In *Proceedings of 2017 AAAI Spring Symposium on Interactive Multi-Sensory Perception for Embodied Agents, 2017*

Robot Behavioral Exploration and Multimodal Perception using POMDPs

Shiqi Zhang[1,2], Jivko Sinapov[2], Suhua Wei[1], and Peter Stone[2]
[1] Department of Electrical Engineering and Computer Science, Cleveland State University
[2] Department of Computer Science, The University of Texas at Austin

1 Introduction

Service robots are increasingly present in everyday environments, such as homes, offices, airports and hospitals. A common task for such robots involves retrieving an object for a user. Consider the request, "*Robot, please fetch me the red empty bottle*". A key problem for the robot consists of deciding whether a particular candidate object matches the properties in the query. For certain words (e.g., *heavy*, *soft*, etc.) visual classification of the object is insufficient as the robot would need to perform an action (e.g., *lift the object*) to determine whether it is empty or not. Furthermore, the robot would need to decide which actions (possibly out of many) to perform on an object, i.e., it would need to generate a behavioral policy for a given request.

Recent research in robotics has shown that robots can learn to classify objects using computer vision methods as well as non-visual perception coupled with actions performed on the objects (Högman, Björkman, and Kragic 2013; Sinapov et al. 2014; Thomason et al. 2016). For example, a robot can learn to determine whether a container is full based on the sounds produced when shaking the container (Sinapov and Stoytchev 2009); or learn whether an object is soft or hard based on the haptic sensations produced when pressing it (Chu et al. 2015). Nevertheless, there has been relatively little emphasis on enabling a robot to efficiently select actions at test time when it is tasked with classifying a new object. The few approaches for tackling action selection, e.g., (Rebguns, Ford, and Fasel 2011; Fishel and Loeb 2012), assume that only one target property needs to be identified (e.g., the object's identity in the case of object recognition) and would not scale to requests such as the one presented earlier.

To address this limitation, we propose to generate behavioral exploration policies for a given request using the partially observable Markov decision process (POMDP) formalism. POMDP (Kaelbling, Littman, and Cassandra 1998) is a general framework that does not have the assumption of full observability over current state, so an agent needs to use its local, unreliable observations to estimate the underlying state and maintains a distribution over possible states. As a result, POMDPs have been used in object exploration in robotics. For instance, hierarchical POMDPs were used for suggesting visual operators for exploring multiple objects on a tabletop scenario (Sridharan, Wyatt, and Dearden

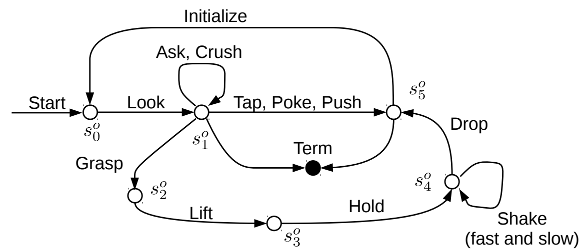

Figure 1: A simplified version of the transition diagram of our POMDP model for object exploration. The transitions led by exploration actions are probabilistic.

2010), and more recent work further used a robotic arm to move objects enabling better visual analysis (Pajarinen and Kyrki 2015). However, the sensing in these research is limited to robot vision and other modalities such as audio and haptics are not used.

Although multimodal perception and POMDP-based object exploration have been studied previously, to the best of our knowledge, there is no research that integrates both in robotics. In this work, given queries about object properties, we dynamically construct POMDPs using a data set collected from a real robot. Experiments on exploring new objects show that our POMDP-based object exploration strategy significantly reduces the overall cost of exploration actions without hurting accuracy, compared to a baseline strategy that uses a predefined sequence of actions.

2 POMDP-based Object Exploration

We construct a POMDP for guiding the robot's exploration behavior based on the robot's sensing and actuating capabilities. A simplified version of the observable aspect of our POMDP's transition diagram is shown in Figure 1, where object properties, as a part of the world state, are not shown. A standard POMDP model is a 7-tuple $(S, A, T, R, \Omega, O, \gamma)$, where γ is the discount factor that represents how much immediate rewards are favored over more distant rewards. In our case, $\gamma = 0.99$, which means the robot has a relatively long horizon in planning.

- $S : S^o \times S^h \cup term$ is the state set. It includes a Cartesian product of sets S^o and S^h, and a terminal state *term*. $s^o \in S^o$ corresponds to one of the non-terminal states $(s_0^o, \cdots, s_5^o)$ in Figure 1. $s^h \in S^h$ is specified by all attributes of a given object: $v_0^p,\ v_1^p,\ \cdots,\ v_{N-1}^p$, where the value of v_i^p is either *true* or *false*. For instance, given a object description "a *red*, *heavy* bottle of *beans*" that includes three properties, S^h will include 8 states. s^o is *fully* observable and s^h is *unreliably* observable, so s^h needs to be estimated through observations. State *term* identifies the end of an episode.

- $A : A^e \cup A^r$ is the action set. A^e includes the object *exploration* actions pulled from the literature of robot exploration, as shown in Figure 1, and $|A^e| = 13$. A^r includes a set of actions that *report* the object's properties and can deterministically lead the state transition to *term*. For $a \in A^r$, we use $s \odot a$ to represent that the report of a matches the underlying values of object properties (i.e., a correct report) and use $s \oslash a$ otherwise.
- $T : S \times A \times S \to [0,1]$ is the state transition function that includes a set of conditional transition probabilities from current state $s \in S$ to next state $s' \in S$ given $a \in A$ being the current action. For instance, $p(s_4^o,\ drop,\ s_5^o) = 0.95$ in our case, indicating there is small probability the object is stuck in robot hand.
- $R : S \times A \to \mathbb{R}$ is the reward function. Each action for object exploration, $a^e \in A^e$, has a cost that is determined by the time required to complete the action. The costs of reporting actions depend on if the report is correct.

$$R(s,a^r) = \begin{cases} r^-, & \textbf{if } s \in S,\ a \in A^r,\ s \oslash a \\ r^+, & \textbf{if } s \in S,\ a \in A^r,\ s \odot a \end{cases}$$

  where r^- is a negative value (penalty) given an incorrect report and r^+ is a big reward given a correct report. Unless otherwise specified, $r^- = -200$ and $r^+ = 100$ in this paper. Costs of other exploration actions are within the range of $[2,10]$ (corresponding reward is negative), except that actions *ask* and *init* have costs of 40 and 20 respectively.
- $\Omega : \Omega^h \cup none$ is a set of observations. Elements in Ω^h include all possible combinations of object properties and have one-one correspondence to elements in A^r and S^h. Actions that produce no information gain (such as *init* and the ones in A^r) will result in a *none* observation.
- $O : S \times A \times \Omega \to [0,1]$ is the observation function that includes a set of conditional observation probabilities. The observation probabilities of actions in A^e are learned from previous experience of object exploration.

We use an approximate, point-based POMDP solver for policy generation (Kurniawati, Hsu, and Lee 2009).

3 Experimental Results

Experiments have been conducted to evaluate our POMDP-based planning strategy for multimodal perception in both accuracy and efficiency. Baseline methods include a *random planner* that suggests an action randomly selected from action set A and a *predefined planner* that suggests actions following a predefined action sequence. Specifically, the predefined action sequence includes all exploration actions (one instance for each action) and the actions are ordered in a way that maximizes information gain. The observation and reward functions of our POMDP are learned from an open-source data set collected in existing research (Sinapov, Schenck, and Stoytchev 2014).

Table 1: Results of multimodal object exploration using POMDP-based and two baseline planners in cost and accuracy.

	Properties	Overall cost	Accuracy
Random	Two	17.56 (30)	0.245
	Three	10.12 (21.77)	0.130
Predefined	Two	37.10 (0.00)	0.583
	Three	37.10 (0.00)	0.373
POMDP	Two	29.85 (12.87)	0.860
	Three	33.87 (8.78)	0.903

For each trial in our experiments, we place an object that has three properties (color, weight and content) on a table and then generate an object description that includes the values of two or three properties. This description matches the object in only half of the trials. The robot needs to take exploration actions (selected by POMDP-based or one of the two baseline planners) to report whether the description is correct or not, while minimizing overall action cost and maximizing report accuracy at the same time.

Preliminary results are reported in Table 1, where each data point corresponds to an average of 100 trials. Not surprisingly, randomly selecting actions produces very low accuracy. The overall cost is smaller in more challenging trials (all three properties are questioned), because in these trials there are relatively less exploration actions, making it more likely to take a reporting action. In the set of experiments using a "predefined" sequence of actions, after executing all exploration actions, the robot selects the reporting action that corresponds to the state with the highest belief. Intuitively, the robot is "forced" to report based on the information collected so far. Finally, our POMDP-based multimodal perception strategy reduces the overall action cost while significantly improving the reporting accuracy.

4 Conclusions and Future Work

In this paper, we investigate using partially observable Markov decision processes (POMDPs) to help robots select actions for multimodal perception in object exploration tasks. Our approach can dynamically construct a POMDP model given an object description from a human user (e.g., "*a blue heavy bottle*"), compute a high-quality policy for this model, and use the policy to guide robot behaviors (such as "look" and "shake") toward maximizing information gain. Experimental results show that our POMDP-based exploration approach enables the robot to identify object properties more accurately without introducing extra cost from exploration actions, compared to a baseline that suggests actions following a predefined action sequence.

In the future, we plan to evaluate our approach using a larger, more recent data set we collected using a real robot (Thomason et al. 2016). Another direction is to better implement the *question-asking* action as a POMDP-based dialog system (Zhang and Stone 2015), and potentially use a single POMDP for both multimodal and language-based perception. Finally, we plan to implement and evaluate this approach on a real mobile robot platform.

References

Chu, V.; McMahon, I.; Riano, L.; McDonald, C. G.; He, Q.; Perez-Tejada, J. M.; Arrigo, M.; Darrell, T.; and Kuchenbecker, K. J. 2015. Robotic learning of haptic adjectives through physical interaction. *Robotics and Autonomous Systems* 63:279–292.

Fishel, J., and Loeb, G. 2012. Bayesian exploration for intelligent identification of textures. *Frontiers in Neurorobotics* 6:4.

Högman, V.; Björkman, M.; and Kragic, D. 2013. Interactive object classification using sensorimotor contingencies. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2799–2805. IEEE.

Kaelbling, L. P.; Littman, M. L.; and Cassandra, A. R. 1998. Planning and acting in partially observable stochastic domains. *Artificial Intelligence* 101(1):99–134.

Kurniawati, H.; Hsu, D.; and Lee, W. S. 2009. SARSOP: efficient point-based pomdp planning by approximating optimally reachable belief spaces. In *Robotics: Science and Systems Conference*, 65–72. The MIT Press.

Pajarinen, J., and Kyrki, V. 2015. Robotic manipulation of multiple objects as a pomdp. *Artificial Intelligence*.

Rebguns, A.; Ford, D.; and Fasel, I. R. 2011. Infomax control for acoustic exploration of objects by a mobile robot. In *Lifelong Learning*.

Sinapov, J., and Stoytchev, A. 2009. From acoustic object recognition to object categorization by a humanoid robot. In *Proc. of the RSS 2009 Workshop-Mobile Manipulation in Human Environments*.

Sinapov, J.; Schenck, C.; Staley, K.; Sukhoy, V.; and Stoytchev, A. 2014. Grounding semantic categories in behavioral interactions: Experiments with 100 objects. *Robotics and Autonomous Systems* 62(5):632–645.

Sinapov, J.; Schenck, C.; and Stoytchev, A. 2014. Learning relational object categories using behavioral exploration and multimodal perception. In *IEEE International Conference on Robotics and Automation (ICRA)*, 5691–5698.

Sridharan, M.; Wyatt, J.; and Dearden, R. 2010. Planning to see: A hierarchical approach to planning visual actions on a robot using pomdps. *Artificial Intelligence* 174(11):704–725.

Thomason, J.; Sinapov, J.; Svetlik, M.; Stone, P.; and Mooney, R. J. 2016. Learning multi-modal grounded linguistic semantics by playing I Spy. In *Proceedings of the Twenty-Fifth international joint conference on Artificial Intelligence (IJCAI)*.

Zhang, S., and Stone, P. 2015. CORPP: Commonsense reasoning and probabilistic planning, as applied to dialog with a mobile robot. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*.