Feature Extraction from Depth Maps for Object Recognition

Caleb Jordan
Department of Computer Science, Stanford University

Motivation

With widespread availability of depth sensors, recognition systems have an exciting new source of information to help bridge the gap between 2D and 3D. This work attempts to more fully explore possible 3D and multimodal features derived exclusively from depth data for the purpose of designing more descriptive image features.

Related Work

There exists a variety of work towards recognizing objects in 3D and recognizing depth maps. Common features generated from depth maps include curvature estimates and scale heuristics. Point cloud registration has been explored for decades to match images as well. The primary motivations for this project come from the following papers:

• K. Lai, L. Bo, X. Ren, and D. Fox. "A Large-Scale Hierarchical Multi-View RGB-D Object Dataset"
• Osada, Robert, et al. "Matching 3D Models with Shape Distributions."
• B. Drost, S. Ilic. "3D Object Detection and Localization using Multimodal Point Pair Features"

Dataset and Features

Tested features include two HOG baseline features, histogram of normals, and several variants of the shape distribution method. Each of these is a distribution over distances between pairs of sampled 3D points—the variation is in the points sampled. The masked version samples only points in the ground-truth image mask, while the edge distribution samples from points in the canny edge mask.

Experimental Results

The histogram of normals performed the best, similar in performance to traditional grayscale HOG. HOG over the depth image as grayscale had typically 20% more error. This demonstrates that at least in this context, object recognition with zero color data is possible.

Surprisingly, all shape distributions performed poorly, with the basic variant best of all. Although sampling from important regions (edges or just the object) seems more appropriate, background may provide context (as in all applications), and minor errors in pixel labeling may dramatically skew the data.

Further research would be well-spent experimenting with these features in more elaborate models. While this project focused on object recognition, these features could certainly be used for many other contexts.