Interactive Categorization of Containers and Non-Containers by Unifying Categorizations Derived From Multiple Exploratory Behaviors

Shane Griffith and **Alexander Stoytchev**
Developmental Robotics Laboratory
Department of Electrical and Computer Engineering
Iowa State University
shaneg@iastate.edu and alexs@iastate.edu

Introduction

The ability to form object categories is an important milestone in human infant development (Cohen 2003). We propose a framework that allows a robot to form a unified object categorization from several interactions with objects. This framework is consistent with the principle that robot learning should be ultimately grounded in the robot's perceptual and behavioral repertoire (Stoytchev 2009). This paper builds upon our previous work (Griffith et al. 2009) by adding more exploratory behaviors (now 6 instead of 1) and by employing consensus clustering for finding a single, unified object categorization. The framework was tested on a container/non-container categorization task with 20 objects.

Initial attempts at robotic object categorization have produced limited results as they assume that robots will explore objects using only a single behavior (Griffith et al. 2009). Research with animals, however, has shown that some birds use almost their entire behavioral repertoire to explore a novel object (Lorenz 1996). This suggests that robots should do the same when categorizing objects. Indeed, an object categorization derived from multiple exploratory behaviors may contain more information compared to one derived from a single behavior. Further work is necessary, however, to determine how a robot can combine its observations from multiple behaviors to come up with one unified categorization for a set of objects, instead of having a separate categorization for each behavior.

This paper tests the hypothesis that a robot can use consensus clustering to form a single categorization for a set of objects after it interacts with them using multiple exploratory behaviors. Our robot performed a sequence of 6 exploratory behaviors during multiple interaction trials with 20 objects (10 containers and 10 non-containers). The robot extracted features from its interaction history with each object and then employed unsupervised clustering to form 6 different categorizations. Consensus clustering was used to combine the 6 different categorizations into a unified object categorization. This resulted in a meaningful separation of containers from non-containers, even in the presence of noisy clusterings.

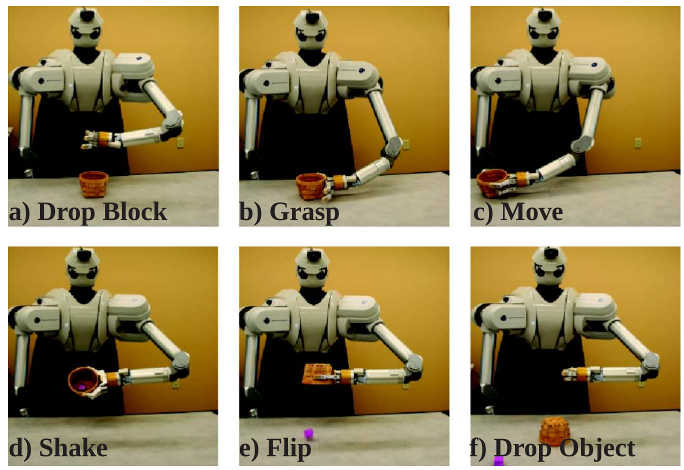

Figure 1: The six exploratory behaviors performed by the robot. Before each trial the block and one of the twenty objects were placed at marked locations on the table. The robot performed each behavior in the order shown after it grasped the block and positioned its arm in the area above the object.

Experimental Setup

The upper-torso humanoid robot shown in Fig. 1 was used for this study. Two 7-dof Barrett Whole Arm Manipulators (WAMs) were used for the robot's arms, each with the 3-finger Barrett Hand as its end effector. A 3-D camera (ZCam by 3DV Systems) was mounted on the robot and used to capture both color and depth images of the environment.

The robot interacted with one block and 20 objects placed on a table in front of it (see Fig. 2). Half of the objects were containers (household containers or children's bucket toys); the other half were the same objects, only flipped over. So while there were only ten real objects the robot was exposed to 20 "different" objects from an interaction point of view. Each trial consisted of six behaviors with the block and one of the objects (see Fig. 1). A total of 12,000 behavioral interactions were performed (6 exploratory behaviors per trial and 100 trials for each of the 20 objects).

Methodology

The robot used the visual co-movement patterns of the block and the object in order to form object categories. A movement was detected when the position of the block or the position of the object changed by more than a threshold, δ, over a short temporal window. A box filter was used to remove noise from the movement detection data.

The movement data for the block and the object was converted into a state sequence. The states of this sequence were four visual co-movement events: 1) neither object moved; 2) the block moved; 3) the object moved; or 4) they both moved. The robot acquired a set of 2,000 state sequences from the 2,000 trials for a given behavior. The robot recursively bi-partitioned the set of state sequences using the spectral clustering algorithm in order to learn visual outcome classes. The visual outcome classes, $C = \{c_1, ..., c_k\}$, are the leaf nodes of the tree created by the recursive algorithm. The robot repeated the same process for each of the six different exploratory behaviors in order to learn six different sets of visual outcome classes.

Next, the robot categorized the objects using the frequency with which different visual outcomes occurred with each object. More formally, given a set of visual outcome classes, $C = \{c_1, ..., c_k\}$, each object, i, is described with a feature vector $H_i = [h_1^i, ..., h_k^i]$ such that h_j^i is the number of outcomes from outcome class c_j that were observed when interacting with object i. Objects are grouped into categories by unsupervised clustering of their H vectors using X-means. This process is repeated for each of the six different exploratory behaviors to form six different sets of object categories.

Unifying the six object categorizations is a necessary step toward identifying a single behavior-grounded categorization of the objects. The best unified categorization is defined as the clustering that has the highest possible total normalized mutual information with the six input clusterings. Finding the best clustering, however, is intractable. Thus, it is necessary to search for a clustering that is approximately the best. For this task, we used the hard consensus clustering algorithm (Strehl and Ghosh 2002). The algorithm takes as input the six clusterings formed for each of the six exploratory behaviors, searches for a good approximation, and outputs the best unified clustering that it finds.

Results

All six behaviors produced visual co-movement patterns that could be used for object categorization. Some behaviors captured the ‘container’ property better than others. The *flip* behavior, for example, led to a categorization that perfectly matched human labels. Next in order were *move* (3 incorrect classifications), *shake* (4 incorrect), *drop object* (5 incorrect), *drop block* (6 incorrect), and *grasp* (7 incorrect).

The fact that some clusterings produced by the robot were noisy was expected. Some behaviors are simply better at capturing certain object properties than others. With 20 objects of various shapes, sizes, and materials there are many ways the robot could have categorized them. No behaviors, however, clearly separated objects by size or material. On the other hand, *flip* and *move* captured the ‘container’ functionality well.

The result of unifying the six object categorizations is shown in Fig. 2. The consolidated clustering separated the objects into two groups, which closely correspond to what a human would call containers and non-containers. Only two

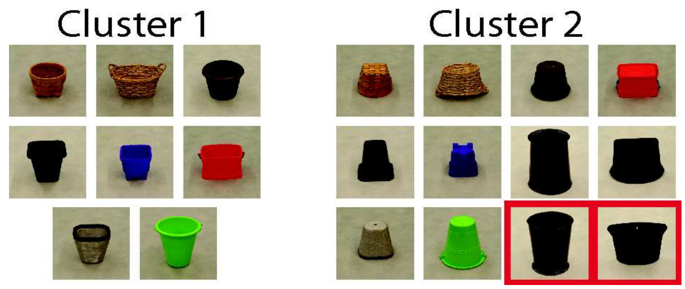

Figure 2: Illustration of the unified object categorization formed by the robot. Only two objects were incorrectly classified (when compared with the ground-truth labels provided by a human and the majority class of the category).

objects were misclassified, which shows that the consensus clustering algorithm was able to find a meaningful categorization even when only two behaviors produced a good clustering of the objects. The noisy clusterings produced by the other behaviors only marginally affected the consolidated clustering. Thus, the consensus clustering method successfully unified the clusterings from multiple behaviors in order to identify a single, behavior-grounded categorization.

Conclusion and Future Work

The experiments show that a robot can derive meaningful object categories by interacting with objects and observing their visual co-movement patterns. Although categorizations derived from some of the interactive behaviors were noisy, the consensus clustering algorithm identified a meaningful object categorization (relative to human labels). Thus, different categorizations derived from different exploratory behaviors can be combined into a single one using consensus clustering. The experiments and the findings of this paper are still preliminary. We plan to extend this study in the future by including auditory events in the framework and by analyzing the effects of different clustering algorithms. For additional details and results, see http://www.ece.iastate.edu/~shaneg/AAAI10/.

References

Cohen, L. 2003. *Early category and concept development*. New York: Oxford University Press. chapter Unresolved Issues in Infant Categorization, 193–209.

Griffith, S.; Sinapov, J.; Miller, M.; and Stoytchev, A. 2009. Toward interactive learning of object categories by a robot: A case study with container and non-container objects. In *Proc. of the 8th IEEE Intl. Conf. on Development and Learning (ICDL)*.

Lorenz, K. 1996. *Learning as Self-Organization*. Mahwah, NJ: Lawrence Erlbaum and Associates, Publishers. chapter Innate Bases of Learning, 1–54.

Stoytchev, A. 2009. Some basic principles of developmental robotics. *IEEE Transactions on Autonomous Mental Development* 1(2):122–130.

Strehl, A., and Ghosh, J. 2002. Cluster ensembles - a knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research* 3:583–617.