Learning to Detect Containers with Human Assistance

Introduction

A long term goal for the field of human-robot interaction is to create robots that assist humans with laborious tasks and to integrate them into society. Assisting the elderly and disabled, helping a person to assemble a large structure, or cleaning and organizing a house are applications waiting for future robots. Giving robots the ability to operate in unstructured human environments is a necessary condition for progress in this area.

Robots that help humans must be able to use everyday objects, but an innumerable amount of object types exist in human environments. Take, for instance, an electric plug and an outlet: the distal end of the plug tightly fits the concave region of the outlet. There are many objects that interact with each other in this respect, but robots must learn to find these object pairs to perform useful tasks. A similar problem is that of finding containers and the objects that fit into them. Containers are larger and make it easier for a robot to learn the fundamental characteristics of objects that fit together.

The process of learning about the environment speeds up by interacting with humans. For example, parents give infants toys that target key concepts in learning, like the “baby’s first blocks” toy set. This set includes several blocks made of various shapes and one container with a hole that matches each block. Parents commonly shake or touch the block to direct their infant’s attention to it, and, with the infant’s attention, put the block in the hole. Infants may lose track of the blocks by throwing them, or lose interest in the set without a parent’s guidance. Similarly, robots in the same situation may overlook important details of the set without a person’s help. Therefore, robots may become better at labeling novel objects with human interaction.

This work addresses a robot’s ability to learn to distinguish containers from non-containers through interaction with them in its environment and through human-assisted learning. Work is presented that shows how a robot may find containers in the environment through interaction. In addition, work is proposed that addresses how a robot’s model of containers may improve with human interaction.

Related Work

Little related work addresses the stated problem. Saxena et al. [2] shows that good places to grasp are frequently identified on the rim of containers. The feature detector used could supplement a robot’s container detection system. Kemp et al. [1] introduced the notion of using task-relevant features to simplify robot manipulation tasks. Task-relevant features may provide one way to detect containers.

Finding Containers through Interaction

Useful robots must have an ability to distinguish between containers and non-containers in the environment. To do this, robots need to find the affordances of containers that make them different from non-containers. These affordances are found only through interaction. In this experiment, a robot is programmed to interact with containers and non-containers to learn to distinguish between them.

A 7-dof Barrett WAM coupled with the Barrett Hand was used in the experiments. The ZCam, an RGB and depth camera in one system, was used to capture video at 30 Hz. Five containers and five non-containers were used – all of which were infant toys. In each trial, the robot performed a sequence of interactions with a toy block and one container or non-container. Through 1000 trials, scripted movements guided the actions of the robot while a person kept the container and block in the robot's range.

Containers have the property that the block moves with the container when dropped inside, thus, block and container movement trajectories were processed to identify containers from non-containers. X-means first clustered movement data to separate move-together events from move-separately events. The result of clustering was a frequency of move-together events versus move separately events for each container and non-container. X-means accurately labeled trials as move-together or move-separately events 97% of the time compared to human-labeled trials. Further, since humans find relationships among statistical data, X-means was applied a second time to cluster objects. To cluster objects, it used the frequency values of each container and non-container. X-means found a strict separation between containers and non-containers with this approach.

A Perceptual Model of Containers Learned through Human Interaction

Using interaction to label novel objects as containers or non-containers takes a lot of time; however, using task-relevant features to label novel objects is much faster. We propose to create a container detection learning system that takes advantage of human assistance. A perceptual model, which combines the task-relevant features from both the RGB and depth images of the ZCam, would determine the label to give to novel objects. An initial container perceptual model would come from containers in the first experiment. The learning experiment would go as follows: 1) a person presents the robot with a novel object; 2) the robot pushes containers left and non-containers right (to group similar objects together); and 3) the person either continues with step 1 if the robot is correct, otherwise the person says "no" and moves the object to the center, which prompts the robot to re-evaluate its perceptual model. The robot will be evaluated on the rate that the container classifier improves across a set of 50 novel objects.

Future Work

With an ability to identify containers, the next step is to learn to manipulate the container to consistently produce a move-together event. Following this, the object properties that allow objects to fit inside a specific container should be investigated; for example, shape, size, and position. A real-time learning architecture, which combines the steps into an autonomous learning system, should be explored once many steps in the process are completed.

Conclusion

This work addressed a robot's ability to separate containers and non-containers using an interaction-based approach. The robot precisely separated containers from non-containers with human guidance during experimentation. A method is proposed that takes advantage of human interaction for learning to detect novel objects as containers or non-containers. Even with this work, little work addresses the problem of giving robots an ability to find congruencies between objects, meaning a wide gap in research prevents integration of robots into society. Although this is only a small step toward integrating robots into society, robots could not operate in unstructured human environments without an understanding of containers.

References

[1] A. Edsinger, and C. C. Kemp, “Robot manipulation of human tools: Autonomous detection and control of task relevant features,” in *Proceedings of the Fifth Annual Conference On Development and Learning, Special Session on Classifying Activities in Manual Tasks*, 2007
[2] A. Saxena, J. Driemeyer, and A. Y. Ng, “Robotic grasping of novel objects using vision,” *International Journal of Robotics Research*, vol. 27, no. 2, pp. 157-173, 2008