Hand-to-Target Guidance Using Multi-Object Tracking

Elise Newman
Department of Computer Science
University of California, San Diego
La Jolla, CA 92093
eanewman@ucsd.edu

Abstract

The GroZi project at UCSD is working toward a complete system that can guide a visually impaired person around a store to collect the items on their shopping list by solving a wide array of computer vision problems. I plan to simultaneously track both the product and the hand in order to produce either sonic or haptic guidance toward the targeted product. I will assume that the person using the system is (a) in the correct aisle, approximately one meter away from the desired product, and (b) looking for an item that is on a list of previously selected items. I will begin with a frame zero that contains (a) a bounding box around the targeted product, generated by one of the abovementioned detectors, and (b) the outstretched hand. This system will be integrated with other GroZi components, including text recognition "in the wild" which will aid aisle sign processing, and product detection.

1. Project Description

The goal of this project is to simultaneously track the hand and the targeted item in order to periodically update sonic or haptic guidance. I will be using Camshift [2], an adaptation of the mean shift algorithm that employs continuously adaptive probability distributions, as a starting point for my system. Since this algorithm was designed to be used as part of a head and face tracker, it only utilizes the hue channel of the HSV colorspace. Modifications to Camshift were employed by [1] and [3] in order to afford more generalized object tracking. Such additional techniques include multidimensional histogram back-projection and using weighted histograms which assign a higher weight to pixels toward the center of the search window. Problems of confusion relating to backgrounds that were similar to the tracked objects were mitigated by assigning lower weights to color features that belong to the background. After the initial challenge of parallelizing two trackers, for the hand and for the product, I plan to investigate how best to solve the problems that crop up as the result of adding clutter by testing the abovementioned techniques. There will be several stages of feedback, starting with the most basic (perhaps a quad-directional text-to-speech output), and culminating in much more fine-grained feedback.

2. Milestones

- Week 1: Track a hand without background clutter, and output its coordinates
- Week 3.5: Integration of hand and product tracking, outputting coordinates of each and simplified guidance
- Week 5.5: Get the system working with background clutter
- Week 7: Explore adaptive background modeling techniques proposed by [1] and [2].
- Week 9: Refine feedback, provide more detailed information (hopefully develop framework that would allow for easy integration of glove-contained haptic feedback system)
- Week 10: Wrap up project and write results

3. Questions

Some questions I intend to answer throughout the duration of this project are:

- How can two Camshift-based trackers be integrated into one smooth tracker?
- How can I effectively guide the hand to the item with information obtained from the joint trackers?
- How often should I update the feedback?

4. Software

I will be using OpenCV for implementations of CAMSHIFT for tracking and perhaps optical flow for camera stabilization.

5. Qualifications

I am a 5th year Computer Science major. I have taken classes covering various topics in artificial intelligence, (CSE 150), Machine Learning (CSE 151), computer vision (CSE 152), and image processing (CSE 166).

References

[1] J. G. Allen, R. Y. D. Xu, and J. S. Jin, Object Tracking Using CamShift Algorithm and Multiple Quantized Feature Spaces. *Proceedings of the Pan-Sydney area workshop on Visual Information Processing*, ACM International Conference Proceeding Series Vol. 100 (Australian Computer Society, Inc., Darlinghurst, Australia, 2004).

[2] G. R. Bradski, Computer Vision Face Tracking for use in a Perceptual User Interface. Intel Technology Journal, 2nd Quarter, 1998.

[3] R. Stolkin, I. Florescu, and G. Kamberov, An Adaptive Background Model for Camshift Tracking with a Moving Camera, in Advances in Pattern Recognition, pp147-151, 2007.