Computer Vision Final Project Proposal

Kevin DuCharme   Max Li

November 13, 2017

1 Motivation

Robots are quickly becoming a staple in everyones living room, with the likes of Alexa and Google Home acting as virtual assistants that aid in everyday tasks. These first-wave household assistant robots primarily focus on verbal interactions, and are fixed to one location. Recent developments have started to branch out, with computer vision bringing a whole new domain of possibilities to these devices. Jibo, a robot out of the MIT Media Lab, is an example of that. Jibo is a natural extension of the Alexa concept, acting as a mobile home assistant. Computer vision is not only used to navigate, but also for the recognition of common faces in the household, among other functions. While the goal of this project is not completely in line with Jibo, it represents one of the many ways that these personal assistants can intersect with computer vision.

The main motivation behind this project is the struggle of finding items on a heavily used desk. On an active workbench, tools and other useful items are frequently getting picked up, used, and placed down somewhere else. While good standard operating procedures will make sure all the parts end up where they're expected to be, it's very easy to quickly lose track of items.

The goal of this project is to start the foundation of a lab assistant robot that has the capability of tracking common items found on a workbench in pseudo real-time. When a user queries the robot, it should be able to tell the user where a specific tool is, in a way that would make sense to the user. Initially, the objects being identified would come from a preset bank of objects that have been pre-trained. However, the end product will likely have to be trainable on new objects by the user.

For this project, the main topic of focus is object tracking. After all, the rest of the features are both fairly difficult (and thus time-consuming), and out of scope of the computer vision task. As such, the goal is to be able to detect a number of different objects, identify what they are, and track them as they move across a camera.

2 Potential Approach

In order to reduce the number of non-idealities to deal with, the test setup will be kept as simple as possible. A top-down viewing camera positioned over a desk will be utilized to track the trained objects. The area the camera covers will be considered the whole workspace, and will be well-lit.

Because of recent advances in GPU processing, machine learning, and in particular deep learning, have become the standard practice in object identification and tracking. As such, it is likely that this project will use these concepts for object detection, identification, and tracking. Because these are becoming the standard, there are plenty of open source libraries that can be used, including OpenCV and Tensorflow. The majority of relevant Machine Learning libraries are applied to Python, and so that is the language that will be used.

Due to the end goals of the robot (including the goals that extend beyond the project), both the training dataset and the test dataset will be self collected if possible. The end goal is to be able to run this tracking on an embedded-class device, such as a Raspberry Pi, but in the scope of the project the tests may be run on a computer. Pseudo real-time performance will be attempted, but the foundational goal will to be able to track the objects once data has been recorded.

3 Distribution of Responsibilities

At this moment, the project is very much still in the exploratory phase. As such, there is not much of a distribution of work. Once the idea is more fleshed out, a more proper distribution will be decided upon.

4 Project Evaluation

The base level of functionality to be achieved is a post-processing interface that can identify and track up to two objects using self-collected data, either trained on self-collected data or on relevant datasets.

The targeted goal is to have a pseudo real-time interface that can identify and track up to five objects, trained on self-collected data.

There are many possible extensions that can happen if time allows, including hand gesture recognition. However, these are reach targets, and will only be approached if the main target is easier than expected.