Tracking-Based Semi-Supervised Learning using Stationary Video

Andrew Chou
Stanford University
akchou@stanford.edu

Abstract

This paper deal addresses the semi-supervised problem of tracking and recognizing objects in videos taken with stationary cameras. Building on work on Stanford's autonomous vehicle using laser range finders to solve the same problem, we aim to develop accurate methods for classifying objects without the additional benefit of 3D laser scans. We set out with three main goals, each building on the previous ones. The first is to perform background subtraction to remove all background objects (those objects that are stationary in the frame of the camera). The second is to track the foreground objects through every frame of the video. Finally, the third goal is to use semi-supervised methods to classify tracked foreground objects. A successful semi-supervised approach will greatly reduce the amount of training data needed for many classification problems.

1. Introduction

The objective of this paper has three subgoals:
(1) Remove background objects.
(2) Track foreground objects.
(3) Classify foreground objects using semi-supervised learning. The combination of these three goals will allow us to train a complex classifier with very little manually labelled data.

1.1. Removal of Background Objects

The background removal stage requires a video, or sequence of images, as input and outputs a binary mask for each frame or image. The ones in a mask (displayed as white in this paper) represent foreground pixels in that mask, while the zeros in a mask (displayed as black in this paper) represent background pixels in the mask. In this context we consider foreground to be any objects that are capable of movement outside of some fixed region that is roughly the size of the object. For example, humans, bicylists, automatic vehicles, and animals are some of the things that are considered to be foreground objects. Conversely, most inanimate objects such as buildings, plants, benches, and poles should ideally be classified as background. More subtle objects obejcts that should be classified as background are bodies of water, fans, fountains, and trees swaying in the wind. Distinguishing these moving background objects from real foreground objects is at the heart of the problem.

1.2. Foreground Object Tracking

The goal of the foreground object tracking step is to take the foreground masks from the background removal step and determine which foreground objects in each mask correspond to objects in other masks. The output should contain labels for each object as well as an outline for the object at each frame. Ideally the foreground object tracking step should be robust to overlap of objects, false positives (Labelling part of a fountain as foreground, for example.), false negatives (failure to detect a foreground object for a few frames), extrance and exit of objects during the sequence, as well as total number of objects needing to be tracked.

1.3. Semi-Supervised Classification

Classification will be done in a semi-supervised way using methods already developed in a previous paper by Teichman and Thrun [4]. Results and classification accuracy will be quantitatively evaluated using a data set provided by Alex Teichman. We realize that it is very important to find a good data set to evaluate our work. Working with Alex Teichman on this issue will be a priority in the coming weeks.

2. Proposed Method

Each of the three steps in our Semi-Supervised learning method build on previous work.

2.1. Background Subtraction

Standard methods for modelling the background include a Mixture of Gaussians method. OpenCV's implementation of Mixture of Gaussians for background subtraction does reasonably well when tested on a scene with swaying trees and fountains [2]. However there are a significant number of spurious points in the tree and fountain that are not adequately modelled by the Mixture of Gaussians. Sheikh and Shah use a Bayesian Modelling method to reduce the error in the subtraction [3]. Their method handles cases, such as the one with trees fountains, with non-stationary backgrounds. After qualitatively evaluating the results of both the Mixture of Gaussians approach and Shah's Bayesian Modelling method, we found that Shah's approach yields many fewer spurious foreground points. Thus we plan to use Shah's method as the background subtraction step of our algorithm. However, once all pieces of steps are complete we will be able to make a final decision by comparing the two methods quantitatively.

2.2. Foreground Object Tracking

Foreground object tracking will initially be done using a variant on the K-Shortest paths method developed by Berclaz *et al*. [1]. Berclaz *et al*. assumes the use of an appearance model to help track objects. As we do not have such a model, and are in fact trying to train a similar classifier, we will use our background subtraction method to give us probabilities of a foreground object at each location and time. In order to solve the overlap problem we will use a low pass filter on the location so that two intersecting paths can still be made continuous by linking them through a lower probability region. If the variant on *et al*. [1] is not sufficient then we will investigate further methods.

2.3. Semi-Supervised Classification

As mentioned above classification will be done in a semi-supervised way using methods already developed in a previous paper by Teichman and Thrun [4].

3. Preliminary Results

So far we have simply used OpenCV's implementation of Mixture of Gaussians to perform background subtraction [2] and compared it against our own implementation of Shah's Bayesian Modelling method [3]. Shah *et al*. performs significantly better as can be seen in figure 1.

References

[1] J. Berclaz, F. Fleuret, E. Turetken, and P. Fua. Multiple object tracking using k-shortest paths optimization. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PP(99):1, 2011.

[2] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.

[3] Y. Sheikh and M. Shah. Bayesian modeling of dynamic scenes for object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27:1778–1792, 2005.

[4] A. Teichman and S. Thrun. Tracking-based semi-supervised learning. In *Robotics: Science and Systems*, Los Angeles, CA, USA, 2011.

4. Appendix

The work in this project builds Alex Teichman's work (advised by Sebastian Thrun) that performs the the same object tracking and semi-supervised classification, but while using a laser range finder. This project aims to perform both steps with only video input.

1. The computer vision components of this project include:
   (a) Background removal
   (b) Clustering and tracking of foreground objects
   (c) Semi-supervised Classification
2. I personally plan to contribute:
   (a) Ideas and implementation (or use of a library like OpenCV) for background removal
   (b) Implementation and testing of object tracking algorithms
   (c) Design and analysis of quantitative experiments
3. Andrew Chou is enrolled in CS231A, and is the sole author of this writeup.

Figure 1. The original frame.

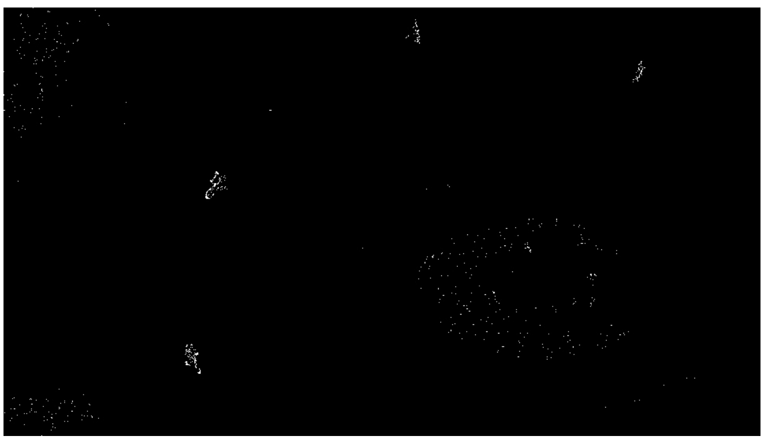

Figure 2. The foreground mask found using OpenCV's Mixture of Gaussians model.

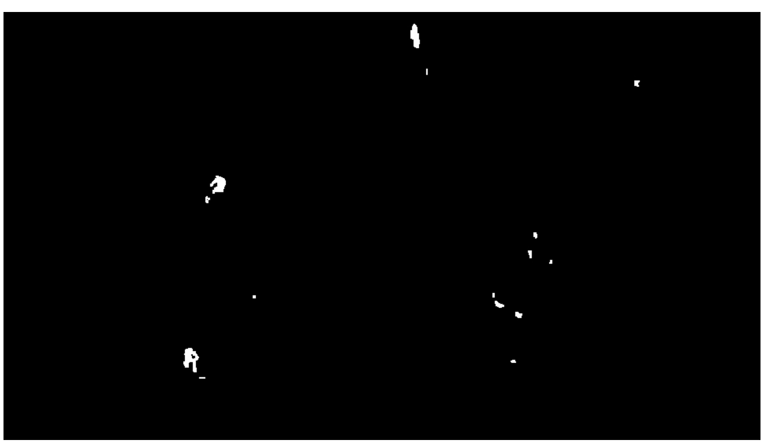

Figure 3. The foreground mask found using Shah *et al*.