Face Detection with HOG - Lab9 - Computer Vision

Jose Yesith Juez
University of the Andes
jy.juez@uniandes.edu.co

Nicolas Florez
University of the Andes
n.florez228@uniandes.edu.co

1. Introduction

Giving the ability of interpreting and recognizing images and object within them to the computers just as humans do is one of the key challenges of computer vision. Histogram of oriented gradient is a tool to recognition of objects within images with a successful trajectory in detecting pedestrians. And Appling to face recognition could be a good approach for a successful recognition.

Face detection is one important task the brain does when interpreting a scene that could give context to the surrounding environments when knowing who is the one in the scene. If we could teach to teach the machines to interpret a person and furthermore, who that person is. We can truly be talking of intelligent machines that could use this information in many applications from just in the scene interpretation to interactivity with humans.

2. Materials and Methods

The hyperparameters of a HOG multiscale are the number of scales that will be made in the image and the size of the window to be used (the HOG cell). In theory, greater scalability allows a greater detection of objects of different sizes.

2.1. Description of the multi-scale HOG

The HOG multiscale is a characteristic space for representing an image in different scales. For each scale, the different oriented gradients that are in a region of the image are calculated. These are used to calculate the histograms in cells (usually 8x8 pixels). The concatenation of the histograms obtained is the HOG descriptor.

The HOG can be used for a detection problem because it allows us to represent the objects that are in an image locally (appearance and shape), by means of the intensity distribution of gradients. But, unlike other methods, the HOG uses a dense grid of cells uniformly spaced in an image.

2.2. Train and test data.

The images that were used for train and validation of the face detection were obtained from the WIDER FACE database. This database is divided into three directories:

- Train: Set of train images that contains at least one face.
- Train Crops: Square images containing the cut out faces of the training images
- Val: Set of validation images that contain at least one face.

Three directories were created called as follows:

- ”myPositives”, which contains 55 images from the ”Train Crops” directory of the WIDER FACE database. Here are located all the training positive data (faces detected).
- ”myNegatives”, which contains 55 images from the ”Train” directory of the WIDER FACE database. Here are located the images that will be used for the train of the data.
- ”myTestImage”, which contains 10 images from the ”Val” directory of the WIDER FACE database. Here are located the images that will be used for the validation of the obtained results.

2.3. HOG determination without data training.

The HOG is calculated using the ”Vl_hog” function of the VLFeat library. The parameters of this function are the size in pixels for each cell of the HOG (for our case it is equal to 8) and an array of image patches, represented by each of the positive samples of the training data. The Figure No. 1 shows the arrangement of ”image patches” and their average. The HOG model obtained without having performed the data training with the support vector machine (SVM) is shown in Figure No. 2.

2.4. Multiscale detection and data training with SVM

Due in an image there can be faces of different sizes, it is necessary to represent the image in different scales. For our case, 15 different scales were used.

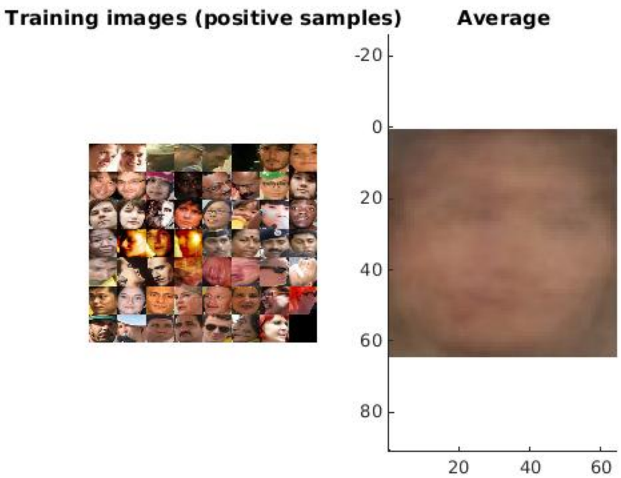

Figure 1. Train Images and Average

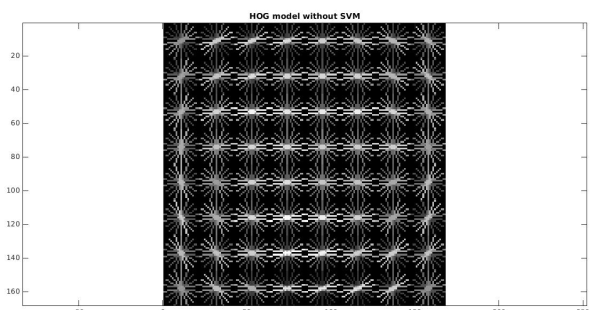

Figure 2. HOG without SVM

In the development of the SVM model, the "vl_svmtrain" function of the VLFeat library was used. For the training with the SVM we take into account the positive samples and the negative samples. In the determination of the negative samples, we used the Hard Negative Mining learning technique. This technique allows finding a small set of key negative samples. For our case we use 7 iterations to find the correct negatives; that is, the training of positive and negative data with the SVM was performed 7 times, where the first iteration was assumed as if there were no negative samples.The Figure 3 shows the HOG obtained with the SVM using the Hard Negative Mining technique.

The detection of the faces in the training images is done using the function "evalmodel". This function is responsible for mapping possible faces that were detected with a certain level of confidence. To eliminate redundant detections, the non-maximum suppression technique is used,that allow to obtain the detection frames with greater confidence.

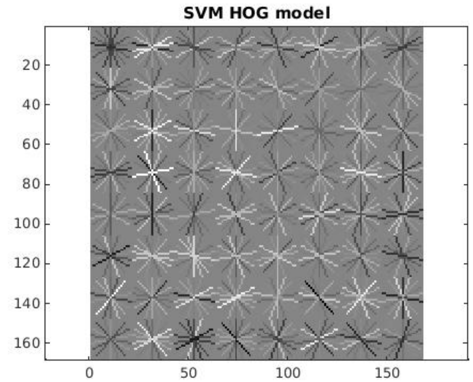

Figure 3. SVM HOG Model

3. Results

3.1. Evaluation with test images.

The detection of the faces in the training images is done using the function "evalmodel". This function is responsible for mapping possible faces that were detected with a certain level of confidence. To eliminate redundant detections, the non-maximum suppression technique is used,that allow to obtain the detection frames with greater confidence.

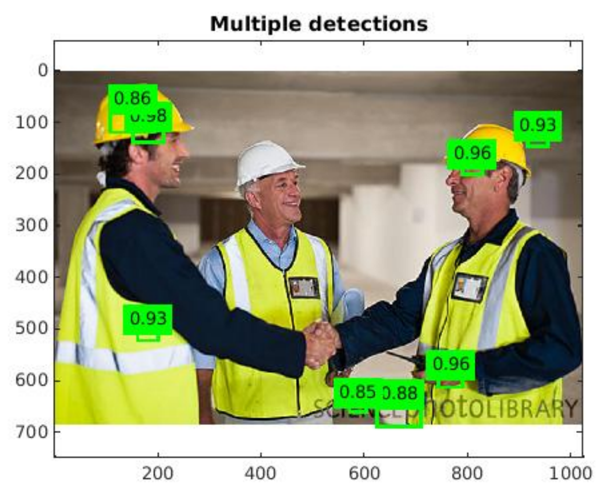

Figure 4. Multiple detections

For each Test image, a file was saved in .mat format taking into account those parameters (coordinates of the detection patches and their corresponding confidence). Each of these files in .mat format were used to obtain the detection parameters in a specific format to perform the evaluation with the given script in class (wider_eval.m). The code used is found in the "EvaluacionFD.m" script

The file where the detection parameters were saved is called "prediccion.txt".

The precision-recall curve obtained by using the ”wider_eval.m” script is shown in Figure 5. It is noted that a comparison is made between other detection methods performed previously.

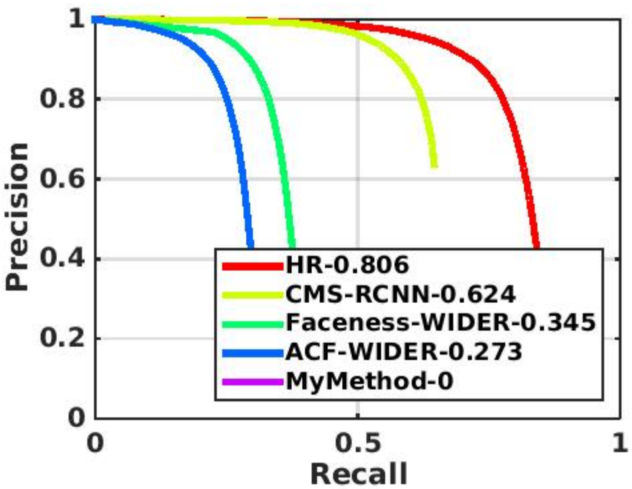

Figure 5. Precision-recall curve

4. Discussion

The result obtained from the developed method was not very encouraging. According to the recall accuracy curve obtained, the Overlap is 0. This could have been presented because the set of test images that was used is very small. Despite this, it can be observed in figure 4, that if the detection of some faces is made, with a high confidence level.

False positives can be presented by the small amount of training data used. False negatives can be presented by the number of iterations that are made using the Hard Negative Mining technique. Another important factor is the suppression of non-maximums, since this depends a lot on a limit that is defined to ignore certain detections that are redundant, so in some cases it can eliminate true positives.

One way to improve the detection strategy used is to use a greater amount of positive data for training. Also, you can use a greater scalability of data, this to detect more objects of different sizes.

5. Conclusions

It is observed that the HOG descriptor is a very powerful detection method and very simple to implement, but it is necessary to make a good selection of training and test data.

The levels of confidence for each face detected depend a lot on the limit defined for the realization of the suppression of non-maximums and the negative data generated by the Hard Negative Mining methodology.

References