Progress Review 9

Rick Shanor

Team D: Human Assistive Robotic Picker

Teammates: Alex Brinkman, Feroze Naina, Abhishek Bhatia, Lekha Mohan

ILR08

March 18, 2016

Individual Progress

Over the past two weeks, I have worked on investigating another end to end vision pipeline that can relieve some of the issues we currently face. Right now, I am working on methods to improve both item segmentation from the shelf as well as item identification.

Our current perception system relies on precise localization. The way our current system is configured, the robot localizes once at startup. From there, the arm orients the camera outside the shelf. Using the initially calculated shelf position, the algorithm removes the image background based on depth data. However, due to inaccuracies in localization and extrinsic camera calibration, a large buffer zone is required and thus often too much of the image is removed to get good perception results. In order to remedy this issue, I trained a convolutional neural network based on the SegNet framework. This framework is set up to label each pixel in an image as either shelf or not shelf. To train the network, I captured 80 images of cluttered shelves and labeled the items using a tool called LabelMe. From there, I rotated, mirrored, and changed image hue to increase training data and decrease overfitting. An example output is shown in figure 1 below. The model classified approximately 95% of the pixels correctly on the test images.

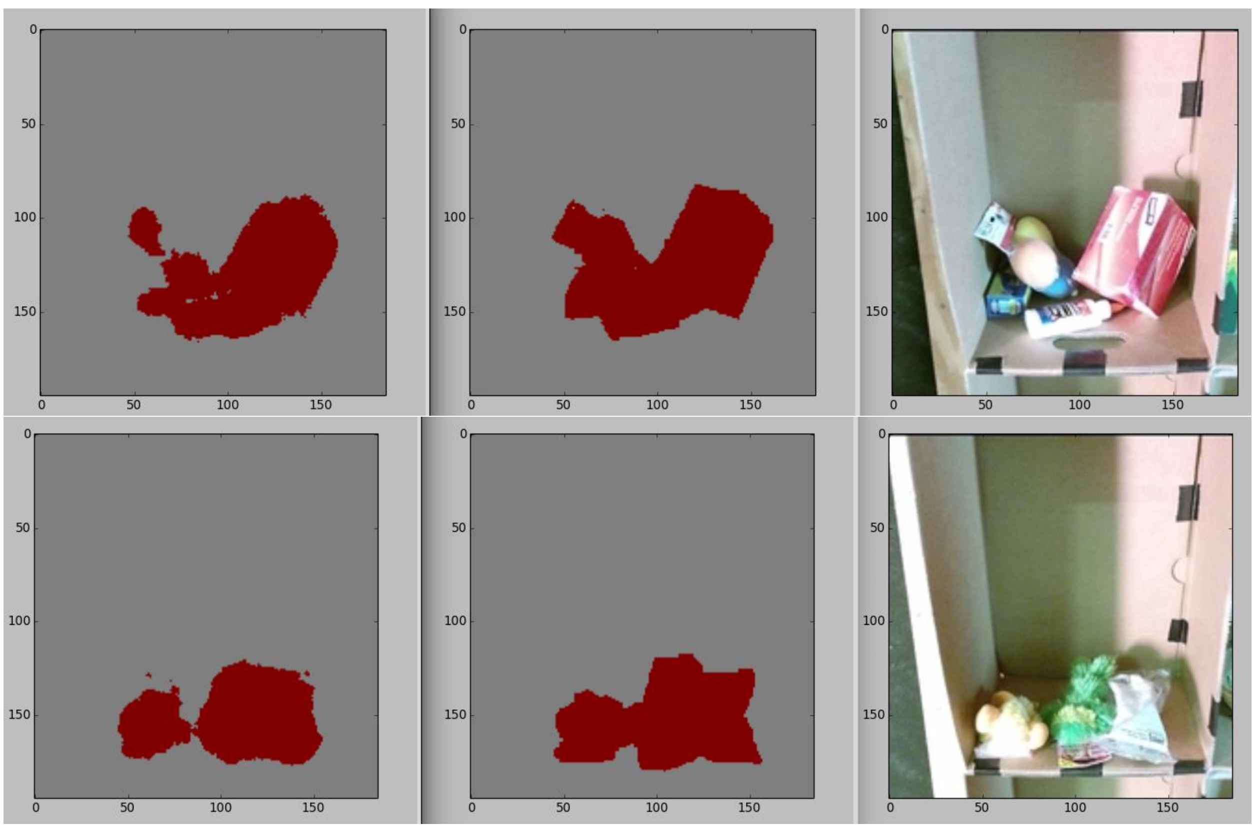

Figure 1 SegNet Output, Ground Truth, and Shelf Image

After having success with this segmentation CNN, I also started working on ways we could use a CNN for the identification task. After talking with many CV specialists, I finalized on the following approach. Given RGBD data of items on a shelf, a region growing segmentation method can create sizable superpixels. By over-segmenting, we ensure that a cluster exists for each item of the shelf. From there. We can input each superpixel into an item-recognizing CNN. The output will be a 40x1 vector that indicating the probability the superpixel belongs to each item in the item database. From there, once each superpixel is labeled, a global policy will optimize the scene energy function. This energy function will merge clusters such that an overall scene probability is maximized.

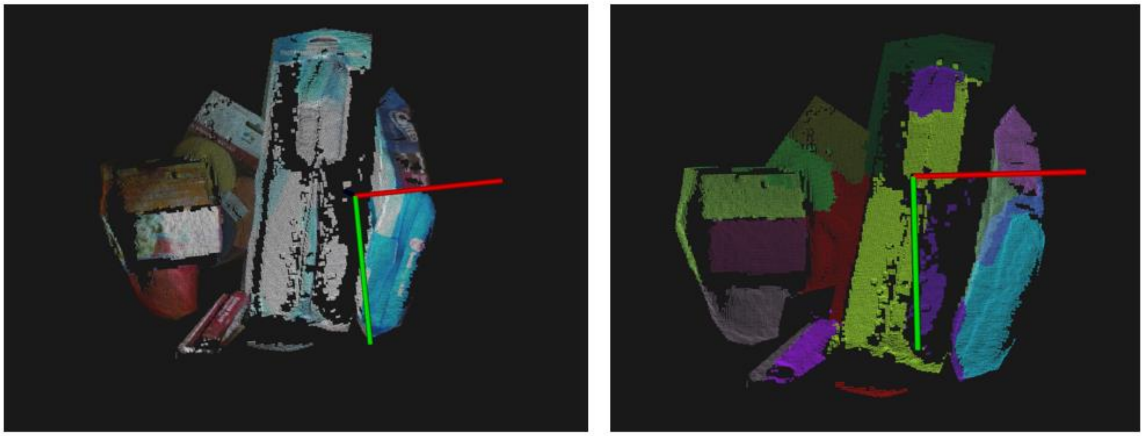

Figure 2: Items (Left) and Superpixels ((Right)

I plan to train the identification CNN using the AlexNet framework. Alexnet was designed for image classification, so it should be perfect for this task. I wrote a program that converts back and forth between point cloud data and images for the neural net. This generated input can be seen in figure 3. Last week, I trained a preliminary model using data released from Rutgers' 2015 Amazon Picking Challenge team. Although the training data was less than ideal, I was getting approximately 75% accuracy during my preliminary tests. Once the new item list has been released, we will use a turntable setup to generate thousands of training images. By implementing this new strategy, we will be able to globally optimize our item predictions and handle dense shelves plus heavy occlusions.

Figure 3: Example network inputs

Challenges

The biggest challenge over the last two weeks has been testing these ideas highlighted above. Specifically, we don't have the 2016 items, and it doesn't make sense to create an extensive 2015 item database. Thus, although we have discussed this approach with CV experts, it is difficult to fully verify our plans. However, the framework is now in place to quickly implement these strategies when the new item list arrives.

Teamwork

Since last progress review, Feroze has been working on an online grasp planner that bases grasp locations based on just the runtime item point cloud. This system couples with any vision algorithm that does not estimate the exact pose of an object on the shelf. Abhishek and Lekha have worked with me on several perception tasks. Lekha has helped outline the overall CNN structure and also helped collect training images for the SegNet segmentation architecture. Abhishek helped write a tool that masks images such that we can 'ground truth' for our training. Finally, Alex has been working on systems integration. Alex and Abhishek did most of the on the single item test. Alex also began work on integrating the SBPL arm planner. This planner should reduce planning time and improve planning constraints.

Plans

Over the past month, I have been exploring lots of potential perception approaches, including geometry-based ICP identification, color-based histogram identification, PERCH scene rendering, and now convolutional neural networks. It is time to hone in on which approach is working best, and from there integrate it in parallel with our current vision pipeline. Hopefully Amazon will release the new item list soon, which will guide this decision process. I am also meeting with one of our technical advisors to hammer out these details on March 17th. In addition, I want to try the SegNet segmentation on the real robot.