Object Detection Using Segmented Images

Naran Bayanbat Stanford University Palo Alto, CA naranb@stanford.edu

Abstract

Object detection is a long standing problem in computer vision. One of the common approaches to object detection is to use segmented images to train. Intuitively, isolated foreground images should provide better training sets and improve the performance of the detection system. However, in practice, there are challenges associated with using segmentation for training data. Features located along segmentation borders in the training image tend to incorporate the clean background of the image, which leads to suboptimal detection performance within noisy background images.

In this paper, we purpose a way of excluding such features in order to obtain a generalized object detector that can perform cluttered background images by training on segmented images. We demonstrate the effectiveness of this approach by comparing the performance of our detector against that of a detector trained using ordinary, "dirty" background images.

1. Introduction

Object detection is one of the oldest problems in computer vision. One of the simplest ways to approach this problem is to training on a set of closely cropped images of the object. However, this approach has the drawback of including background pixels that are not necessarily a relevant part of the object itself. The image noise tends to degrade the performance of the detector. An alternative to this approach is to use segmentation to tightly crop out the object in the training set images and effectively remove the background clutter. As long as a set of clean images were chosen, it is possible to achieve a tight, accurate segmentation using the state of the art segmentation methods.

Unfortunately, there are also drawbacks associated with using segmented images. Features that are detected along the frame of the object, on its outer edges and corners, detract from the performance of the detector. Training on these features tunes the detector to clutter-free backgrounds. When applied to normal, noisy environment images, the Jason Chen Stanford University Palo Alto, CA jasonch@stanford.edu

detector performs poorly, even compared to a classifier that was trained on non-segmented images.

Our approach is to identify features (SIFT descriptors) that are located around the perimeter of the object segment, and exclude these from the training data. We hypothesize that removing the detracting features will improve the overall performance of the detector.

For the training set, we will use ImageNet to find clutter-free pictures of the object to train on. ImageNet has pre-categorizes images into "synsets," or semantic categories, that averages about 1000 images per category. A significant portion of the images on ImageNet are clean background images that will allow us to segment out the object easily.

We will evaluate our performance by training a SVM classifier on the modified set of features, and measuring the precision and recall rates of detection on a test set images. Additionally, we will also evaluate the performance of a classifier trained on "dirty", or non-segmented images, and provide plots of the performance metrics of both detectors for comparison. We hypothesize that our modified detector will outperform the "dirty" classifier.

1.1. Related Work

For the task of object detection, bag-of-features methods have been successfully applied in many instances. These approaches reduce an image into a collection of local features without preserving the geometrical structure of the underlying objects. Their application has largely been successful, allowing them to outperform more sophisticated methods that preserve the structure of objects [1, 2, 3]. However, they carry limited descriptive data, and are unable to segment an object out of its background. By first applying segmentation, then excluding the "bad" features, our approach seeks to improve on the bag-of-features methods.

Future Distribution Permission

The author(s) of this report give permission for this document to be distributed to Stanford-affiliated students taking future courses.

Comment [M1]: PLEASE INCLUDE THIS SECTION AT THE END OF YOUR REPORT.

We ask the given statement be present in your paper before submission. In case you would prefer that your project report is not posted on the course website, please email us in advance on the staff list.

References

- S. Lazebnik, C. Schmid, and J. Ponce. Beyond bag of features: Spatial pyramid matching for recognizing natural scene categories. In *Proc. CVPR*, 2006.
- C. Wallraven, B. Caputo, and A. Graf. Recognition with local features: the kernel recipe. In *Proc. ICCV*, volume 1, pages 257–264, 2003.
 J. Willamowski, D. Arregui, G. Csurka, C. R. Dance, and L.
- [3] J. Willamowski, D. Arregui, G. Csurka, C. R. Dance, and L. Fan. Categorizing nine visual classes using local appearance descriptors. In *ICPR Workshop on Learning for Adaptable Visual Systems*, 2004.

2. Appendix

This project is part of the ImageNet research effort in providing a large-scale image database for researchers and educators around the world. The detection system developed in this paper aims to improve the tagging of relevant portions of the images in each semantic category ("synsets").