CS 231A Project Proposal

Segmentation of RGB-Z Images of Cluttered Scenes Using Kinect

Budi Waskita
Stanford University
bwaskita@stanford.edu

Abstract

I am interested in exploring the problem of image segmentation using color-depth sensor like Kinect. Low-cost depth camera opens up many possibilities in the application of computer vision techniques, since depth data previously could only be collected using stereo cameras or through other expensive, specialized sensors such as laser or acoustic ranging devices. In particular, I would like to explore the use of Kinect in segmenting cluttered scenes, in which many objects of interest are present in a relatively small space. These objects present a difficult problem for segmentation, since they are mutually occluded in the 2D image. However, the use of Kinect could greatly help even traditional 2D segmentation algorithm through the use of depth cues. The segmented image could then be fed to another object recognition algorithm or used to help a robot manipulating objects in cluttered environments.

Approach and Related Work

I am planning to explore and test the methods described by Bleiweiss and Werman [1] using Mean-Shift Algorithm, Taylor and Cowley [2] using RANSAC, as well as other methods such as using applying edge detector on depth data.

Data for this project will be collected using a Kinect sensor, using both its color and depth camera. Several exposures may be averaged to reduce the noise uncertainty from the depth camera, and possibly an additional color camera may be used if Kinect's RGB camera performs unsatisfactorily. I plan to collect the data on my own, using household objects that are packed closely together, whether on a table or in a plastic bin to simulate a clustered scene that a robot might encounter.

The results should be in the form of an annotated image, where each object will be surrounded by border lines separating it from other objects and the background. I plan to evaluate my results by comparing it to human-generated ground truth, other implemented RGB-Z algorithms, and algorithms that only use RGB data. The planned algorithm should be able to pinpoint the borders and locations of objects within a tenth of an inch for effective robot manipulation, with errors (such as merged borders and multiple splitting of an object) to be kept below 10% (i.e. only 1 error for every 10 objects correctly segmented).

References

[1] Bleiweiss, Amit, and Michael Werman. *Fusing Time-of-Flight Depth and Color for Real-Time Segmentation and Tracking*. Web.

[2] Taylor, Camillo J., and Anthony Cowley. *Segmentation and Analysis of RGB-D Data.* Web.