Paper Review: Segmentation via Manipulation

Alvin Sun

November 1, 2021

1 Paper Summary

This paper presents a interactive manipulation method that can adjust grasp plans on the fly based on sensory feedback from interacting with the environments. This paper focuses particularly on the domain of decomposing heaps of partially observable objects in a post office settings. The heap of objects is observed using a real-time range image sensor and represented as a growing graph with edges indicating spatial relation among the objects. The control strategy is formulated as a non-deterministic finite-state Turing machine where the state transitions involves actuation and sensing that modifies the object graph. This paper proposed four specific state machines with progressively higher level of interaction, carried out experiments for each of the strategies, and recorded the success rates. The results indicated that the interactive strategy is the most effective for this heap sorting task.

2 What I Learned

1. The concept of using interaction to progressively build up the understanding of a partially observable scene is really useful, especially when the scene is dynamic, where interaction changes the observability of the objects of interests.

2. Feedback control is the way to go for any type of control task, as it adapts to environmental changes that is almost certainly going to occur in reality.

3 Opinions

3.1 Up Votes

I agree about how this work sets up the experiments with strategies derived with progressively higher level of interaction. The experimental results perfectly validate their hypothesis that interaction can be very effective for heap segmentation tasks.

3.2 Down Votes

Given that this work is carried out in the last century when just compute is a scarce resource, I believe it is quite amazing how much this paper accomplishes. However, one slight disagreement I have is how the evaluation uses state transition number to compare across the different strategies. Given that the different strategies use different set of actions for state transitions, I don't think the number of transitions is a good metric for representing the grasp plan efficiency. A more appropriate metric could be the average time to complete the sorting since all experiments are conducted with the same hardware platform.

4 Evaluations

The goal of this paper is to propose a novel paradigm of scene segmentation that takes advantage of interactive manipulation and interaction. This is a perfectly valid objective as this work is presented in the pre-deep-learning era, where pure vision based segmentation methods are intractable due to compute and resolution bottlenecks. The domain of this work also presents possibilities for real-world deployment as it does not require too much a priori assumptions on the geometries of the objects to be manipulated and segmented.

The overall quality of this work is exceptional in my opinion. Given such limited compute and sensing hardware, this paper really show cases how explicitly exploiting interaction can play an important role in scene understanding as well as manipulation. This work also clearly states their assumptions. Even though it is far from what a day-to-day manipulation complexity a human has to deal with, those assumptions are quite realistic in controlled domains such as an assembly line or post office as mentioned in the paper. Under those assumptions, the experimental results also demonstrated very high success rate given the scene is composed of partially observable overlapping objects with unknown dynamics. Again I would like to emphasize how impressive this is given the limited compute and sensing resource they had at the time.

5 Questions

1. In strategy 4, if the partially visible objects are first displaced out of the scene, wouldn't that make them lost objects?

2. In the evaluation metrics, what does the number of top-most surface segments identified per scene mean?