Histogram of Oriented Gradients (HOG)

Natalia Andrea Durán Castro
Universidad de los Andes
na.duran@uniandes.edu.co

Ana M. Velosa Orduz
Universidad de los Andes
am.velosa@uniandes.edu.co

Abstract

This paper shows the use of the classifier at multiple scales to which its hyper parameters were modified to solve the problem of facial recognition in the Caltech database which has a performance of

1. Introduction

1.1. Dataset

In this paper was used the Caltech 10,000 Web Faces. The images in this dataset were collected from *Google Images*. The ground truth is a text file that contains the coordinates of the center of eyes, nose and mouth. In total, there are 10,524 human faces with different resolutions and settings [1]. However, in this work it is not going to be used all the images. In this case, the train set is conformed by 6,173 cropped images with dimensions 36x36. Also, it is going to be used a negative sample set that are images without faces. In the case of the test set, it is going to be used 130 random sized images with lot of faces. In the figure 2 is shown some images belonging to the data used in this paper.

1.2. HOG

Histogram of Oriented Gradient (HOG) it's a widely used descriptor that evaluates the distribution of the gradient density or the edges direction to be able to characterize the apparent object and shape. [2] HOG, divides the image in to cells,where the gradient direction histogram it's calculated. Once you have the histogram for each cell, you build the descriptor by making a combination of these histograms.[2] A multi-scale HOG consist on extend the histograms across the diferent scale of image structures, scale space.[3] The descriptor explain before it's useful to apply it to a detection problem because you are training your model with 36x36 images but when the model has to be tested it will have to work with images with larger dimension. For that reason having a multi-scale HOG could help to have a better performance on our algorithm. On the other hand, multi-scale HOG has hyper parameter such as the cell size or the step which provide the model with a greater amount of information regarding the appearance of the object being analyzed. This is because these two hyper parameters are those that divide the image into a determined cell size and indicate the number of cells that can fit in the image..

1.3. Viola-Jones

After reading the Viola-Jones paper, an extremely rapid face recognition algorithm , and compare it with the algorithm that we are implementing we could observe that both solutions to the face recognition problem didn't use the color information of the image, but they use the space information of the object that's is been analyzed.[4] Viola-Jones select relevant features from a huge library of potential features using the AdaBoost algorithm which selects different weak classifiers to be able to select the most relevant features of each image.[4] We, on the other hand use a multi-scale HOG to obtain the features of the images that we consider our group of positives and of the images that we consider our group of negatives in order to train a binary classification model where we can distinguish between face or no face.

Also, Viola-Jones combine what was said before with a complex classifier in a cascade structure that amplify the speed of the algorithm by focusing only on the important parts of the image, centering only on the parts that have faces. Another contribution of Viola-Jones work was a new image representation called Integral image representation that allowed fast feature evaluation that is similar to texture mapping.[4]

2. Methodology

After downloading the images from the database, we After downloading the images from the database, we proceeded to load and obtain the features through HOG of the two groups. These groups consisted of positives, faces of people and negatives, any object other than a face. These features were organized in a bidimentional matrix in which the rows correspond to the number of images of the group in which they are working and the columns correspond to the values obtained after performing HOG on each of the images of the group.

To create the bidimensional matrix, a reshape process of the features obtained from HOG was made taking into account that the positive features were all of the same size and that the negative features had a special reshape because the obtained HOG's had different sizes.

Finally, the model is trained with a linear classifier with the positive and negative examples and then a classifier at multiple scales was made for each of the images, suppressing the non maximums. These classifiers have as hyper parameters the size of the window, the step and the threshold

For the evaluation of our algorithm we use a ROC curve which is a graphic representation of the sensitivity versus the specificity for a binary classifier system as the discrimination threshold is varied, in other words, it is a comparison of two operational characteristics (TP and FP) as we change the threshold for the decision.

3. Results and Discussion

As for the images belonging to the group of positives, a variation of the size of the cell was made to observe how HOG behaved and could be observed as shown in 6 and 7 that with a value of the size of the smallest cell a higher level of detail could be obtained on the image that is being analyzed. Similarly, the size of the cell was changed in the images of the negative group and results consistent with those mentioned above were observed, as can be seen in 3,4 and 5.

As expected, a 100 percent accuracy in the training images was achieved while with the test images we find different accuracy. In the first approach we used a step of 6, a number of windows 10 and a threshold of 1, which were indexed by parameter in the function. With this we did not find any positive truth and had an accuracy of 0 percent. The results obtained with this method can be observed in 8

So, the algorithm was changed to one that used a multiscale detector based on a scale vector for the size of the windows of odd numbers from 0 to 1 in a step of 0.2. The average pressure obtained with this algorithm was 0.007 as can be seen in. It is believed that this is because the algorithm is recognizing a greater number of false positives, assigning windows of faces to zones of the image that does not present faces In the figure 11 is presented the best true positive of this algorithm, because the big window only was presented in the face. However, there are a lot of flase positives presented as little windows.

4. Conclusions

The limitation that we found in our method was linked to that we did not find a way to correctly assign the number of windows that should be selected per image, so a standard value was assigned for each one of them. This standard value generated a large number of true negatives, so the value of our average pressure is very low. On the other hand, no clear pattern was observed with respect to the true negatives, while the true positives follow patterns which can be observed 13, where the orientations of the gradient form a circular trace.

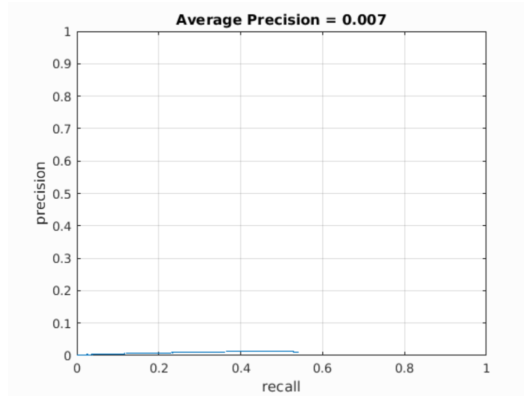

Figure 1. Precision-recall curve with an average precision of 0.007

References

[1] P. P. Anelia Angelova, Yaser Abu-Mostafa. Pruning training sets for learning of object categories. Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2005. 1

[2] M. R. Guedira, A. El Qadi, M. R. Lrit, and M. El Hassouni. A novel method for image categorization based on histogram oriented gradient and support vector machine. In *2017 International Conference on Electrical and Information Technologies (ICEIT)*, pages 1–5. IEEE, 2017. 1

[3] 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. 1

[4] P. Viola and M. J. Jones. Robust real-time face detection. *International journal of computer vision*, 57(2):137–154, 2004. 1

5. Annex

5.1. Images

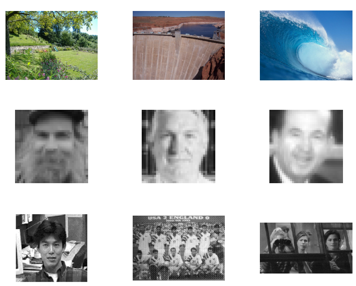

Figure 2. Examples of the images used in this article. The first row are part of the negative sample. The second row are part of the training images, and the third row are part of the test set.

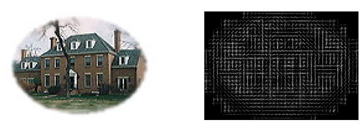

Figure 3. Anex 2. Image of a House with it's HOG with a cell Size of 3

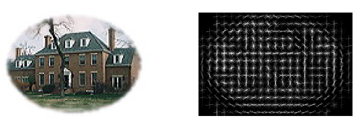

Figure 4. Anex 3. Image of a House with it's HOG with a cell Size of 6

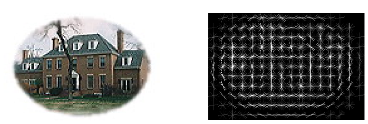

Figure 5. Anex 4. Image of a House with it's HOG with a cell Size of 3

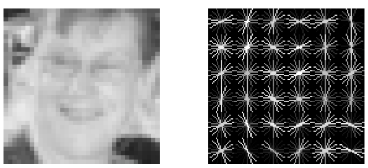

Figure 6. Anex 5. Image of a female face with it's HOG with a cell Size of 6

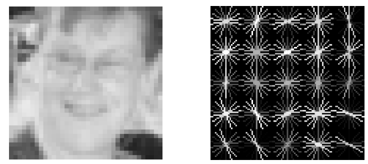

Figure 7. Anex 6. Image of a female face with it's HOG with a cell Size of 8

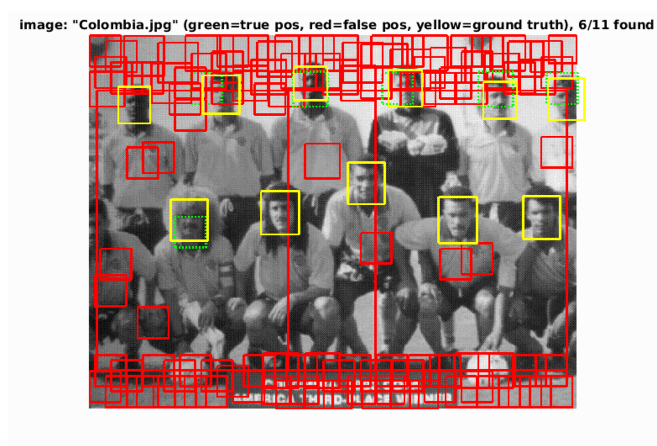

Figure 8. Anex 7. First algorithm with a step of 6, a number of windows 10 and a threshold of 1

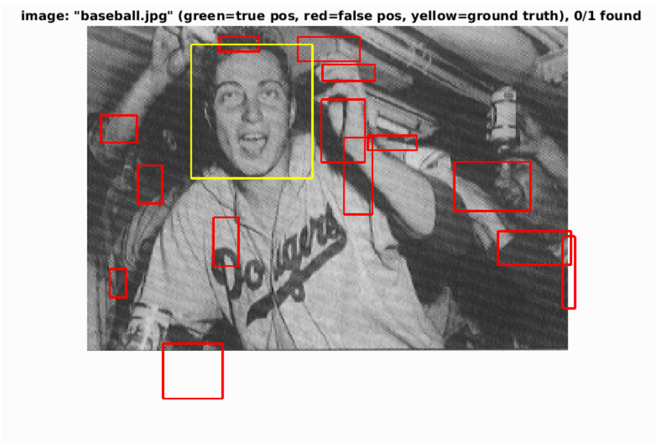

Figure 9. Anex 8. Image of a female face with it's HOG with a cell Size of 8

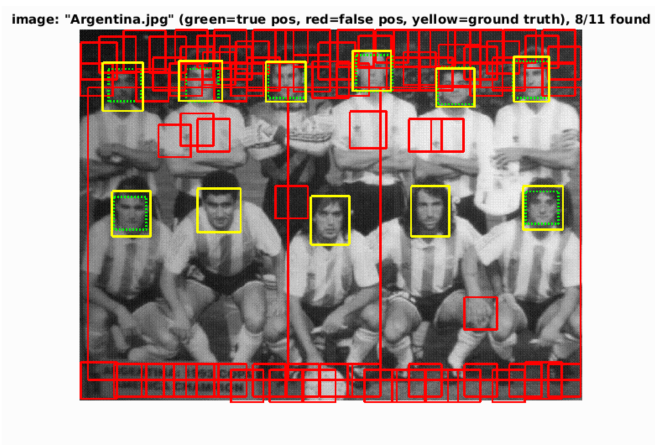

Figure 10. Anex 9. Image of a female face with it's HOG with a cell Size of 8

Figure 11. Anex 10. Image of a female face with it's HOG with a cell Size of 8

Figure 12. Anex 11. Image of a female face with it's HOG with a cell Size of 8

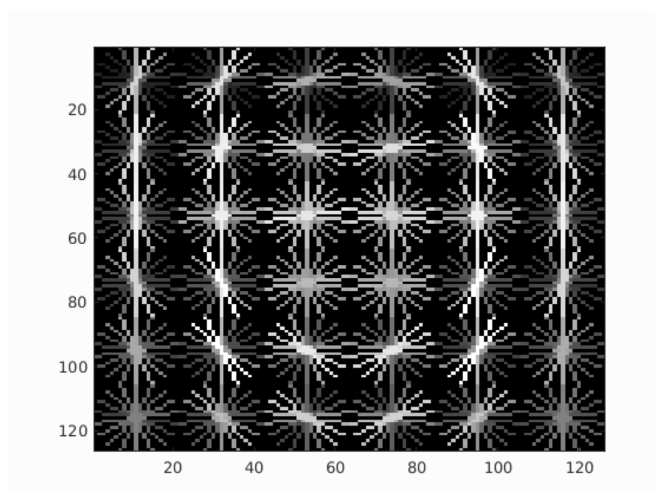

Figure 13. Anex 13. HOG of one of the images that were in the group with positive characteristics