Real Time, Scalable Object Recognition with Linemod and Winner Take All

Abi Raja
Stanford University
abii@stanford.edu

Ivan Zhang
Stanford University
zhifanz@stanford.edu

1. Project Proposal

1.1. Problem

We will be investigating general, real-time object recognition. This problem is interesting because of its various applications in robotics. A good real-time object recognition algorithm would enable robots to perform complex tasks such as identifying mugs in the close vicinity in heavily occluded scenarios and fetching it for the human user.

1.2. Data

We will start with image data and masks provided by Dr. Gary Bradski. Later we may collect new image data and masks using our own camera.

1.3. Algorithm and Implementation

The algorithm that we are working on involves the combination of two existing techniques that have proved quite useful at object recognition, "Linemod" and "winner take all". Existing implementations of the Linemod algorithm in C++ was provided by Dr. Gary Bradski. We will use and improve upon his implementation and develop a working version of "winner take all" ourselves. Then we would put together the two algorithms into one working system so that they complement each other to achieve fast and scalable object recognition.

1.4. Context and Background

We will first focus our attention to the following two papers:

Multimodal Templates for Real-Time Detection of Texture-less Objects in Heavily Cluttered Scenes. IEEE International Conference on Computer Vision (ICCV), Barcelona, Spain, November 2011.

The Power of Comparative Reasoning. Jay Yagnik , Dennis Strelow, David Ross , Ruei-Sung Lin. International Conference on Computer Vision (ICCV), 2011.

Both papers are recent publications and have yet to be formally presented at conferences. Hence, there are no existing papers that develop off them. However, we are also looking into reading into other prior art related to template-matching.

1.5. Evaluation

Since our project produces very visual results, qualitatively we can evaluate our algorithm by observing how well it performs the recognition task on various image data.

We will also use standard quantitative methods to evaluate our results, such as constructing the precision and recall curves when the system identifies various objects. To resolve the ambiguity involved with a "correct" result, we define a recognition to be correct if the bounding box overlaps at least 50% with the bounding box a reasonable human would use to bound the object, i.e. the "hard truth". We will also explore whether the Pascal VOC Challenge benchmark is applicable in this recognition problems.