

The Aspect Transition Graph: An Affordance-Based Model

Li Yang Ku^(✉), Shiraj Sen, Erik G. Learned-Miller,
and Roderic A. Grupen

School of Computer Science, University of Massachusetts Amherst,
Amherst, MA 01003, USA
{lku,shiraj,elm,gruppen}@cs.umass.edu

Abstract. In this work we introduce the Aspect Transition Graph (ATG), an affordance-based model that is grounded in the robot’s own actions and perceptions. An ATG summarizes how observations of an object or the environment changes in the course of interaction. Through the Robonaut 2 simulator, we demonstrate that by exploiting these learned models the robot can recognize objects and manipulate them to reach certain goal state.

Keywords: Robotic perception · Object recognition · Belief-space planning

1 Introduction

The term affordance first introduced by Gibson [2] has many interpretations, we prefer the definition of affordance as “the opportunities for action provided by a particular object or environment.” Affordance can be used to explain how the “value” or “meaning” of things in the environment is perceived. Our models are based on this interactionist view of perception and action that focus on learning relationships between objects and actions specific to the robot. Some recent work [6] [8] [13] in computer vision and robotics extended this concept of affordance and applied it to object classification and object manipulation. Affordances can be associated with parts of an object as, for example in the work done by Varadarajan [16] [15], where predefined base affordances are associated with surface types. In our work, we build models that inform inference in an extension of Gibson’s original ideas about direct perception [3] [5].

Affordances describe the interaction between an agent and an object (or environment). For example [2], a chair that is “sittable” for a grown-up might not be “sittable” for a child. In this work we introduce the Aspect Transition Graph (ATG), an affordance-based model that is grounded in the robot’s own actions and perceptions. Instead of defining object affordances from a human perspective, they are learned by direct interaction on the part of the robot. Using the Robonaut 2 simulator [1], we demonstrate that by exploiting these learned models the robot can recognize objects and manipulate them to reach goal states.

2 Aspect Transition Graph

Aspect Graphs were first introduced to represent shape [9] [4] in the field of computer vision. An Aspect Graph contains distinctive views of an object captured from a viewing sphere centered on the object. The Aspect Transition Graph introduced in this paper is an extension of this concept. In addition to distinctive views, the object model summarizes how actions change viewpoints or the state of the object and thus, the observation. Besides visual sensors, extensions to tactile, auditory and other sensors also become possible with this representation. The term Aspect Transition Graph was first used in [12] but redefined in this work.

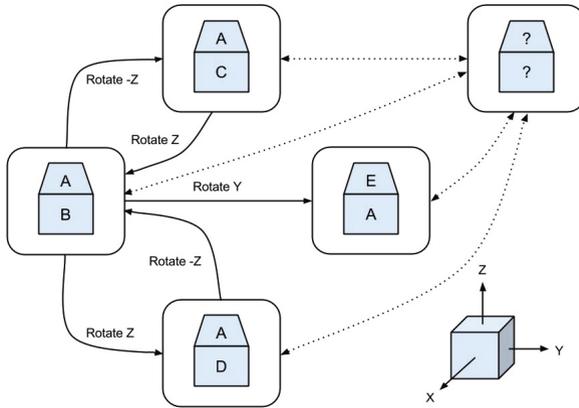


Fig. 1. An example of an incomplete aspect transition graph (ATG) of a cube. Each aspect consists of an observation of two faces of the cube. The lower right figure shows the coordinate frame of the actions and the aspect in the upper right is the “collection node” representing all unknown aspects of the object that may be present. Each solid edge represents a transition between aspects associated with a particular action. Each dotted edge is a transition that may not yet have been observed.

An object in our framework is represented using a directed graph $G = (\mathcal{X}, \mathcal{U})$, composed of a set of aspect nodes \mathcal{X} connected by a set of action edges \mathcal{U} that capture the probabilistic transition between the aspect nodes. Each aspect $x \in \mathcal{X}$ represents the features of an object that are measurable given a set of sensors and their relative geometry to the object. The ATG summarizes empirical observations of aspect transitions in the course of interaction.

The ATG of an object is complete if it contains all possible aspect nodes and node transitions. However, in practice, when ATGs are learned through exploration they are almost always incomplete. In addition, an object might be represented by multiple (incomplete) ATGs. A complete model is more informative but harder to learn autonomously. In this paper, we will focus on handling incomplete object models. Each of our ATG models have a single collection node

representing all unobserved aspects. Figure 1 shows an example of an incomplete ATG on a cube object with a character on each face.

3 Modeling and Recognition

The robot memory \mathcal{M} is defined as a set of ATGs that the robot created through past interaction. Each ATG in the robot memory represents a single object presented to the robot in the past. An ATG is added to the \mathcal{M} only if the presented object is judged to be novel. Although the robot might not have seen all the objects or all the aspects of each object, to simplify this problem we make this very limiting assumption that the robot knows that $|\mathcal{O}|$ objects exist in the environment and each object has $|G|$ aspects. If the robot assumes that there are more objects in the environment or more aspects of an object than there actually are, it will bias the judgment toward novelty.

Let \mathcal{S}_{T-1} denote the set of objects that have been presented to the robot in the first $T - 1$ trials. Given a sequence of observations $z_{1:t}$ and actions $a_{1:t}$ during trial T , the probability that the object presented during trial T , O_T , is novel can be calculated;

$$\begin{aligned} & p(O_T \notin \mathcal{S}_{T-1} | z_{1:t}, a_{1:t}, \mathcal{M}) \\ &= \sum_{o_i \notin \mathcal{S}_{T-1}} p(O_T = o_i | z_{1:t}, a_{1:t}, \mathcal{M}) \\ &= \sum_{o_i \notin \mathcal{S}_{T-1}} \sum_{x_t \in \mathcal{X}_i} p(x_t | z_{1:t}, a_{1:t}). \end{aligned} \tag{1}$$

Where o_i is an element of set \mathcal{O} designating all of the objects in the environment. Element x_t of set \mathcal{X}_i describes all the aspects comprising object o_i . The conditional probability $p(x_t | z_{1:t}, a_{1:t})$ of observing an aspect can be inferred using a Bayes filter. Object O_T is classified as novel if $p(O_T \notin \mathcal{S}_{T-1} | z_{1:t}, a_{1:t}, \mathcal{M}) > 0.5$.

If a particular object is judged to be a previously observed object, we associate it with the ATG that is most likely to generate the same set of observations. The posterior probability of object o_i is calculated by summing the conditional probability of observing aspect x_t over all aspects comprising object o_i ,

$$p(O_T = o_i | z_{1:t}, a_{1:t}, \mathcal{M}) = \sum_{x_t \in \mathcal{X}_i} p(x_t | z_{1:t}, a_{1:t}). \tag{2}$$

The posterior probability of an aspect $p(x_t | z_{1:t}, a_{1:t})$ is calculated after each measurement and control update using the Bayes Filter Algorithm [14].

4 Action Selection Strategy

The challenge of integrating task-level planners with noisy and incomplete models requires that we confront the partial observability of the state while building

plans. Since the true state of the system cannot be observed, it must be inferred from the history of observations and actions. Our planner belongs to a set of approaches (for example [7][10]) that select actions to reduce the uncertainty of the state estimate maximally with respect to the task.

Object recognition can be viewed as one such task in which the uncertainty over object identities (as quantified by the object entropy) is reduced with each observation. Our task planner selects the action a_t that minimizes the expected entropy of the distribution over elements of set O_T representing the object identity [11];

$$\begin{aligned} & \operatorname{argmin}_{a_t} E(H(O_T|z_{t+1}, a_t, z_{1:t}, a_{1:t-1})) \\ & = \operatorname{argmin}_{a_t} \sum_{z_{t+1}} H(O_T|z_{t+1}, a_t, z_{1:t}, a_{1:t-1}) \times \\ & \quad p(z_{t+1}|a_t, z_{1:t}, a_{1:t-1}). \end{aligned} \quad (3)$$

Once the object entropy is lower than a threshold, the robot has high certainty regarding the ATG in robot memory that represents the same object. All aspect nodes in this ATG that are reachable from the current aspect node represent the set of aspects that the robot can observe by executing a sequence of actions. If a goal aspect is in one of these aspects, the actions on the shortest path from the current aspect node to the goal aspect node represents an optimal sequence of actions for achieving the goal state.

5 Experiments

We evaluated the capabilities of the proposed affordance models and planner using the Robonaut 2 simulator shown in Figure 2. The simulation contains 100 unique objects called ARcubes that consist of a 28cm cube with unique combinations of ARtags on the six faces; 12 different ARtag patterns are used in this experiment. In an ATG for an ARcube, an aspect consists of ARtag features observed on 2 faces. Each ATG has 24 unique aspects and each aspect has 132 different pattern combinations. For the sake of simplicity, we assume that an object does not have two faces with the same ARtag. The robot can perform 3 different manipulation actions on the object: 1) flip the top face of the cube to the front, 2) rotate the left face of the cube to the front, and 3) rotate the right face of the cube to the front. We emphasize that our model is not restricted to cube like structures and that every inference is based on the combination of observation and action.

Table 1 shows the result of using the planner to recognize the object presented. Each test involves 100 trials and starts with an empty robot memory \mathcal{M} . In each trial, the task is to decide which ATG in memory the experiment corresponds to or to declare it to be novel. For each trial, an object is chosen at random and presented to the robot. The robot observes the object and executes an action. This process is repeated 10 times. At the end of each trial the robot



Fig. 2. The simulated Robonaut 2 interacting with an ARcube

determines the likelihood that the presented object is novel and the most likely existing object in memory is identified.

The last row in Table 1 presents the results averaged over all the tests. The success rate is the percentage of objects correctly classified, that is, correctly identified in memory or declared as a novel object. The system correctly recognizes the object 90.7% of the time, and correctly determines if the presented object is novel or not 81.6% of the time.

Table 1. The success rate of an information theoretic planner in recognizing the object (10 actions per trial)

Test	Correct Identification	Correct Recognition	Success Rate
1	80/100	20/21	79%
2	79/100	25/27	77%
3	87/100	21/25	83%
4	78/100	26/28	76%
5	84/100	24/27	81%
average	81.6%	90.7%	79.2%

We also tested the efficiency of the planner against a random policy. The number of actions executed per trial were varied from 4 to 20. Figure 3 shows how the success rate of a test varies with the number of actions executed per trial. As is evident from the plots, the information theoretic planner outperforms a random exploration policy for all cases except when the number of actions per trial is low. Both algorithms perform equally poor when not enough information is provided.

To demonstrate how ATGs can be used to reach certain goal state. We set up an environment where 3 ARcubes are located in front of the simulated Robonaut 2 as shown in Figure 2. The goal is to rotate the cubes till certain faces are observable. The robot starts with a robot memory learned through interacting with 20 different ARcubes including the 3 ARcubes located in the test environment. To achieve the goal state, Robonaut 2 manipulates the object to condense

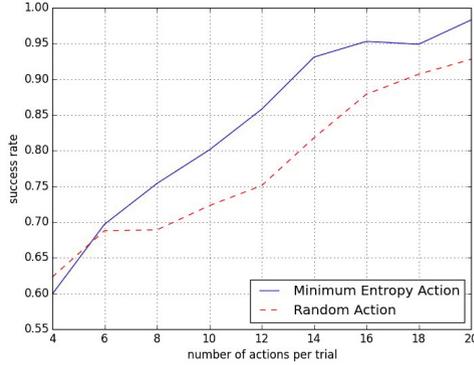


Fig. 3. The plot shows the average success rate of 10 tests as the number of actions per trial are increased. Selecting actions that minimize entropy leads to a higher success rate than selecting actions at random.

belief over objects. Once the object entropy is lower than a threshold, Robonaut 2 tries to execute the sequence of actions that is on the shortest path from the current aspect node to the goal aspect node in the corresponding ATG if such goal aspect node exists. In this experiment, the simulated Robonaut 2 successfully reaches the goal state by manipulating the cubes so that the observed aspects match the given goal aspects.

6 Discussion

This paper introduces an affordance-based model and demonstrates that it can be learned and used to support inference in a simulated environment with discrete actions and observations. To apply this model to real world applications, several challenges need to be addressed. First, in this work we assume that all actions lead to aspect transitions for all objects. A more realistic assumption will relate actions to new aspects probabilistically. Second, unlike ARcubes, real objects do not have a set of unique aspects; metrics such as the deviation of a new observation to past observations can be used to determine if a new aspect is observed. For future work, we plan to address these difficulties and test the ATG model in a more realistic environment. We are also exploring how to represent interactions between multiple objects in the scene and extensions of the idea that can incorporate multi-modal sensory features like tactile data.

Acknowledgments. This material is based upon work supported under Grant NASA-GCT-NNX12AR16A and a NASA Space Technology Research Fellowship.

References

1. Dinh, P., Hart, S.: NASA Robonaut 2 Simulator (2013). http://wiki.ros.org/nasa_r2_simulator (accessed on July 7, 2014)
2. Gibson, J.: The Theory of Affordance. In: *Perceiving, Acting, and Knowing: Toward an Ecological Psychology*. Lawrence Erlbaum Associates, Michigan (1977)
3. Gibson, J.J.: *The Ecological Approach to Visual Perception*. Houghton Mifflin, Boston (1979)
4. Gigus, Z., Malik, J.: Computing the aspect graph for line drawings of polyhedral objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **12**(2), 113–122 (1990)
5. Goldstein, E.B.: The ecology of j.j. gibson's perception, pp. 191–195. *Leonardo* (1981)
6. Grabner, H., Gall, J., Van Gool, L.: What makes a chair a chair? In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1529–1536. IEEE (2011)
7. Kaelbling, L.P., Littman, M.L., Cassandra, A.R.: Planning and acting in partially observable stochastic domains. *Artificial Intelligence* **101**(1), 99–134 (1998)
8. Katz, D., Venkatraman, A., Kazemi, M., Bagnell, J.A., Stentz, A.: Perceiving, learning, and exploiting object affordances for autonomous pile manipulation (2013)
9. Koenderink, J.J., van Doorn, A.J.: The internal representation of solid shape with respect to vision. *Biological Cybernetics* **32**(4), 211–216 (1979)
10. Platt, R., Tedrake, R., Kaelbling, L., Lozano-Perez, T.: Belief space planning assuming maximum likelihood observations. In: *Proceedings of Robotics: Science and Systems, Zaragoza, Spain* (2010)
11. Sen, S., Grupun, R.: Manipulation planning using model-based belief dynamics. In: *Proceedings of the 13th IEEE-RAS International Conference on Humanoid Robots, Atlanta, Georgia* (October 2013)
12. Sen, S.: Bridging the gap between autonomous skill learning and task-specific planning. Ph.D. thesis, University of Massachusetts Amherst (2013)
13. Stoytchev, A.: Toward learning the binding affordances of objects: a behavior-grounded approach. In: *Proceedings of AAAI Symposium on Developmental Robotics*, pp. 17–22 (2005)
14. Thrun, S., Burgard, W., Fox, D.: *Probabilistic robotics*. MIT Press (2005)
15. Varadarajan, K.M., Vincze, M.: Object part segmentation and classification in range images for grasping. In: 2011 15th International Conference on Advanced Robotics (ICAR), pp. 21–27. IEEE (2011)
16. Varadarajan, K.M., Vincze, M.: Afrob: The affordance network ontology for robots. In: 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1343–1350. IEEE (2012)