Histogram of Oriented Gradients and Support Vector Machines for Face Detection

Javier Perez
Universidad de los Andes
Bogota, Colombia
je.perezl@uniandes.edu.co

1. Introduction

Face detection is a problem in computer vision that has been studied for several years. In 2001, Viola and Jones [1] developed an algorithm that recognize frontal faces in real-time, and this work has been recognized as one of the first one in his type. Since then, many approaches to tackle this problem has been proposed. For example, face detection algorithms can be encompassed in feature-based, template-based and appearance-based [2]. In this laboratory, we focused on the appearance-based technique.

2. Materials and Methods

In this laboratory, we considered a part of three datasets: Caltech 10,000 Web Faces, SUN database and MIT+CMU test set. The Caltech 10,000 Web Faces contains photos of people obtained from Google Images [3]. The images appear in different setting, for example, group of people, single-person or portrait. The position of the eyes, nose and mouth were provided as the groundtruth. From this dataset, some faces were cropped and stored as an individual image, and resized to 36x36 pixels. These individual face images were used in the training stage and this resulted in 6,713 cropped faces.

From SUN database, we used several images as the non-faces scenes [4]. This dataset contains a variety of images, like outdoor, animals, cities, in different contexts. Some random cuts of these images were used as the negative example of faces in the training stage. The MIT+CMU is a common benchmark for face detection. This dataset contains 130 images with 511 faces [5]. The images are in grayscale, and contains faces at different scales, for example, because of perspective.

The algorithm lies on Histogram of Oriented Gradients (HOG) and Linear Support Vector Machines (Linear SVM). For obtaining better results, a multi-scale face detector is design. In the training phase, we get the HOG features for the face and non-face images. Then, we trained a Linear SVM where the +1 data correspond to the face images, and the -1 data to the non-face images. Next, from the Linear SVM, we learned a model that can be summarised in the normal vector w and the bias b. In the test phase, we used w and b in a linear equation to get the confidences of every data in the test images, i.e., patches from images at different scales. The multi-scale face detector consists in resizing the test image to a particular scale, and running the detector in the HOG space with a sliding window of step size of 1, this last parameter remains constant in all the experiments done.

3. Results

First, we maintained constant the HOG template size, the number of examples for the non-face, the confidence in the prediction with the SVM. We said that the default HOG template size and cell size is 36 and 6, respectively, the number of examples for the non-face is 10000, the confidence in the prediction with the SVM is -0.2.

In this part, we evaluated the influence of and the scale, the data augmentation andt he HOG cell size. For the scale, we considered 10 scales that went from 100% of the image's size to 10%, in step of 10%. We also considered 3 scales, i.e., 100%, 50% and 10%, and a single scale at 100%. Figures 1, 2 and 3 show the PR (Precision-Recall) curve of the single, three, ten scales detector, respectively. After that, we added more scale from 200% of the image's size to 10%, in step of 10%, and compared its PR curve, as shown in Figure 4, with the ten scale detector.

We noticed that the multi-scale detector improved the Average Precision (AP) over a single-scale detector. We obtain that a detector with 10 different scales gives the better AP. This detector is more robust than a single-scale detector because it can find faces from different sizes. From the single-scale detector PR curve, we saw that the algorithm can find only certain faces. In particular, the maximum recall that it obtained is near 0.55, this means that 45% of the faces are not detected, i.e., they have different scales. If we consider three scales, we can improve the AP from 0.386 to 0.698 and the maximum recall from 0.55 to 0.85. This

result suggests that there are three different size of faces in this test set that correspond to a little more of 2/3 of all the faces' sizes. Also, with ten different scales we got a AP of 0.867 and a maximum recall of 0.95.

For the training stage, we evaluated the influence of the positive samples number for the SVM. We duplicate this number by performing data augmentation with the same images. In this case, we considered the reflected image (by the y-axis) of each one. We compared the PR curve obtained previously with the 3 and 10 scales with the same scales detectors after data augmentation in the SVM training. Figures 5 and 6 shows the RP curve of the 3 and 10 scales detector with data augmentation. We improve the AP comparing with the previous data used for training. This tool can enhance the predictor doing it more robust to face orientations.

We evaluate the effect of changing the HOG cell size, in the 3 and 10 scales with data augmentation. We consider a HOG cell size, half of the previous size. Figure 7 and 8 shows the RP curve of the 3 and 10 scales detector with data augmentation and HOG cell size of 3. The HOG visualisation, see Figures 12 and 13, shows more details to the positive detector, like fine-grained information of the faces. With the 10 scales detector, data augmentation a a HOG cell size of 3 we get a AP of 0.930 and a maximum recall of 0.975. Table 1 summarizes the different results.

Figures 9 and 10 show qualitative result of the multi-scale detector that performs better results in the test set (AP of 0.930). In both cases, the detector found all the faces that were in the images, even though there were different scales and orientations. Figure 14 is the ROC curve of this detector, as in [1]. Figure 11 shows an image taken from the social media. In this case, the detector found 9/17 faces. We consider that probably the clutter disposition may affect but we saw similar images on the test set.

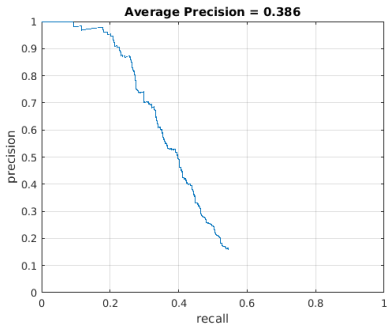

Figure 1. Precision-Recall curve and the Average precision with a single scale detector, with the default HOG features, number of examples for the non-face, confidence in the prediction with the SVM.

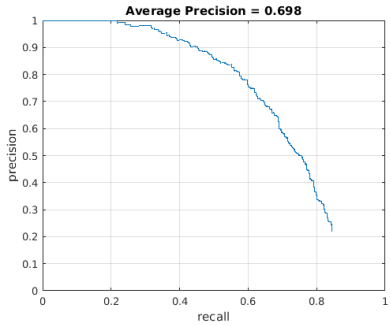

Figure 2. Precision-Recall curve and the Average precision with three scale detector, with the default HOG features, number of examples for the non-face, confidence in the prediction with the SVM.

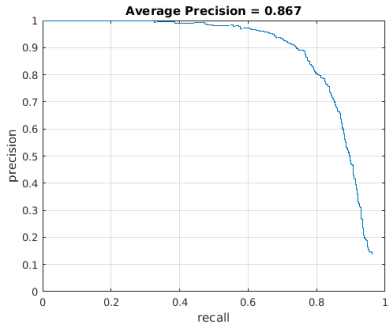

Figure 3. Precision-Recall curve and the Average precision with ten scale detector, with the default HOG features, number of examples for the non-face, confidence in the prediction with the SVM.

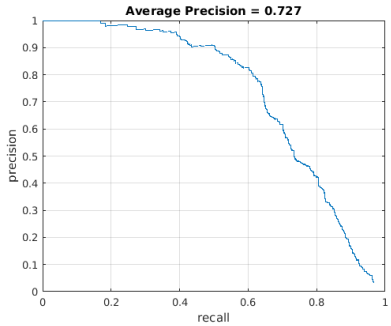

Figure 4. Precision-Recall curve and the Average precision with twenty scale detector, with the default HOG features, number of examples for the non-face, confidence in the prediction with the SVM.

References

[1] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, IEEE Comput. Soc.

[2] M.-H. Yang, D. Kriegman, and N. Ahuja, "Detecting faces in images: a survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34–58, 2002.

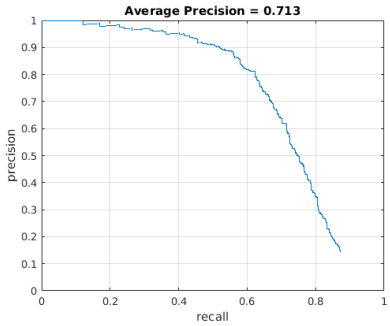

Figure 5. Precision-Recall curve and the Average precision with three scale detector and data augmentation, with the default HOG features, number of examples for the non-face, confidence in the prediction with the SVM.

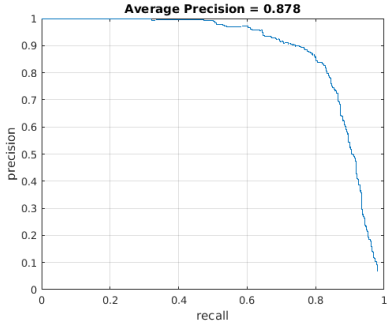

Figure 6. Precision-Recall curve and the Average precision with ten scale detector and data augmentation, with the default HOG features, number of examples for the non-face, confidence in the prediction with the SVM.

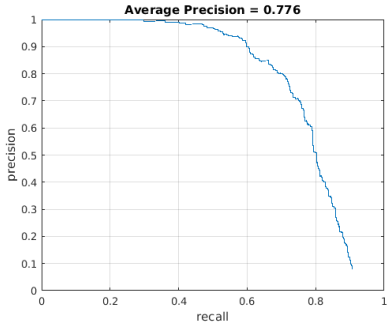

Figure 7. Precision-Recall curve and the Average precision with three scale detector, data augmentation and HOG cell size of 3, with the default number of examples for the non-face, confidence in the prediction with the SVM.

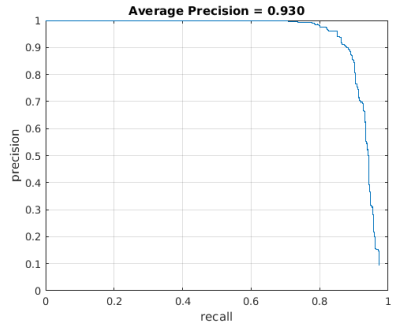

Figure 8. Precision-Recall curve and the Average precision with ten scale detector, data augmentation and HOG cell size of 3, with the default number of examples for the non-face, confidence in the prediction with the SVM.

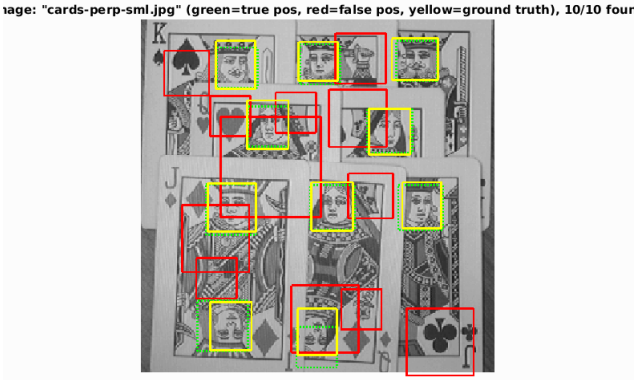

Figure 9. Example of the best face detector in the test set.

Figure 10. Example of the best face detector in an on-the-wild image with a good performance.

puter Vision and Pattern Recognition, IEEE, jun 2010.

[3] L. Fei-fei, R. Fergus, and P. Perona, "One-shot learning of object categories," *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, vol. 28, p. 2006, 2006.

[4] J. Xiao, J. Hays, K. A. Ehinger, A. Oliva, and A. Torralba, "SUN database: Large-scale scene recognition from abbey to zoo," in *2010 IEEE Computer Society Conference on Com-*

[5] S. Liao, A. K. Jain, and S. Z. Li, "A fast and accurate unconstrained face detector," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, pp. 211–223, feb 2016.

Figure 11. Example of the best face detector in an on-the-wild image with a bad performance.

Number of scales	HOG cell size	Data aug.	AP
1	6	No	0.386
3	6	No	0.698
3	6	Yes	0.713
3	3	Yes	0.776
10	6	No	0.867
10	6	Yes	0.878
10	3	Yes	0.930
20	6	No	0.727

Table 1. Summary of different results in the CMU-MIT test set.

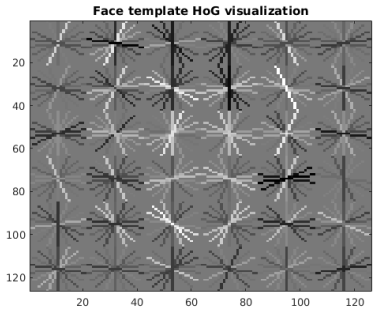

Figure 12. HOG visualization with cell size of 6.

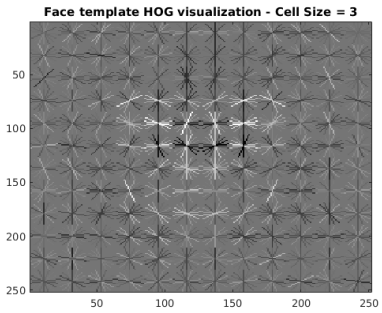

Figure 13. HOG visualization with cell size of 3.

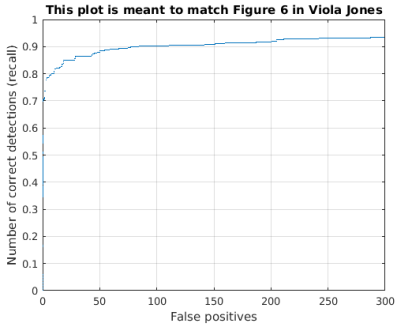

Figure 14. The ROC curve for the best method (multi-scale, data augmentation, HOG cell size of 3) as shown in Figure 6 of [1].