CS231A Course Project Proposal

Phumchanit Watanaprakornkul
Stanford Univerisity
Stanford University. Stanford, California 94305
yiam@stanford.edu

Abstract

One problem in supervised learning is that we need a lot of data in order to get a reliable classifier. In object recognition, there is a way to reduce the size of manually labeled object needed by using track information of an object. This semi-supervised learning method using track information has already been done on data from laser range finder. The goal of this project is to extend such result to camera-based system. We could do background subtraction and segmentation on frames and then applied the aforementioned semi-supervised learning algorithm.

1. Introduction

Supervised learning object recognition algorithm has one main weakness: the need for large amount of labeled training data. One way to remedy the weakness is to generate more labeled data from a few manually labeled ones. We can label object of one type in one frame then use a model-free segmentation and tracking to find the place for the same object in the next frame. We then label the object in the next frame of the same type.

The idea of using tracking information to generate more labeled data for supervised learning is already done in [1]. However, the data from LIDAR sensor used in [1] is very different from frames that we get from normal camera. We cannot apply exact same process because the segmentation and track classification will be very different. In addition, camera is cheaper than LIDAR sensor.

2. Dataset

Our dataset is a video of overhead view from Hoover tower to the fountain in front of it. We are interested in recognizing pedestrians and bikers. We label a few objects in a few frames of the video for training data.

3. Algorithm

There are mainly four steps for doing semi-supervised learning method using track information

3.1. Background subtraction

We start by subtracting background from every frame. We will first try openCV BackgroundSubtractor.

3.2. Image Segmentation

We then have to divide the image into segments of possible objects. This model-free segmentation is the main part of the project. There are many existing methods for this part. We will have to find the one that work best together with our tracking part.

3.3. Track Classification

In this step, we add the tracking information into each segment in the frame, determining that this segment come from which segment in the previous frame. Since we know that the object in a track is obviously of the same type, we get more labeled training data for object classifier. We will mainly use Kalman filter to get track information.

3.4. Object Classification

We then use the extra training data labeled by tracking information to train our usual object classifier. This part is not the main point of the project. We probably use open source SVM classifier on HOG features.

4. Evaluation

The goal of our experiment is to label only a few objects (about ten labels from our 10 minutes video with pedestrians and bikes as our objects of interest). Then, show that using only those few manually labeled data, our algorithm produce similar accuracy to common object recognition algorithm. We probably test the accuracy by eye at first (let the classifier draw a box on recognized object in each frame). Hopefully, we could manually label some part of the video as test data. (For test data, it's more work since we need to label everything in every frame.)

References

[1] Alex Teichman, and Sebastian Thrun. Tracking-based semi-supervised learning.

[2] Thanarat Horprasert, David Harwood, Larry S. Davis. A Robust Background Subtraction and Shadow Detection

[3] P. Felzenszwalb, R. Girshick, D. McAllester, D. Ramanan. Object Detection with Discriminatively Trained Part Based Models. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 32, No. 9, September 2010

5. Appendix

This project extends the result of tracking-based supervised learning done on LIDAR data[1] to data from normal camera.

Future Distribution Permission

The author of this report gives permission for this document to be distributed to Stanford-affiliated students taking future courses.