Object Recognition and Computer Vision: Final Project
Object Category Detection and Localization

Antoine Labatie
antoine.labatie@gmail.com

Abstract

In this assignment we deal with object category detection and localization using GentleBoost which is a simple Boosting variant to train a detector. The types of weak classifiers are simple stumps while the training and detection are made with more robust method by working with filtered images.

1. Work done

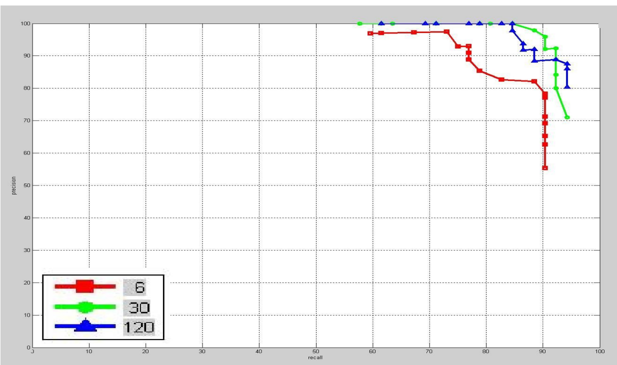
1.1. Overview of the work

In this assignment I worked on the improvement of the detection by implementing a multi-scale version. The idea is to perform the same strategy and threshold at different scales by up-sampling or down-sampling the image.

1.2. Choices of parameters

Due to the long time for training and detections phases I restricted to specific parameters to test the extension. I decided to use the pre-trained detector since it had 200 training examples which is very long to obtain.

Concerning the number of weak learners, I used generally 6 in my tests since the curve exhibits the same trends for 6, 30 or 120 learners as it is shown here.



The more classifiers the better the performance but we see that after 30 classifiers the detection rate improves slowly.

1.3. Details of the extension

I implemented the multi-scale in the *runDetector* function by running the *singleScaleBoostedDetector* function with different image sizes. I used 4 Scale Steps of 0.8 down-sampling, and started with an up-sampled image of a factor 1.2. This led to versions of the image of scale 1.2, 1.02, 0.86 and 0.74. Due to the scale factor I had to rescale the bounding boxes after each scale step.

Yet it produces redundancies since an object may be detected at different scales so I implemented a non maxima suppression by looking at the overlapping between bounding box. In the case it was over 0.5 for the smaller box I removed the bounding box with the lower score.

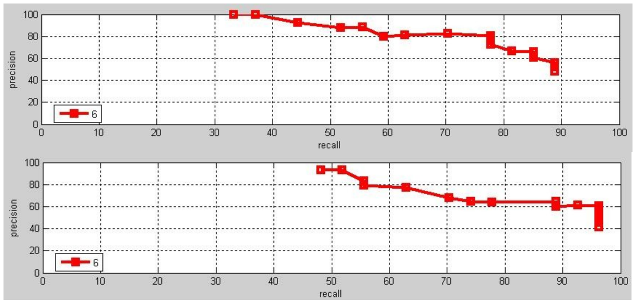
So with this version we have better chances to find objects which have bigger or smaller size than the average. And this is indeed what we observe by founding higher recall rates and observing it on specific examples.

This example shows the detected object for 3 different thresholds of 0, 40 and 80 for the classifier function, for a single-scale version (top) and the multi-scale (bottom). For a low threshold the object is also detected in the single-scale version. Yet since the object is smaller than the average training size, it is detected with a low score which is not the case anymore with a multi scale algorithm. By looking at the size of the bounding box we see indeed that it has been detected in an up-sampled version of the image.

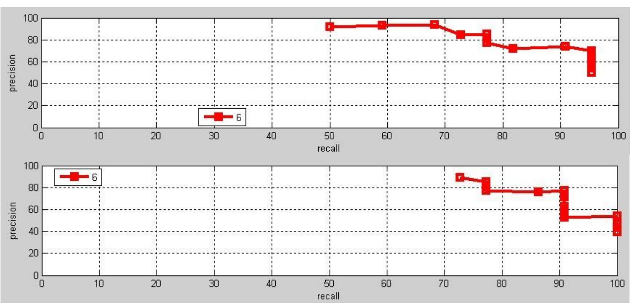
2. Evaluation of the extension

2.1. Quantitative evaluation

We present here the precision-recall curves obtained with 50 test images and 6 weak classifiers. The results are shown: basic detector on top, extended detector on bottom.



Precision-Recall curves for ‘car’



Precision-Recall curves for ‘screen’

The curves cannot be very properly compared since the percentage of overlapping in term of recall rate is limited. Indeed the extended detector has higher recall rate for given threshold, yet it has also a lower precision.

To have a fair comparison we need to look at the common parts in terms of recall and we see that performance depends on the object category. For the ‘screen’ category the extended detector performs better for recalls between 70% and 90%. Yet on the ‘car’ category it seems there is a loss of precision between 55% and 80% recall rates. However it has a better precision in small regions (less than 55% and more than 80%).

For both versions there is a saturation (for the lowest thresholds) where we cannot anymore increase the recall but only lower the precision.

2.2. Qualitative evaluation

This complex pattern can be explained by looking at the modification added in the extension. Since we look at more scales we have better chances of having false alarms which lowers globally the precision. On the other, by merging overlapping windows we increase the precision since we avoid multiple bounding boxes on the same object. There is a similar attempt in the function *singleScaleBoostedDetector which* clusters centers which are not far apart but it is less effective as we see below.

This example (basic detector on top and extended on bottom) shows that we may have a better recall by finding more bounding boxes which are finally merged which is not the case in the basic detector.

Those examples favor the extension but we have seen we may have a noticeable loss of precision due to the false alarms. For the ‘screen’ category it seems uniform regions lead easily to false positives as we see below. And it is easy to find uniform regions on up-sampled images, which correspond to smaller regions in the normal image.

Concerning the recall rate, we find more objects when they have variations in size. Even though there are not so much variations in the dataset and objects can be detected at low threshold, small variations affect the score obtained and thus the detection for higher thresholds as illustrated in Section 1.

2.3. Further improvements

The improvement may depend on situations but we can think of ways to make it more stable. We may take advantage of the recall improvement while not diminishing the precision by adding other detectors behind it (for example a similar detector based on HOG).

We may also use a prior knowledge on the images, which could be that there are not a lot of object instances in the image (for example less than 2 or 3) which would increase the precision.