Real Time Detection of Groceries Proposal

Tess Winlock
Department of Computer Science and Engineering
University of California San Diego
La Jolla, CA 92093
twinlock@cs.ucsd.edu

Abstract

The problem of detecting objects in a scene with limited prior data is becoming a more well explored area of computer vision. The GroZi project aims at allowing a blind user to independently navigate a grocery store and collect the items on their grocery list. Using a shoulder mounted camera the project aims to detect objects and direct the user to the object. There are several major problems presented in trying to detect objects in such scenes: occlusion, motion blur, focus, and pose to name a few. This problem has been explored previously by [3], where they explored detecting objects in a grocery scene using images collected from 'ideal' situations. This projects aims to continue this work, applying detection algorithms for real time recognition of products from a shopping list. I intend on taking 4 promising algorithms and comparing their performance across a large number of items and video to determine if there is one algorithm which performs best, or if a combination of approaches need to be undertaken. After comparing the performance of various detection algorithms individually on various scenes, I plan on applying the classifiers in parallel to try to detect which of the desired objects exist in a given image. This will rely heavily on which algorithms will have been found to be best for this problem.

1. Qualifications

Tess Winlock is a second year masters student, and is currently working on the GroZi project. I have previously completed two graduate level classes in machine learning, and a course in computer vision. In my undergrad I worked on a project to build a cheap device to assist those with cerebral palsy communicate. I have also completed 2 internships at Google working on a team that uses machine learning to detect payment fraud on Google Checkout and Google AdWords. Further, Though unrelated during my undergraduate career I studied human navigation in a virtual reality environment. Overall this has given me a good background in machine learning and working with the disabled.

2. Project Outline

The final goal of this project is to be able to accurately detect the location of one object from a shopping list in an image provided from a video. The algorithms that I plan on exploring are:

- SURF - I currently use SURF [1] to match keypoints in the training image to those in the scene, and then try to find a homography for planar objects.
- Boosted HAAR-like Features - Using the Viola Jones method [5] explored by Merler's and Galleguillos on this dataset, I intend on expanding their approach by not only training the entire object, but also on detecting sub features (like the brand label) once the location of the object has been detected, and then applying geometric constraints to these.
- Ferns - Ozuysal [4] has used ferns to supposedly quickly recognize keypoints in a scene, it seems to be an interesting approach that may compete with SURF.
- One Way Descriptors [2] - Another real time approach which attempts to match patches taken from keypoints in the input image to those in the scene, shows promise on being quick and providing a homography upon detection.

I intend on using training data from the GroZi-120 dataset to compare the performance of all these algorithms, and then compare their performance across the products. The criteria for the final algorithm will be heavily driven by these findings, as well as the speed of detection. For our final product we will need to detect multiple objects at once, this could require several classifiers to run in parallel, which means performance is a non-trivial concern. Application of these classifiers in parallel will rely heavily on the results of the

first milestone, there may not be a single algorithm which performs best under all situations. Thus we may need to run a classifier capable of detecting an object's presence, and then running the more exhaustive detector. Once this has been built, the final test will be run on the complete GroZi-120 dataset with randomly generated shopping lists of 10 items.

2.1. Milestones

1. Week 2: Application of all algorithms - I am currently working on this section of the project, so I hope to be able to apply all algorithms to the data by the end of the second week of the quarter.

2. Week 5: Comparison of performance of 3 different algorithms - Once I have the various approaches complete, I intend on testing their performance on the complete GroZi-120 dataset. I may also expand this on to videos that show the products from angles other than fronto parallel

3. Week 8: Development of Parallel Algorithm - Once the algorithm(s) have been selected I can try to detect one of many objects in a scene.

4. Week 10: Test of Complete application - If/when the parallel approach to application has been determined I will run the application on the GroZi-120 dataset with a shopping list of 10 items.

3. Questions

- Which algorithm shows the best performance at recognizing items?

- Will one algorithm work for all items or is there some criterion for using different approaches for different items (say for planar objects vs. cylindrical objects)

- What is should be used to run these algorithms in parallel?

- How to pass to the detection data to tracker?

4. Software

I am currently developing the detection algorithms in OpenCV, as they have many useful algorithms already implemented. OpenCV has implementations of SURF, Cascade classifiers using HAAR, and the OneWayDescriptor. I plan on continuing to use this throughout the duration of the project.

5. Data

I will use video shot in the old UCSD sunshine stores as my primary data set, and possibly videos from Von's if we get permission to shoot new videos. The GroZi-120 dataset will provide a good base for comparing the different algorithm's performance under ideal conditions, but we will likely need to expand the data to include perspectives that are not just fronto parallel. This will require some significant gathering and labeling of videos with the various visible products of interest. To accommodate this I have a simple program, written using OpenCV, that allows for multiple labeling of objects in a video.

References

[1] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features (surf). 110(3):346–359, June 2008. 1

[2] S. Hinterstoisser, O. Kutter, N. Navab, P. Fua, and V. Lepetit. Real-time learning of accurate patch rectification. 2009. 1

[3] M. Merler, C. Galleguillos, and S. Belongie. Recognizing groceries in situ using in vitro training data. In *CVPR*, 2007. 1

[4] M. Ozuysal, M. Calonder, V. Lepetit, and P. Fua. Fast keypoint recognition using random ferns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009. 1

[5] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. pages 511–518, 2001. 1