Learning to Pick Up Objects Through Active Exploration

John Oberlin and Stefanie Tellex
Computer Science Department, Brown University

I. 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. Transparent or reflective surfaces that are not visible in IR or RGB make it difficult to infer grasp points [4]. 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.

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 [8] where the robot aims to minimize simple regret after a finite exploration period [2]. 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.

Existing algorithms for best arm identification require pulling all the arms as an initialization step [5, 1, 3]; in the case of identifying grasp points, where each grasp takes more than 90 seconds and there are more than 1000 potential arms, this is a prohibitive expense. To address this problem, we present a new algorithm, Prior Confidence Bound, based on Hoeffding races [6]. 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 pick success rate from 55% to 75% on our test set of 30 household objects, shown in Figure 1(c). 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 [7].

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 **??** describes our evaluation in simulation and on the real robot with 30 household objects, Section **??** covers related work, and Section **??** concludes.

II. 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

(a) Before learning. (b) After learning. (c) Objects in our evaluation, which improves grasping performance from 55% (before learning) to 75% (after learning).

Fig. 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.

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.

Instance-based approaches mean that the robot must acquire data about a specific object before manipulating it; however by adapting itself to the object it can obtain higher accuracy. We aim to aggregate data collected via our instance-based system to create a new data set of images, RGB-D, and grasp success rates at various poses to provide supervision for general-purpose category models for grasping. Our long term vision is a system that can infer high quality grasps for any object, but adaptively recover if the initial grasp attempt is unsuccessful, leading overall to robust and accurate pick-and-place.

REFERENCES

[1] 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.

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

[3] 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.

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

[5] 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.

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

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

[8] 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.