Bandit-Based System Adaptation for Grasping

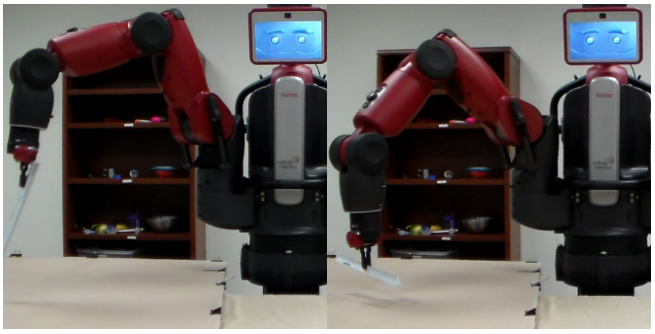
Abstract

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. The robot performs best arm identification in the N-armed bandit problem by using a policy that computes confidence bounds while incorporating prior information, enabling it 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 [38]. 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 [28].

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 [44] where the robot aims to minimize simple regret after a finite exploration period [8]. 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.



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

Figure 1: Before learning, the robot grasps the ruler near the end, and it twists out of its gripper and falls onto the table when it lifts; after learning, the robot knows to grasp it near the center of mass.

Existing algorithms for best arm identification require pulling all the arms as an initialization step [30, 2, 11], 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 two new algorithms: Ordered Confidence Bound, based on Hoeffding races [31], and Reckless Ordered Confidence Bound, an approximation of Ordered Confidence Bound that explores faster and performs similarly empirically. In our approaches, 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.

Although Ordered Confidence Bound is more appealing from a theoretical point of view, it explores significantly more slowly than Reckless Ordered Confidence Bound. We evalu-

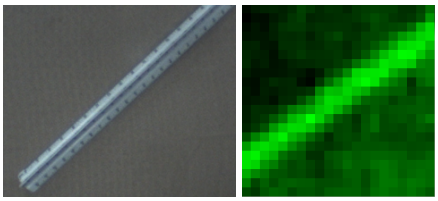


Figure 2: Automatically acquired RGB image for one of the objects in our training set, combined with the IR raster scan.

ated Reckless Ordered Confidence Bound on a Baxter robot, demonstrating that our adaptation step improves the overall 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 [43].

We first give an overview of our object detection and localization, then 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 able to detect, localize, and grasp the specific instances for which it has models with much higher speed and accuracy.

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 $(x, y, \theta) \rightarrow \mathbb{R}$ generated by convolving the 1cm depth map with 4 linear filters which correspond to the 4 orientations we consider. Their 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

Our model 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. Ordered Confidence Bound only makes use of order of the arms as ranked by the prior, trying promising ones first and ignoring the specific values of π. Reckless Ordered Confidence Bound, on the other hand, actually makes use of the probabilities specified by the prior in order to explore more aggressively.

Our first algorithm, Ordered Confidence Bound, iterates through each arm, and tries it until it has identified that it either is above k with probability c (in which case it terminates) or it is below k with probability c (in which case it moves to the next arm in the sequence. 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:

$$= \int_k^1 \mu^S (1-\mu)^F d\mu \quad (3)$$

This integral is the CDF of the beta distribution, and is called the regularized incomplete beta function [34]. 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.

`OrderedConfidenceBound(`$armOrder$, k, δ_{accept}, δ_{reject}`)`
Initialize $S_0 \dots S_n$ to 0
Initialize $F_0 \dots F_n$ to 0
for $i \in armOrder$ **do**
  $r \leftarrow sample(arm_i)$
  if $r = 1$ **then**
    $S_i \leftarrow S_i + 1$
  else
    $F_i \leftarrow F_i + 1$
  $p_{below} \leftarrow \int_0^k \Pr(\mu_i = \mu | S_i, F_i) d\mu$
  $p_{above} \leftarrow \int_k^1 \Pr(\mu_i = \mu | S_i, F_i) d\mu$
  $p_{threshold} \leftarrow \int_{k-\epsilon}^{k+\epsilon} \Pr(\mu_i = \mu | S_i, F_i) d\mu$
  if $p_{above} \geq \delta_{accept}$ **then**
    return i ; `// accept this arm`
  else if $p_{threshold} \geq \delta_{accept}$ **then**
    return i ; `// accept this arm`
  else if $p_{below} \geq \delta_{reject}$ **then**
    break ; `// go to the next arm`
  else
    pass ; `// keep pulling this arm`
  end
end

Algorithm 1: Ordered Confidence Bound for Best Arm Identification

Our second algorithm, Reckless Ordered Confidence Bound, approximates Ordered Confidence Bound using heuristics on prior for untried grasps and the observed success probabilities for grasps that have been attempted.

JGO: Here an algorithm figure for Reckless Ordered Confidence Bound, and also a side by side policy map for Ordered Confidence Bound and Reckless Ordered Confidence Bound restricted to one arm.

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, comparing both algorithms to Thompson sampling as a baseline. Next we describe our robotic evaluation, which assesses our system's ability to learn and adaptively improve its ability to grasp objects, end-to-end.

4.1 Simulation

Our simulated results compare our two algorithms to Thompson sampling, assessing the trade-offs inherent in our choice of parameters and algorithms. We present results in

`RecklessOrderedConfidenceBound(`π, k, δ_{accept}, δ_{reject}, $maxTries$`)`
for $i \in 0...n$ **do**
  $S_i \leftarrow \pi(i)$
end
Initialize $F_0 \dots F_n$ to 0
Initialize $M_0 \dots M_n$ to 0
$totalTries \leftarrow 0$
while $true$ **do**
  $totalTries \leftarrow totalTries + 1$
  for $i \in 0...n$ **do**
    $M_i \leftarrow S_i/(S_i + F_i)$
  end
  $maxI \leftarrow -1$
  $maxIval \leftarrow -1$
  for $i \in 0...n$ **do**
    $p^i_{below} \leftarrow \int_0^k \Pr(\mu_i = \mu | S_i, F_i) d\mu$
    if $p_{below} < \delta_{reject}$ *and* $M_i > maxIval$ **then**
      $maxI \leftarrow i$
      $maxIval \leftarrow M_i$
    end
  end
  $j \leftarrow maxI$
  $r \leftarrow sample(arm_j)$
  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**
    for $i \in 0...n$ **do**
      $M_i \leftarrow S_i/(S_i + F_i)$
    end
    $maxI \leftarrow -1$
    $maxIval \leftarrow -1$
    for $i \in 0...n$ **do**
      if $M_i > maxIval$ **then**
        $maxI \leftarrow i$
        $maxIval \leftarrow M_i$
      end
    end
    return $maxI$; `// return the best arm encountered`
  else
    pass ; `// keep trying`
  end
end

Algorithm 2: Reckless Ordered Confidence Bound for Best Arm Identification

Figure 3: The objects used in our evaluation.

JGO: This should be updated for whatever we decide to do. Here and previously we need to make sure to refer to algorithms because now we have two.

Table 4. We simulate picking performance by creating a sequence of 20 bandits, where each arm pays out at a rate of 0.1 except for one, which pays out at 0.9. We move the location of the best arm to a uniformly sampled random position in the sequence. Thompson Sampling always uses all trials in its budget and, given enough trials, reliably finds the optimal arm. We present two versions of Ordered Confidence Bound with different parameters. The tight bound uses a confidence interval defined by the union bound to decide when to move on; this results in the agent pulling each arm many times before achieving a high confidence. It almost always terminates with the optimal arm, but takes many trials to do it. In contrast, the parameter settings used in this paper with a lower confidence bound takes many fewer trials on average, but sometimes terminates with a non-optimal arm. Because of the high cost of running on the robot, we use less conservative settings in our evaluation.

4.2 Robotic Evaluation

JGO: In this section we use algorithm singular because we only evaluated one of them.

We have implemented our approach on the Baxter robot. It is equipped with a seven-degree-of-freedom arm with a camera and IR depth sensor, which we use as a one-pixel depth camera to acquire our models.

The robot acquired visual and RGB-D models for 30 objects using our autonomous learning system. 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 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 2.

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

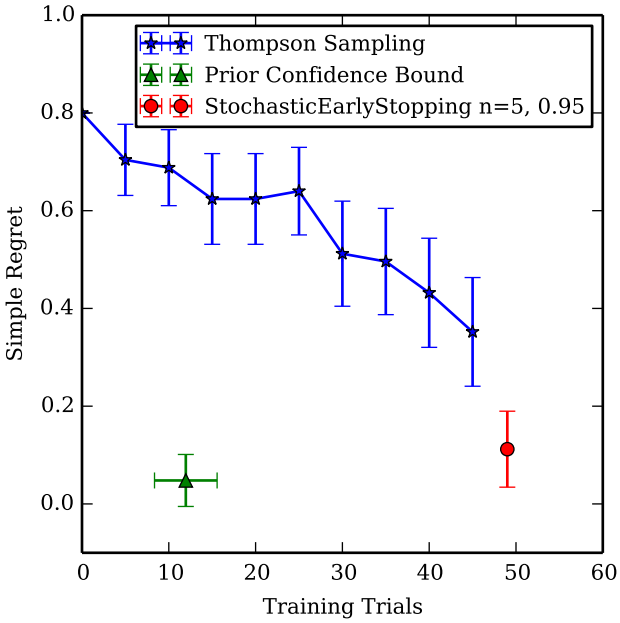


Figure 4: Results comparing our approach to various baselines in simulation.

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. The policy for exploring an arm appears in Figure **??**.

4.3 Discussion

Consider the top block in Figure **??**, the objects for which the initially estimated grasp succeeds 3 times in a row. Note that when 10 trials are performed, it becomes clear that 3 consecutive successes do not assure even near perfect performance. Even though Reckless Ordered Confidence Bound rejects more quickly than Ordered Confidence Bound, they behave identically for these objects and accept faster than Thompson sampling does.

	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 Reckless Ordered 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.

The first few grasps suggested by the prior for the Triangle Block were infeasible 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. For 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 and Ordered Confidence Bound. Reckless Ordered 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 and Ordered Confidence Bound 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. It is likely that Reckless Ordered Confidence Bound rejected good grasps a little too quickly, as a result perhaps even taking longer than Ordered Confidence Bound would have and settling for a worse solution.

The Red Bucket exhibited interesting behavior. Its gradient profile is rectangular, but at the chosen height only two opposing sides are visible. Due to symmetry it was able to find a good grasp despite the degeneracy in the model. Most of its failures were because the object is tall and the robot did not back up enough before checking if the gripper was empty, which triggered false negatives for grasp release. Finally, some objects, such as the Helicopter and Dragon, barely fit in the gripper and would therefor benefit from additional servo iterations to increase localization precision.

5 Related Work

Bohg et al. [5] survey data-driven approaches to grasping. Our approaches can be thought of as a pipeline for automatically building an experience database consisting of object models and known good grasps, using analytic approaches to grasping unknown objects to generate a grasp hypothesis space and bandit-based methods for trying grasps and learning 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 ap-

plied; when they do fail, the system can automatically recover by trying grasps and adapting itself based on that specific object. Additionally, we can use other grasp detectors to seed our prior, since it is only based on an ordering and does not require an explicit probability distribution. We can also interface with manual labeling similar to the forklift demarcation: circle an area and we will restrict exploration to that region.

Ude et al. [45] described an approach for detecting and manipulating objects to learn models. It uses a bag of words model and learns to detect the objects. It does not learn a model for grasping. Schiebener et al. [40] describes an extension that also does model learning. The robot pushes the object and then trains an object recognition system. It does not use a camera that moves and does not grasp. Schiebener et al. [39] discovers and grasps unknown objects.

Summary:

- People doing SLAM. Wang et al. [47], Gallagher et al. [14],
- People doing 3d reconstruction. Krainin et al. [23], Banta et al. [3]
- People doing big databases for category recognition. Kent et al. [21], Kent and Chernova [20], Lai et al. [26], Goldfeder et al. [15]
- Object tracking in vision (typically surveillance).
- POMDPs for grasping. Platt et al. [35], Hsiao et al. [17]
- People doing systems. Hudson et al. [18], Ciocarlie et al. [12]

Crowd-sourced and web robotics have created large databases of objects and grasps using human supervision on the web [21, 20]. These approaches outperform automatically inferred grasps but still require humans in the loop. Our approach enables a robot to acquire a model fully autonomously, once the object has been placed on the table.

Zhu et al. [48] created a system for detecting objects and estimating pose from single images of cluttered objects. They use KinectFusion to construct 3d object models from depth measurements with a turn-table rather than automatically acquiring models.

Chang et al. [9] created a system for picking out objects from a pile for sorting and arranging but did not learn object models.

next-best view planning [24]

Nguyen and Kemp [33] learn to manipulate objects such as a light switch or drawer with a similar self-training approach. Our work learns visual models for objects for autonomous pick-and-place rather than to manipulate objects.

Developmental/cognitive robotics [29**?**]

Banta et al. [3] constructs a prototype 3d model from a minimum number of range images of the object. It terminates reconstruction when it reaches a minimum threshold of accuracy. It uses methods based on the occluded regions of the reconstructed surface to decide where to place the camera and evaluates based on the reconstruction rather than pick up success. Krainin et al. [23] present an approach for autonomous object modeling using a depth camera observing the robot's hand as it moves the object. This system provides a 3d construction of the object autonomously. Our approach uses vision-based features and evaluates based on grasp success. Eye-in-hand laser sensor. [**?**]

ST: Need to find the instance-based work that Erik mentioned when he said it was a "solved problem."

Velez et al. [46] 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.

Methods for planning in information space [16, 1, 36] have been applied to enable mobile robots to plan trajectories that avoid failures due to inability to accurately estimate positions. Our approach is focused instead on object detection and manipulation, actively acquiring data for use later in localizing and picking up objects. **ST: May need to say more here depending on what GRATA actually is.**

Early models for pick-and-place rely on has been studied since the early days of robotics [7, 27]. These systems relied on models of object pose and end effector pose being provided to the algorithm, and simply planned a motion for the arm to grasp. Modern approaches use object recognition systems to estimate pose and object type, then libraries of grasps either annotated or learned from data [38, 15, 32]. These approaches attempt to create systems that can grasp arbitrary objects based on learned visual features or 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.

Visual-servoing based methods [10] **ST: Need a whole paragraph about that.**

ST: Ciocarlie et al. [12] seems highly relevant, could not read from the train's wifi. Existing work has collected large database of object models for pose estimation, typically curated by an expert [25]. Kasper et al. [19] created a semiautomatic system that fuses 2d and 3d data, but the setup requires a special rig including a turntable and a pair of cameras. Our approach requires an active camera mounted on a robot arm, but no additional equipment, so that a robot in the home can autonomously acquire new models.

?] describes an approach for lifelong robotic object discovery, which infers object candidates from the robot's perceptual data. This system does not learn grasping models and does not actively acquire more data to recognize, localize, and grasp the object with high reliability. It could be used as a first-pass to our system, after which the robot uses an active method to acquire additional data enabling it to grasp the object. Approaches that integrate SLAM and moving object tracking estimate pose of objects over time but have not been extended to manipulation [47, 14, 37, 41].

Our approach is similar to the philosophy adopted by Rethink Robotic's Baxter robot, and indeed, we use Baxter as our test platform [13]. **ST: Haven't actually read this paper, just making stuff up based on Rod's talks. Should**

read the paper and confirm. Baxter's manufacturing platform is designed to be easily learned and trained by workers on the factory floor. The difference between this system and our approach is we rely on the robot to autonomously collect the training information it needs to grasp the object, rather than requiring this training information to be provided by the user.

Robot systems for cooking [6, 4] or furniture assembly [22] use many simplifying assumptions, including pre-trained object locations or using VICON to solve the perceptual system. We envision vision or RGB-D based sensors mounted on the robot, so that a person can train a robot to recognize and manipulate objects wherever the robot finds itself.

Approaches to plan grasps under pose uncertainty [42] or collect information from tacticle sensors [17] using POMDPs. **?**] describe new algorithms for solving POMDPs by tracking 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. [18] 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.

6 Conclusion

We presented a software stack for autonomously acquiring instance based models of objects for the Baxter Research Robot. Conventional MAB algorithms value reduced run time complexity over small constants because large numbers of trials can be tolerated. Our MAB 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. We provided the Ordered Confidence Bound algorithm for solving pure exploration bandits with low constant factors by exploiting prior knowledge. This allows us to obtain an approximately optimal solution without pulling too many arms.

Our stack gathers feedback from the environment, which it uses to learn models to detect, localize, and manipulate previously unseen objects. The platform and example we provide open the door for more extensive investigation in the same direction and suggest that it might be possible to similarly learn parameters for a 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.

Objects which failed entirely did so because of reasonable limitations of the system induced by compromises we made for global compatibility. For instance, the small Vaseline container only has 3mm of play between the container and the gripper in the only feasible grasp location. If we double the number of iterations during gradient 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.

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] Joseph E Banta, LR Wong, Christophe Dumont, and Mongi A Abidi. A next-best-view system for autonomous 3-d object reconstruction. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 30(5):589–598, 2000.

[4] Michael Beetz, Ulrich Klank, Ingo Kresse, Alexis Maldonado, L Mosenlechner, Dejan Pangercic, Thomas Ruhr, and Moritz Tenorth. Robotic roommates making pancakes. In *Humanoid Robots (Humanoids), 2011 11th IEEE-RAS International Conference on*, pages 529–536. IEEE, 2011.

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

[6] Mario Bollini, Stefanie Tellex, Tyler Thompson, Nicholas Roy, and Daniela Rus. Interpreting and executing recipes with a cooking robot. In *Proceedings of International Symposium on Experimental Robotics (ISER)*, 2012.

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

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

[9] Lillian Chang, Joshua R Smith, and Dieter Fox. Interactive singulation of objects from a pile. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 3875–3882. IEEE, 2012.

[10] François Chaumette and Seth Hutchinson. Visual servo control. i. basic approaches. *Robotics & Automation Magazine, IEEE*, 13(4):82–90, 2006.

[11] 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.

[12] 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.

[13] C Fitzgerald. Developing baxter. In *Technologies for Practical Robot Applications (TePRA), 2013 IEEE International Conference on*, pages 1–6. IEEE, 2013.

[14] Garratt Gallagher, Siddhartha S Srinivasa, J Andrew Bagnell, and Dave Ferguson. Gatmo: A generalized approach to tracking movable objects. In *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*, pages 2043–2048. IEEE, 2009.

[15] 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.

[16] 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.

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

[18] 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.

[19] Alexander Kasper, Zhixing Xue, and Rüdiger Dillmann. The kit object models database: An object model database for object recognition, localization and manipulation in service robotics. *The International Journal of Robotics Research*, 31(8):927–934, 2012.

[20] 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.

[21] 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.

[22] Ross A. Knepper, Stefanie Tellex, Adrian Li, Nicholas Roy, and Daniela Rus. Single assembly robot in search of human partner: Versatile grounded language generation. In *Proceedings of the HRI 2013 Workshop on Collaborative Manipulation*, 2013.

[23] Michael Krainin, Peter Henry, Xiaofeng Ren, and Dieter Fox. Manipulator and object tracking for in-hand 3d object modeling. *The International Journal of Robotics Research*, 30(11):1311–1327, 2011.

[24] Simon Kriegel, T Bodenmlller, Michael Suppa, and Gerd Hirzinger. A surface-based next-best-view approach for automated 3d model completion of unknown objects. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 4869–4874. IEEE, 2011.

[25] Kevin Lai, Liefeng Bo, Xiaofeng Ren, and Dieter Fox. A large-scale hierarchical multi-view rgb-d object dataset. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 1817–1824. IEEE, 2011.

[26] Kevin Lai, Liefeng Bo, Xiaofeng Ren, and Dieter Fox. A scalable tree-based approach for joint object and pose recognition. In *Twenty-Fifth Conference on Artificial Intelligence (AAAI)*, August 2011.

[27] 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.

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

[29] 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.

[30] 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.

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

[32] 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.

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

[34] 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. Print companion to [**?**].

[35] 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.

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

[37] Renato F Salas-Moreno, Richard A Newcombe, Hauke Strasdat, Paul HJ Kelly, and Andrew J Davison. Slam++: Simultaneous localisation and mapping at the level of objects. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 1352–1359. IEEE, 2013.

[38] 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.

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

[40] David Schiebener, Jun Morimoto, Tamim Asfour, and Aleš Ude. Integrating visual perception and manipulation for autonomous learning of object representations. *Adaptive Behavior*, 21(5): 328–345, 2013.

[41] Antonio HP Selvatici and Anna HR Costa. Object-based visual slam: How object identity informs geometry. 2008.

[42] 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.

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

[44] 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.

[45] Ales Ude, David Schiebener, Norikazu Sugimoto, and Jun Morimoto. Integrating surface-based hypotheses and manipulation for autonomous segmentation and learning of object representations. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 1709–1715. IEEE, 2012.

[46] 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.

[47] Chieh-Chih Wang, Charles Thorpe, Sebastian Thrun, Martial Hebert, and Hugh Durrant-Whyte. Simultaneous localization, mapping and moving object tracking. *The International Journal of Robotics Research*, 26(9):889–916, 2007.

[48] M. Zhu, N. Atanasov, G. Pappas, and K. Daniilidis. Active Deformable Part Models Inference. In *European Conference on Computer Vision (ECCV)*, 2014.