurate grasp statistics during training which it can leverage later at inference time.

Crowd-sourced and web robotics have created large databases of objects and grasps using human supervision on the web [14, 13]. These approaches outperform automatically inferred grasps but still require humans in the loop. Our approach can incorporate human annotations in the form of the prior. If the annotated grasps work well, then the robot will quickly converge and stop sampling; if they are poor grasps, our approach will find better ones.

Nguyen and Kemp [23] learn to manipulate objects such as a light switch or drawer with a similar self-training approach. Our work autonomously learns visual models to detect, pick, and place previously unencountered rigid objects by actively selecting the best grasp point with a bandit based system, rather than acquiring models for the manipulation of articulated objects. We rely on the fixed structure of objects rather than learning how to deal with structure that can change.

Methods for planning in information space [10, 1, 26] have been applied to enable mobile robots to plan trajectories that avoid failures due to inability to accurately estimate positions. Velez et al. [33] created a mobile robot that explores the environment and actively plans paths to acquire views of objects such as doors. However it uses a fixed model of the object being detected rather than updating its model based on the data it has acquired from the environment. Our approach is focused instead on identifyng the best grasp point by actively experimenting with objects in the world.

Other approaches plan grasps under pose uncertainty [29] or collect information from tactile sensors [11] using POMDPs. Platt et al. [25] describe new algorithms for solving POMDPs by tracking the belief state with a high-fidelity particle filter, but using a lower-fidelity representation of belief for planning, and tracking the KL divergence. Hudson et al. [12] used active perception to create a grasping system capable of carrying out a variety of complex tasks. Using feedback is critical for good performance, but the model cannot adapt itself to new objects. Our approach could be used to improve any of these systems by defining a space of parameters to optimize, such as potential grasp points, and then assessing the performance from experience.

We formalize the problem as an N-armed bandit [31] where the robot aims to perform best arm identification [2, 6], or alternatively, to minimize simple regret after a finite exploration period [5]. Audibert and Bubeck [2] explored best arm identification in a fixed budget setting; however a fixed budget approach does not match our problem, because we would like the robot to stop sampling as soon as it has improved performance above a threshold. We take a fixed confidence approach as in Chen et al. [6], but their fixed confidence algorithm begins by pulling each arm once, a prohibativly expensive operation on our robot. Instead our algorithm estimates confidence that one arm is better than another, following Hoeffding races [20] but operating in a confidence threshold setting that incorporates prior information. By incorporating prior information, our approach achieves good performance without being required to pull all the arms.

6 Conclusion

We presented a formalization of the grasping problem as best arm identification in an N-armed bandit. Our bandit problem has 1764 levers, so even if we could pull a lever every 10 seconds it would take around 5 hours to explore all of the arms for a single object. To address this problem, we created a new algorithm, Prior Confidence Bound, which explores promising arm first by exploiting prior knowledge, significantly reducing constant factors.

Our stack gathers feedback from the environment, which it uses to learn models to detect, localize, and manipulate previously unseen objects. In the future, we plan to explore learning parameters for a wider variety of tasks. For instance, different objects can be located and oriented better or worse at different heights. One could learn which heights work well for which objects. Likewise, we use color gradients for localization, but some objects would work better with other quantities. One could learn the appropriate map to use when localizing each object, or even further, the map might also depend upon the robot's current environment.

Our stack currently runs on Baxter, but the requirements are not stringent. In fact, it would be possible to execute all scanning and some training for crane grasps on a modified 3D printer. Furthermore, the parallel electric gripper for Baxter is more difficult to infer grasps for than the gripper on the PR2. Even though the PR2 lacks the IR rangefinder we use, that data could be gathered by a printer converted from a scanner, and the PR2 could perform its own grasp training.

It is clear that the system's accuracy and precision would benefit from the use of more sophisticated imaging equipment such as the Kinect 2. Better and faster point clouds acquisition would allow the use of more precise physical models for grasps. It would also open the way for additional grasp types, such as side and handle grasps.

References

[1] Nikolay Atanasov, Jerome Le Ny, Kostas Daniilidis, and George J Pappas. Information acquisition with sensing robots: Algorithms and error bounds. *arXiv preprint arXiv:1309.5390*, 2013.

[2] Jean-Yves Audibert and Sébastien Bubeck. Best arm identification in multi-armed bandits. In *COLT-23th Conference on Learning Theory-2010*, pages 13–p, 2010.

[3] Jeannette Bohg, Antonio Morales, Tamim Asfour, and Danica Kragic. Data-driven grasp synthesis—a survey. 2013.

[4] Rodney A Brooks. Planning collision-free motions for pick-and-place operations. *The International Journal of Robotics Research*, 2(4):19–44, 1983.

[5] Sébastien Bubeck, Rémi Munos, and Gilles Stoltz. Pure exploration in multi-armed bandits problems. In *Algorithmic Learning Theory*, pages 23–37. Springer, 2009.

[6] Shouyuan Chen, Tian Lin, Irwin King, Michael R Lyu, and Wei Chen. Combinatorial pure exploration of multi-armed bandits. In *Advances in Neural Information Processing Systems*, pages 379–387, 2014.

[7] Matei Ciocarlie, Kaijen Hsiao, Edward Gil Jones, Sachin Chitta, Radu Bogdan Rusu, and Ioan A Şucan. Towards reliable grasping and manipulation in household environments. In *Experimental Robotics*, pages 241–252. Springer, 2014.

[8] Alvaro Collet, Bo Xiong, Corina Gurau, Martial Hebert, and Siddhartha S. Srinivasa. Herbdisc: Towards lifelong robotic object discovery. *The International Journal of Robotics Research*, 2014. doi: 10.1177/0278364914546030.

[9] Corey Goldfeder, Matei Ciocarlie, Hao Dang, and Peter K Allen. The columbia grasp database. In *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*, pages 1710–1716. IEEE, 2009.

[10] Ruijie He, Sam Prentice, and Nicholas Roy. Planning in information space for a quadrotor helicopter in a gps-denied environment. In *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, pages 1814–1820. IEEE, 2008.

[11] Kaijen Hsiao, Leslie Pack Kaelbling, and Tomás Lozano-Pérez. Task-driven tactile exploration. Robotics: Science and Systems Conference, 2010.

[12] Nicolas Hudson, Thomas Howard, Jeremy Ma, Abhinandan Jain, Max Bajracharya, Steven Myint, Calvin Kuo, Larry Matthies, Paul Backes, Paul Hebert, et al. End-to-end dexterous manipulation with deliberate interactive estimation. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 2371–2378. IEEE, 2012.

[13] David Kent and Sonia Chernova. Construction of an object manipulation database from grasp demonstrations. In *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*, pages 3347–3352. IEEE, 2014.

[14] David Kent, Morteza Behrooz, and Sonia Chernova. Crowdsourcing the construction of a 3d object recognition database for robotic grasping. In *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pages 4526–4531. IEEE, 2014.

[15] Dirk Kraft, Renaud Detry, Nicolas Pugeault, Emre Baseski, Frank Guerin, Justus H Piater, and Norbert Kruger. Development of object and grasping knowledge by robot exploration. *Autonomous Mental Development, IEEE Transactions on*, 2 (4):368–383, 2010.

[16] Tomás Lozano-Pérez, Joseph L. Jones, Emmanuel Mazer, and Patrick A. O'Donnell. Task-level planning of pick-and-place robot motions. *IEEE Computer*, 22(3):21–29, 1989.

[17] Ilya Lysenkov, Victor Eruhimov, and Gary Bradski. Recognition and pose estimation of rigid transparent objects with a kinect sensor. *Robotics*, page 273, 2013.

[18] Natalia Lyubova, David Filliat, and Serena Ivaldi. Improving object learning through manipulation and robot self-identification. In *Robotics and Biomimetics (ROBIO), 2013 IEEE International Conference on*, pages 1365–1370. IEEE, 2013.

[19] Shie Mannor and John N Tsitsiklis. The sample complexity of exploration in the multi-armed bandit problem. *The Journal of Machine Learning Research*, 5:623–648, 2004.

[20] Oded Maron and Andrew W Moore. Hoeffding races: Accelerating model selection search for classification and function approximation. *Robotics Institute*, page 263, 1993.

[21] Andrew T Miller and Peter K Allen. Graspit! a versatile simulator for robotic grasping. *Robotics & Automation Magazine, IEEE*, 11(4):110–122, 2004.

[22] Antonio Morales, Eris Chinellato, Andrew H Fagg, and Angel Pasqual del Pobil. Experimental prediction of the performance of grasp tasks from visual features. In *Intelligent Robots and Systems, 2003.(IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on*, volume 4, pages 3423–3428. IEEE, 2003.

[23] Hai Nguyen and Charles C Kemp. Autonomously learning to visually detect where manipulation will succeed. *Autonomous Robots*, 36(1-2):137–152, 2014.

[24] F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark, editors. *NIST Handbook of Mathematical Functions*. Cambridge University Press, New York, NY, 2010.

[25] Robert Platt, Leslie Kaelbling, Tomas Lozano-Perez, and Russ Tedrake. Simultaneous localization and grasping as a belief space control problem. In *International Symposium on Robotics Research*, volume 2, 2011.

[26] Samuel Prentice and Nicholas Roy. The belief roadmap: Efficient planning in belief space by factoring the covariance. *The International Journal of Robotics Research*, 2009.

[27] Ashutosh Saxena, Justin Driemeyer, and Andrew Y Ng. Robotic grasping of novel objects using vision. *The International Journal of Robotics Research*, 27(2):157–173, 2008.

[28] David Schiebener, Julian Schill, and Tamim Asfour. Discovery, segmentation and reactive grasping of unknown objects. In *Humanoids*, pages 71–77, 2012.

[29] Freek Stulp, Evangelos Theodorou, Jonas Buchli, and Stefan Schaal. Learning to grasp under uncertainty. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 5703–5708. IEEE, 2011.

[30] Stefanie Tellex, Ross Knepper, Adrian Li, Daniela Rus, and Nicholas Roy. Asking for help using inverse semantics. In *Robotics: Science and Systems (RSS)*, 2014.

[31] William R Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, pages 285–294, 1933.

[32] Leslie G Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.

[33] Javier Velez, Garrett Hemann, Albert S Huang, Ingmar Posner, and Nicholas Roy. Active exploration for robust object detection. In *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, volume 22, page 2752, 2011.