Learning to Pick Up Objects Through Active Exploration

John Oberlin and Stefanie Tellex
Computer Science Department, Brown University

Robots need to perceive and manipulate objects in its environment, yet robust object manipulation remains a challenging problem. Many aspects of a perception and manipulation system need to be customized for a particular object and environment, such as where to grasp an object, what algorithm to use for segmentation, and what height to visually servo above an object on the table. To address these limitations, we propose an approach for enabling a robot to learn about objects through active exploration and adapt its grasping model accordingly. We frame the problem of model adaptation as a bandit problem, specifically the identification of the best of the arms of an N-armed bandit, [5] where the robot aims to minimize simple regret after a finite exploration period [1]. 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. This paper provides an overview of our previous work [3] using this approach to actively infer grasp points and adds a description of our efforts learning the height to servo to an object.

We use our robot's eye-in-hand camera to visually servo to an object and identify it; however, objects have different sizes. Some objects are large, and need to be observed from far away so that the entire object fits in frame, but not too far because the precision of the alignment process falls with the distance from the object as pixels in the image span larger physical distances. On the other hand, small objects should be observed more closely to obtain a higher-resolution image and finer-grained pose estimation, but not too closely because distortions in the transformations between camera and physical coordinates become more dramatic when the object is nearer the lens.

Our previous work [3] presented a new algorithm, Prior Confidence Bound, based on Hoeffding races [2]. 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 exploring by bounding the confidence in the result. Figure 1(a) shows the robot's performance on two differently sized objects; after training it servos at a higher height for the epipen, which is larger, and a lower height for the egg, improving the success rate. The contribution of this short paper is to apply our algorithm to learn what height to observe an object.

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 decisions about which arm to explore.

Our approach 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.) 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: $\Pr(\mu_i > c|S, F) = 1 - \int_0^k \Pr(\mu_i = \mu|S, F)d\mu$, assuming a beta distribution on μ: $\int_k^1 \mu^S(1-\mu)^F d\mu$. This integral is the CDF of the beta distribution, and is called the regularized incomplete beta function [4].

In our previous work, the space of possible arms was defined by different grasp points on the object. In this paper, we optimize an additional parameter from experience: the height above the table. Each arm is the height above the table, and reward is defined by the ability to servo twice and end up at the same x,y location, when the object hasn't moved. If the robot can servo consistently to the same location on the object, this signal is an indication that it is a good height.

Evaluation

We 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 1(b). After acquiring visual and IR models for the object at different poses of the arm and camera, the robot performed the bandit-based adaptation step. After the robot detects an initially successful grasp, 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

(a) The robot learns to servo for the epipen at a higher height, so the entire object fits in frame; it servos the egg at a lower height to obtain higher positional accuracy.

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

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

(c) Quantitative results.

Fig. 1. Results from our evaluation.

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 evaluation demonstrates 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(b), after doing pick and height training. Height servoing allows the robot to adjust to different heights as shown in Figure 1(a). Video showing our approach and evaluation can be seen at https://www.youtube.com/watch?v=xfH0B3g782Y.

Conclusion

Our robotic system gathers feedback from the environment, which it uses to learn models to detect, localize, and manipulate previously unseen objects. This paper demonstrates that our approach can learn to servo at different heights for different objects. In the future, we plan to explore learning parameters for a wider variety of tasks. For instance, 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.

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

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

[3] John Oberlin and Stefanie Tellex. Bandit-based adaptation for robotic grasping. In *IJCAI*, 2015. Under review.

[4] 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.

[5] 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.