

Face Detection using multi-scale HOG

Laura Munar Acosta
Universidad de los Andes

l.munar10@uniandes.edu.co

Maria Ana Ortiz
Universidad de los Andes

ma.ortiz1@uniandes.edu.co

1. Introduction

Recently, problems such as object detection or image classification have received an increasing amount of attention in the computer vision community. Faces and human bodies are among the most important objects in images and videos. Therefore, face detection and human detection have attracted considerable attention in applications of video surveillance, biometrics, smart rooms, driving assistance systems, social security, and event analysis. Detecting humans in images is challenging due to the variable appearance, illumination, and background. The complexity of these problems is often such that an extremely large set of examples is needed in order to learn the task with the desired accuracy [2].

Histograms of Oriented Gradients (HOG) plus Support Vector Machine (SVM) is one of the most successful human detection algorithms. The HOG+SVM algorithm employs sliding window principle to detect humans in an image. It scans the image at different scales and at each scale examines all the subimages. In each subimage, a 3780-dimensional HOG feature vector is extracted and SVM classifier is then used to make a binary decision: human or non-human [1].

The success of the HOG+SVM human detection algorithm lies in its discriminative HOG features and margin-based linear SVM classifier. The HOG+SVM algorithm concentrates on the contrast of silhouette contours against the background. Different humans may have different appearances of wears but their contours are similar. Therefore the contours are discriminative for distinguishing humans from non-humans. It is worth noting that the contours are not directly detected. It is the normal vector of the separating hyperplane obtained by SVM that places large weights on the HOG features along the human contours. While most of the classification algorithms are based on the idea of minimizing the training error, which is usually called *empirical risk*, SVMs operate on another induction principle called *structural risk minimization*, which minimizes an upper bound on the generalization error. That is why the HOG+SVM detection algorithm works very well in test [3].

It is important to recognize that there are some hyperparam-

eters that are needed to be tuned in multiscale HOG. One of them is the size of the sliding window because it will determine if a face will be recognized by the detector. On the other side, the c parameter on the SVM algorithm is important because it will allow to the regression to fit more or less to the data. In this paper, other hyperparameters will be considered in order to see their influences.

One of the most recognized detectors founded in literature is the one implemented by Viola et.al [4]. The main highlight from their proposed algorithm is that is capable of processing images extremely rapidly, as they called it "real time face detection". This was achieved because of the introduction of a new image representation called "Integral Image" which is an intermediate representation for the image that allows to compute features in a faster way. This is one of the main difference between these algorithm and the one proposed in this study. Otherwise, they combine different classifiers in a cascade way that allows background regions of the image to be quickly discarded while spending more computation on promising face-like regions [4]. In the present study, a face detector is implemented using HOG+SVM algorithm. The influence of some hyperparameters is evaluated in the Caltech web faces dataset.

2. Materials and Methods

2.1. Dataset

Dataset used was Caltech web faces, that contained 6,713 cropped images by 36×36 for training set and 130 randomly sized images with lot of faces for test set, on Figures 1 - 2 examples from dataset can be observed.

2.2. Strategy

As HOG was described on previous section the strategy for obtaining better results was to evaluate the effect on the different parameters such as:

- **Lambda:** parameter on SVM (Support Vector Machine) was varied from 1 to 0.0001 (for VL feat functions lambda is the regularization parameter)
- **Confidence:** for multiscale hog with SVM trained was varied to found the lesser false negative



Figure 1. Images from dataset used for trained

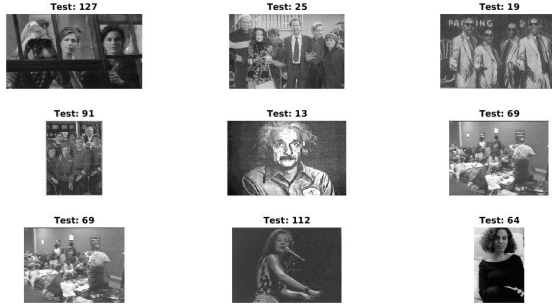


Figure 2. Images from dataset used for test

- **Pixel Step:** for sliding window on detector the step in which windows moves was detected
- **HOG cell size:** was varied to observed the effect on detection, knowing that it will varied the size of the features collected.

Additionally, it was also tested if using histogram equalization had a significant effect on detection. Multi-Scales for sliding window on testing part was set fixed, on a range of 1.2 – 0.1.

2.2.1 Final Model

From previous strategy the final model was defined taking into account the best Average Precision (AP) values, being for Lambda 0.0001, for Confidence initially 0.5 then experimented to be change, for Pixel Step 3 and for HOG cell size was 3.

3. Results

3.1. Variation of Lambda

Lambda of SVM was variated, Lambda parameter for SVM function its related to C parameter, results from variation it are shown on Table 3.1. It can be observed that for Lambda value 1 Average Precision (AP) is very low, and that lower values AP are very similar, with low variability.

Lambda	1	0.1	0.01	0.001	0.0001
AP	0.004	0.667	0.801	0.812	0.813

Table 1. Performance on dataset with variation of Lambda, values of confidence, negative samples, pixel step and HOG cell size were kept constant

3.2. HOG cell size variation

HOG Cell size variation results can be observed on Table 3.2, and it can be observed that on bigger cell size AP tends to be lower, however 3 size cell had a very similar behaviour as cell size 6.

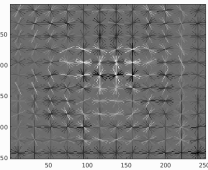
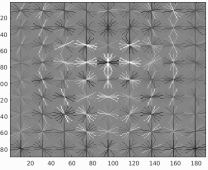
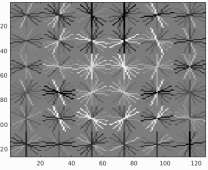
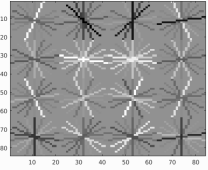
HOG Cell Size	HOG trained	AP
3		0.82
4		0.852
6		0.813
9		0.676

Table 2. Result on variation HOG cell size, values of confidence, negative samples, pixel step and Lambda were kept constant

3.3. Variation of Pixel Step

Results of variation of Pixel Step can be observed on Table 3.3, and an increase in AP can be observed as step lowers.

Step	8	6	4	3	2
AP	0.725	0.852	0.880	0.897	0.890

Table 3. Performance on dataset with variation of Pixel Step on detector, values of confidence, negative samples, Lambda and HOG cell size were kept constant

3.4. Confidence of the trained model

In table 3.4 are shown the results of the variation of the confidence of the trained model, where it can be seen that lower confidence results in greater average precision. The figure 3 presents the Precision-Recall curve for the best AP founded in table 3.4.

Table 4. Confidence value variations and its corresponding performance

Confidence	0.25	0.5	0.75	0.95	1.25
AP	0.817	0.812	0.788	0.799	0.754

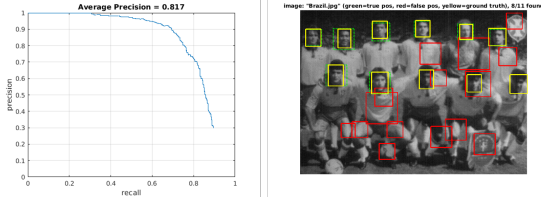


Figure 3. Precision-Recall curve and qualitative results for the best AP founded for a 0.25 confidence

3.5. Final Model

Results of precision and recall curve for best model can be observed on figure 4

3.6. Examples on test images

Examples of performance on extra test images (not on test set) are shown on figures 5 and 6, and performance on original test set are shown on figures 7 and 8. For extra images it is observed how some non-faces objects are marked as faces and mostly all faces are found. And on original test set examples shown some are missed and other are not missed but does not match exactly with annotation.

Additional Results Examples can be found on section 5

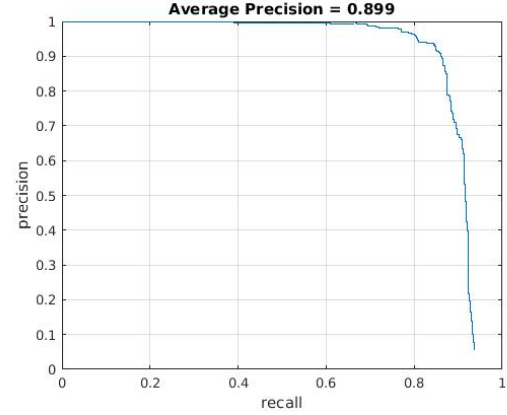


Figure 4. Precision and Recall Curve



Figure 5. Result on Extra test images

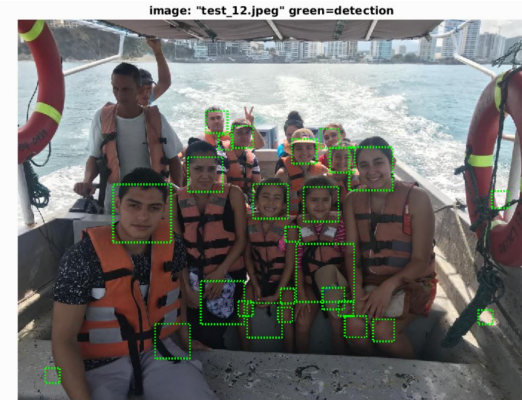


Figure 6. Result on Extra test images

3.6.1 Example Results

4. Discussion

4.1. Variation of Lambda

The effect of Lambda on detection as observed on Table 3.1 mainly happens because as Lambda is SVM regu-

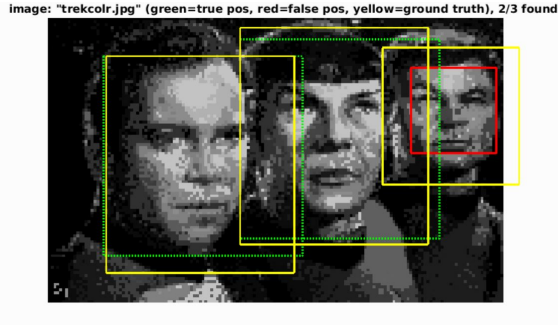


Figure 7. Result on Original Test images

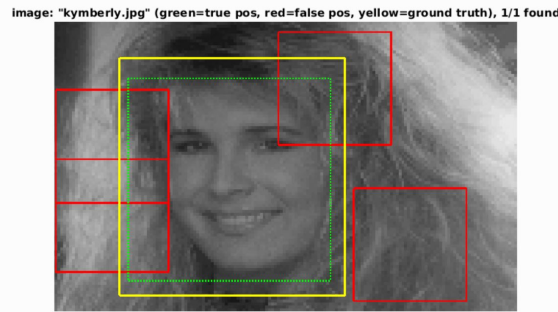


Figure 8. Result on Original Test images

larization parameter it allows some error on classification, its behaviour usually depends on the data, which for these problems it is observed that value of 0.0001 is the best fitted. These means some error is allow in detection on the SVM hyperplane.

4.2. Variation of HOG Cell Size

On Table 3.2 the results observed on varying the HOG cell occurs mainly because as train images are 36x36 bigger cells tend to ignore more detailed information of shape and not characterize good enough the face so detector has a poor performance. Additionally for a value of cell size 3 detector develop very similar to size 6, which indicates that probably because of the size of image it took irrelevant shape information. Best cell size performance was for cell size 4, which as it can be observed on the table the learned HOG descriptor has a more defined face appearance.

4.3. Variation of Pixel Step

On table 3.3 the results from changing the pixel step are listed, and as already mentioned the smaller the step it tends in very small magnitude to increase the Average Precision, and on the opposite bigger steps decrease it. The main issue is that computational time increases for smaller pixel steps and for this problem AP did not had a significant increase, mainly because step only changes amount of candidates per image or how is the sliding window detecting on image,

that is why very high values ignore true positive candidates lowering AP.

4.4. Confidence of the trained model

As it can be seen on table 3.4, a lower confidence results on a greater AP. This results are expected because we are selecting all the results above an specific confidence. The highest this number is, the lowest the AP because we are being more strict and we can be leaving some faces behind. In figure 3 it can be seen that as the recall increases, the precision decreases and this is explained because more precise results need a higher confidence value of the model and this can leave behind some faces affecting the recall. The explanation on why a lower number of confidence gives a better Average Precision is because the overall recall increases in a greater way than the precision decreases. That means that in a global performance, the new boxes that are being allowed with the lowest confidence are more face-like than background and that is increasing the overall AP. This hyperparameter is considered very important and should be always tuned in terms of the problem and the dataset.

4.5. Final Model

According to the figure 4, the average precision on the best model found in this study is 0.897. In this figure, it can be seen that the precision keeps steady for almost the entire recall region, until the recall goes above 0.8. Telling us that almost all of the instances selected by the detector were faces, even though it not predicted all the groundtruth faces. This implies that the detector has great quality selecting front faces but has some troubles in the selection of faces that are with a different orientation, thus decreasing the precision in order to increase the recall.

In order to improve the method, is highly recommended to use features that include different face orientations. Because that is one of the main limitations of this model. The faces can be oriented different and the trained model can only recognized front faces. In this way, the trained model can have more information about the shapes of a face and can produce a better outcome. For example, in figure 3 can be seen that some of the false negatives follows a pattern that is similar to a face in HOG representation. So, this false negative or false positive patterns are circular HOG representations and that is why are being considered as faces. In order to overcome this, another shape feature can be used to describe the face and in this way, even though the HOG feature recognize the instance as a face, another feature can predict that it is not face. It was found in literature that a Gaussian-like function is being used to account for the center bias and can work to differentiate a face from another instance [3].

On the other side, the computation of the dimensional HOG features in each detection window is time-consuming and this is another limitation of the model. Considering that in a scaled image there are a lot of detection windows, the total computational time is therefore very large. Suppose that the image size is 320×240 and the scaling factor is 1.1. Then there are totally 2000 detection windows to compute HOG. This is the bottleneck for a real-time human detection system. When multi-scale HOG is performed there is a large overlapping area for two neighboring depending on the window size. If the step between the two neighboring detection windows is proper, the two detection windows intersect with a lot of blocks and the HOG features of these blocks are the same. So independently computing the HOG features in two neighboring detection window is redundant. This redundant information should be reduced in future work. [3].

5. Conclusions

Face detection is a problem that according to our results can be addressed with HOG shape descriptor because it obtains good overall results not only over test set data of data-set but on extra test set.

Main problem of the algorithm proposed is classifier confidence mainly because its value is very sensible to change of parameters, as evidenced on main model that best confidence was not best for best model.

Finally, face detection errors occur mostly over data that have similar features to a face such as circular shape, or curve wrinkles that make classifier get false detections, on the future algorithm should have into account specific similar features for training. As future work, the computational time of the algorithm should be addressed.

References

- [1] A. J. Newell and L. D. Griffin. Multiscale histogram of oriented gradient descriptors for robust character recognition. In *2011 International Conference on Document Analysis and Recognition*, pages 1085–1089. IEEE, 2011.
- [2] E. Osuna, R. Freund, F. Girosi, et al. Training support vector machines: an application to face detection. In *cvpr*, volume 97, page 99, 1997.
- [3] Y. Pang, Y. Yuan, X. Li, and J. Pan. Efficient hog human detection. *Signal Processing*, 91(4):773–781, 2011.
- [4] P. Viola and M. J. Jones. Robust real-time face detection. *International journal of computer vision*, 57(2):137–154, 2004.

Annex

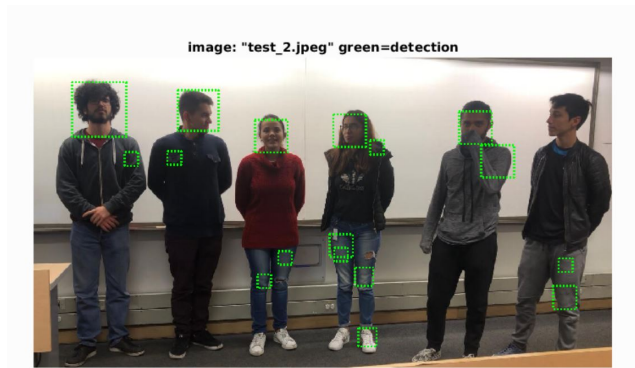


Figure 9. Result on Extra test images



Figure 10. Result on Extra test images



Figure 11. Result on Extra test images

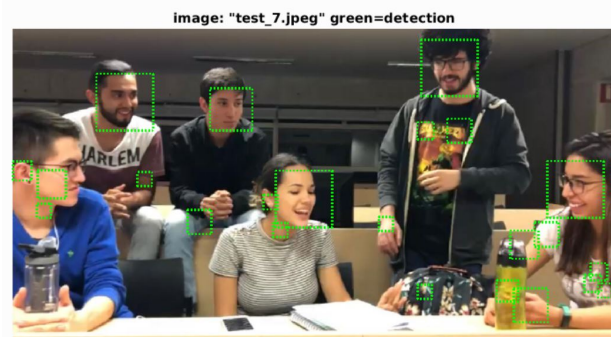


Figure 12. Result on Extra test images

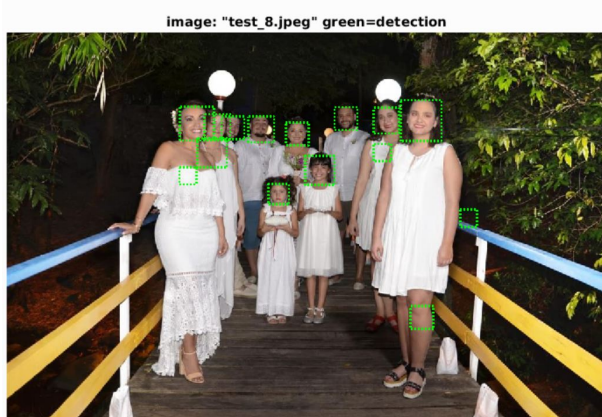


Figure 13. Result on Extra test images



Figure 14. Result on Extra test images



Figure 15. Result on Extra test images

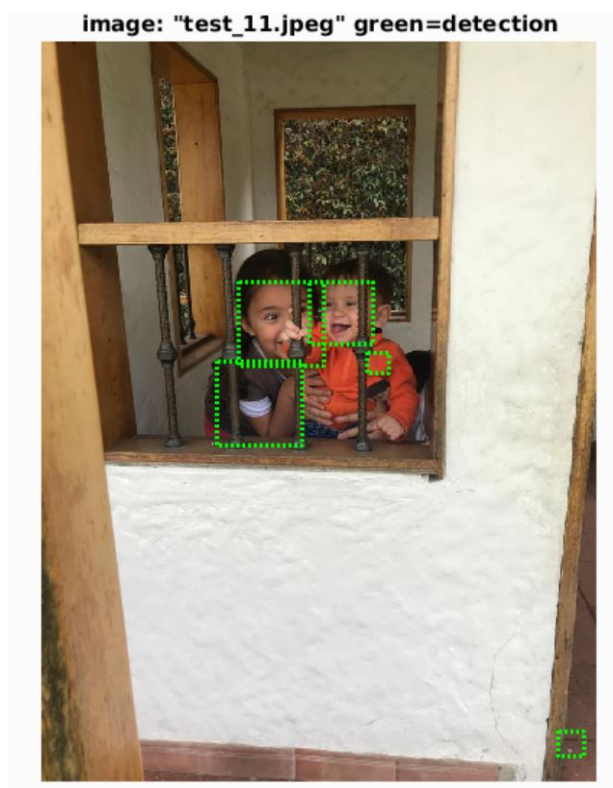


Figure 16. Result on Extra test images

image: "test_3.jpeg" green=detection



Figure 17. Result on Extra test images

image: "test_6.jpeg" green=detection



Figure 18. Result on Extra test images