INTERNATIONAL JOURNAL
OF SOFT COMPUTING

MEDWELL PUBLICATIONS
SINCE 2000

Study of Haar AdaBoost (VJ) and HOG AdaBoost (PoseInv) Detectors for People Detection

[1]Nagi OULD TALEB, [1]Mohamed Larbi BEN MAATI, [1]Mohamedade Farouk NANNE and [2]Aicha Mint Aboubekrine

[1]*Department of Computer Sciences, University of Abdelmalek Essaadi, Tetouan, Morocco*

[2]*Department of Computer Sciences, University of Nouakchott Al Aasriya, Nouakchott, Mauritania*

Key words: Pedestrian detection, learned-based methods, behavior analysis

Corresponding Author:
Nagi OULD TALEB
Department of Computer Sciences, University of Abdelmalek Essaadi, Tetouan, Morocco

Page No.: 134-147
Volume: 16, Issue 6, 2021
ISSN: 1816-9503
International Journal of Soft Computing

Abstract: The detection of objects in general and pedestrians in particular in images and videos is a very popular research topic within the computer vision community, it is an issue that is currently at the heart of much research. In this study, we will present a comparative study of the performance of the two detectors Haar AdaBoost and HOG AdaBoost in detecting people in the INRIA image database of people. An evaluation of the experiments will be presented after making certain modifications to the detection parameters.

INTRODUCTION

The detecting people in images is a very important subject in the field of computer vision. The detection of pedestrians is therefore a main concern of several researchers in the field of computer vision. These applications, ranging from surveillance, retail data mining and automatic pedestrian detection in the automotive industry have fueled research over the past decade, leading to a growing number of approaches on the subject[1].

Many factors can influence the human figure, such as the constantly changing appearance, crowds, obscuration by objects, the type of environment and the unpredictability of pedestrians[2, 3].

In the literature, we find techniques that require segmentation or subtraction of the background and others directly detect the person without such preprocessing. These techniques use many characteristics to describe human appearance (shape, color, movement) in order to build shape models used on explicit or learning-based detection techniques.

Several systems have been developed in this context with dynamic methods such as Phantom and P finder[4]. Other methods have been conducted[5, 6] for the detection of people with a measure of their activity in video sequences. Shooting with a fixed camera allows background subtraction to reduce search space. Finally, we find the system that performs fast and accurate human detection by integrating the cascade approach with histograms of gradient directions[7, 8].

Among these approaches, we find a so-called global one that has a principle of using the shape of the whole body as a source of information without taking into account local characteristics[9, 7, 10, 11, 5] also proposed a detector based on Haar filters and the boosting algorithm. There are some aspects of this algorithm, based on infrared vision to detect a human in a room and provide a history of occupancy of the room. Another so-called local approach uses local characteristics. Here, we extract the characteristics from the image base and then, we build the discriminating model, for example, Papageorgiou *et al*.[12] that proposed a detector based on the Haar wavelet.

Another so-called local approach uses local characteristics. Here, we extract the characteristics from the image base and then we build the discriminating model as for example, Papageorgiou *et al*.[12] which proposed a detector, based on the Haar wavelet.

The latest so-called hybrid approach combines local and global characteristics to improve recognition performance[13].

The research work proposed in this article aims to contribute to the modeling of methods of recognition of shapes (or objects) and more particularly of pedestrians by classification of descriptors containing the most relevant information of an object and to apply the models found to the detection of the human silhouette (people or pedestrians) in images or multimedia streams (video).

MATERIALS AND METHODS

Learning methods for human detection: In The main approaches based on discriminant learning train different types of classifiers on a large number of samples of negative and positive images where humans are well framed.

Each method must extract the appropriate characteristics and the main information captured from the training data is the spatial recurrence of local shape events. If the trained classifier does not detect an object (missing the object) or mistakenly detects the missing object (false detection), it is easy to make an adjustment by adding the corresponding positive or negative samples to the learning set.

However, due to the complexity of articulated human poses and variable visualization conditions, training data becomes very large (especially positive samples) therefore the generalization ability of the trained classifier may be compromised. Illustration of the common basic data formation process in Fig. 1.

Study of Haar-AdaBoost and HOG-AdaBoost detectors for the detection of people: The study which we carried out in the paper[14] of 14 traditional techniques resulting from the literature and representing the state of the art allowed us to choose the two most popular methods in the detection of the objects Haar AdaBoost (VJ)[11,15] and HOG AdaBoost (PoseInv) which constitutes a variant of HOG SVM to study their feasibility for detecting people.

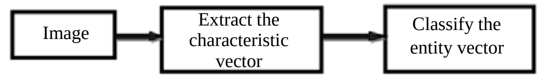


Fig. 1: Common learning process

In this approach, we will present the experimental results carried out in the Computer Science and Systems Engineering Laboratory of the Faculty of Sciences of Tetouan for the evaluation of the detection of people in images using the two detectors Haar AdaBoost (or VJ) and HOG AdaBoost (or PoseInv).

The performance analysis of these two detectors was carried out on the database of images of people from INRIA Person Dataset (http://pascal.inrialpes.fr/data/human/). This database provides 460 color bitmap (BMP) images of people at 640×480 resolution.

The study thus made is based on the plotting of the VP IoU, FP IoU, FN IoU, IoU sensitivity curves and the evaluation of the AR (Average Recall) metric. Plotting the precision sensitivity curve and evaluating the AP (Average precision) metric cannot be performed in this study because the Haar AdaBoost and HOG AdaBoost detectors do not return a confidence score but rather an indication whether an object detected belongs to the desired class or not.

The OpenCV library offers a list of classifiers in XML format already trained to respectively detect faces, eyes, profile heads, human bodies, etc. These classifiers are located in the OpenCV\data\haarcascades\folder.

Among the classifiers provided by OpenCV, we have chosen to study the performance of two of them which are already trained for detecting people in images, haarcascade_fullbody.xml and hogcascade_pedestrians.xml which provide two models for detecting people in the images obtained, respectively by the implementations under OpenCV of the cascade classifier Haar AdaBoost of Viola and Jones and HOG AdaBoost of Lin and Davis (a variant of the detector of Dalal and Triggs).

INRIA person dataset image database and manual labeling of images: Among the 460 images in the INRIA person dataset, we manually tagged 187 images using the object marker annotation program, resulting in a total of 481-ground truth framing boxes.

Figure 2 shows four images from the INRIA Person Dataset labeled using the object marker program, each person presented in these images is manually framed using a rectangle, called a truth framing box ground. These boxes give precise positions of the people in the images, they are presented by the coordinates (x, y) of the upper left point of the rectangle, its width and its height.

Fig. 2: Four images from the INRIA person dataset labeled using the object marker program

Comparison of the two detectors Haar AdaBoost and HOG AdaBoost using examples of detection of people in the INRIA person dataset database: In this paragraph, we will present a preliminary comparative study of the two detectors Haar AdaBoost and HOG AdaBoost. This comparison will be based on the application of these two detectors on the images of the INRIA person dataset database. We will discuss the strengths of each of these two detectors as well as their failing.

Figure 3 shows some examples of people detection obtained, respectively by the application of the two Haar AdaBoost and HOG AdaBoost detectors on the images in the INRIA Person Dataset database. The first column corresponds to the application of the Haar AdaBoost detector while the second column corresponds to the application of the HOG AdaBoost detector. To compare the two detectors and discuss their performance, we have chosen to show some of the most significant detection results obtained on a sample of well-selected INRIA person dataset images. In the images below, the blue frame corresponds to the ground truth frame box produced by manual labeling using the objectmarker program. The boxes in green correspond to the boxes predicted, respectively by the two detectors Haar AdaBoost and HOG AdaBoost.

Experimentation with Haar AdaBoost and HOG AdaBoost detectors on images from the INRIA Person Dataset database allowed the following conclusions to be drawn:

The two detectors generally fail to detect people on a small scale (or very far away). Likewise, on a very large scale or when people are very close and fill almost the entire image, the two detectors generally do not succeed in detecting them or sometimes generate, in particular by the HOG-AdaBoost detector, boxes Small predicted framing whose IoU with their associated ground truth framing boxes is of small value. Sometimes the shape of the clothes (especially if a person is wearing a coat or a gown) can also cause a person on a medium scale to not be detected by the Haar AdaBoost and HOG AdaBoost detectors.

The HOG AdaBoost detector is overall better than the Haar AdaBoost detector for detecting people on a medium scale (i.e., people who are slightly close) and large scale (i.e., people who are close) but unfortunately it generates a lot of false detections (or false positives) than the Haar AdaBoost detector. On a medium and large scale, the Haar AdaBoostr detector sometimes sends two predicted boxes corresponding to the detection of the same person. This does not happen with the HOG AdaBoost detector which does a good job of eliminating duplicates and typically returns a single predicted box for each person detected.

The choice of the IoU threshold is very important so as not to miss some correct detections. We have observed that with the IoU threshold set at 0.5, the Haar AdaBoost detector sometimes returns detections which are correct but which have an IoU <0.5 which leads to an erroneous interpretation of the results of the detections obtained. This situation rarely happens with the HOG AdaBoost detector where the IoU of detection is often >0.5.

We have also found that the minimum value of the IoU threshold that must be set depends on how to label people, in fact, if the field truth framing boxes are manually drawn too tight to the people that they frame, the detector can sometimes generate an IoU with the predicted framing box <0.5.

Based on the analysis of the detection results obtained by the Haar AdaBoost and HOG AdaBoost detectors, we found that for the minimum IoU threshold value set at 0.4, almost all detections that give rise to true positive are correctly determined. Therefore, to study the performance of the two detectors, it will be preferable to vary the minimum threshold of the IoU between 0.4 and 1 instead of 0.5 and 1, this is what we used in the plotting of the true positive as a function of the IoU (VP IoU), false positive as a function of the IoU (FP IoU) and false negative as a function of the IoU (FN IoU) and sensitivity as a function of the IoU (Sensitivity-IoU).

Simple version of the metric evaluation algorithm VP, FN and FP: To evaluate the VP (True Positives), FN (False Negatives) and FP (False Positives) metrics, we will start by presenting a first simple version of an algorithm for calculating these values.

For simplicity's sake, let's assume that each person detected in an image is located only once using a predicted box. In other words, there is a single predicted framing box associated with the ground truth framing box framing that detected person (Algorithm 1).

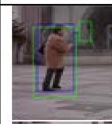
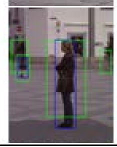
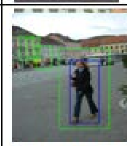
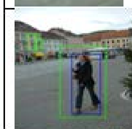
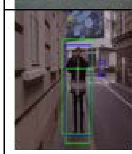
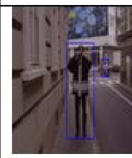
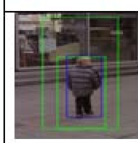
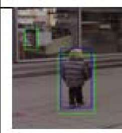
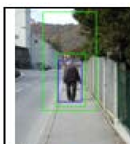
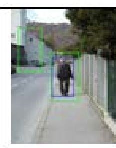
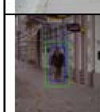
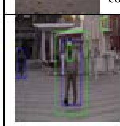
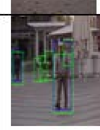
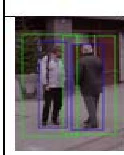
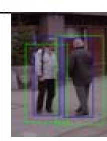
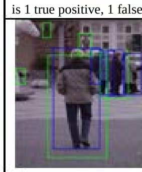
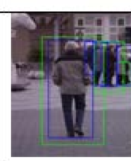
Détecteur Haar AdaBoost	Détecteur HOG AdaBoost
 The person on a large scale (or at a close distance) is not detected. This is therefore, a false negative	 Here, the same large-scale person is indeed detected with an IoU equal to 0.755259, this is a true positive. The image also shows a false detection that matches the green frame on the side of the display case, so this is a false positive
 The mid-scale child is well detected (IoU equals 0.580247). The large-scale lady goes undetected. There is a true positive (the child) and a false negative (the lady) here	 Here, the medium-scale child and the large-scale lady are well detected (the respective IoUs are 0.613888 and 0.482886). Despite the lady being detected, the IoU between the green and blue frame is 0.482886 which is less than 0.5. In this case, if the IoU threshold is set to 0.5, then the lady's blue and green frames will be considered a false negative and a false positive respectively which is incorrect. There are also two false detections (or false positives)
 The large-scale lady is well detected (IoU is 0.636414), but the small-scale people are not detected. There is therefore 1 true positive and 3 false negatives	 Same thing here, the large-scale lady is detected (IoU equals 0.70401) but, the small-scale people are not detected. There is therefore 1 true positive, 3 false negatives and 1 false positive (the statue)
 A person is detected with an IoU equal to 0.527575. There is therefore here 1 true positive and 2 false positives (or false detections) which correspond to the green boxes on the side of the windows of the buildings	A person is detected with an IoU equal to 0.465565. If the IoU threshold is taken equal to 0.5, then the ground truth box in blue will be considered as a false negative and the predicted box will be considered as a false positive which is wrong. Here, there are also 3 false positives which correspond to the green boxes on the side of the buildings
 The same person is detected twice, the two boxes predicted in green have respectively for IoU O, 622019 and 0.793165. Only the box with the maximum IoU, i.e., 0.793165 should be counted as a true positive, the other should be removed and it should not be counted. The small-scale person on a motorcycle is not detected, so in this example there is 1 true positive (one of the two predicted boxes is not counted) and one false negative (the person on the motorcycle)	 Both large-scale and small-scale motorcycle people are not detected. There are 2 false negatives here. Apparently, here is the shape of the gown worn by the person who trained it to go undetected by the HOG AdaBoost detector
 This example also shows that the same large-scale person is detected twice, but here the two boxes of frames have respectively an IoU of 0.607078 and 0.253846 (the large green frame). In this case, the box with the IoU of 0.607078 will be considered a true positive, whereas the one with an IoU of 0.253846 will be considered a false positive	 The large-scale child is well detected with an IoU equal to 0.692701. There is also a false detection corresponding to the green frame on the display case, this is a false positive
 Same thing as the previous example, the same person at medium scale is detected twice with two boxes of predicted frames having respectively for IoU of 0.623512 and 0.181492. In this case the box having the IoU of 0, 623512 will be considered a true positive, however the one with an IoU of 0.181492 will be considered a false positive	 The medium scale person is well detected with an IoU equal to 0.653686. There is 1 true positive and one false positive here
 The mid-scale person is detected with a predicted box having for IoU equal to 0.457869. If the threshold of the IoU is taken equal to 0.5, then the field truth framing box in blue will be considered as a false negative and the predicted framing box will be considered a false positive which is wrong	 The medium scale person is well detected with an IoU equal to 0.580978. There is therefore 1 true positive here
 A large-scale person is detected twice using the green boxes predicted for IoU 0.48906 and 0.0806955 respectively, the box with IoU 0.0806955 will be rejected and considered as a false positive. Likewise, the box with the IoU of 0.48906 will also be rejected if the IoU threshold is set to 0.5 and it will also be considered as a false positive	 The large-scale person is detected twice using the green boxes predicted in green which have respectively for IoU 0.7074 and 0.153458.The box with the IoU of 0.153458 will be rejected and it will be considered as a false positive. On the other hand, the box with the IoU of 0.7074 will be accepted and considered as a true positive. The mid-scale lady is also detected using a predicted box with the IoU of 0.612489
 Both people are detected but they have respectively 0.437131 (person on the left) and 0.518307 (person on the right) for IoU. If the IoU threshold is set to 0.5, then only the box predicted for the person on the right with the IoU of 0.518307 will be considered a true positive. On the other hand, the predicted box and the ground truth box for the person on the left will be respectively considered as a false positive and a false negative. So, for the IoU threshold set at 0.5, there is 1 true positive, 1 false positive and 1 false negative which is not correct	 Both people are well detected with IoUs of 0.543696 (person on the left) and 0.5461 (person on the right), respectively. There are therefore, 2 true positives here
 Two people are detected, the large-scale man and a lady in the medium-scale crowd. The IoUs obtained are 0.674091 and 0.522472 respectively. Note that the crowd side predicted framing box overlaps with multiple ground truth framing boxes, but only the ground truth framing box having the IoU with the highest predicted box will be taken. the others will be considered false negatives, in this example there are 2 true positives, 3 false negatives and 3 false positives	 Here, three people are detected, the large-scale man and two ladies in the medium-scale crowd, the obtained IoUs are 0.650545 and 0.449743 and 0.48198, respectively. If the IoU threshold is set to 0.5, the two detections in the crowd will be considered false positives and the corresponding ground truth framing boxes will be considered false negatives. 'IoU taken equal to 0.5, the detection in this example gives 1 true positive, 4 false negatives and 2 false positives

Fig. 3: Some people detection results obtained respectively by applying the two Haar AdaBoost and HOG AdaBoost detectors to the images in the INRIA Person Dataset database

Algorithm 1; Evaluate the number of true positives, false negatives and false positives:

Input:

- Database of labeled images
- For each image in the database, we have the list of field truth framing boxes and the list of predicted frame boxes
- The minimum threshold of the IoU

Output: VP, FN et FP

We initialize: VP = 0 et FP = 0

For each image of the data bass:

For each detection (or predicted frame box) in the current image:

Choose from all the field truth framing boxes labeled in the image, the one that has the highest IoU with the predicted framing box

If all the field truth framing boxes in the current image have an IoU below the minimum IoU threshold (typically 0.5), then:

Detection is a false positive and increments FP : FP = FP+1

If else:

The detection is a true positive and we increment VP : VP = VP+1

Since, each predicted framing box corresponds to one and only one field truth framing box (or a person in the image), one can easily calculate FN by:

FN = Total number of field truth frames boxes in the database-VP

This simple algorithm has the advantage of quickly calculating VP, FN and FP metrics but unfortunately it is only suitable if the detector effectively returns a single predicted box for each object detected in an image. In our case, the considered object is a person labeled manually using a ground truth framing box.

This algorithm is therefore suitable for the HOG AdaBoost detector but not for the Haar AdaBoost detector, since this denier can sometimes return two predicted border boxes for the same person and therefore this box will be counted twice as a true positive. Whereas normally only one predicted box should be counted as a true positive and the other should be ignored.

Subsequently, we will propose a general algorithm making it possible to correctly calculate the VP, FN and FP metrics. This second version of the algorithm is unfortunately slower in computing time than the previous algorithm but it has the advantage of working regardless of the number of predicted border boxes returned by a detector for the same object (or person) tagged in an image using a ground truth framing box.

General version of the VP, FN and FP metric evaluation algorithm: It is assumed that the same person can be detected more than once, that is, there are several predicted framing boxes which may correspond to the ground truth framing box framing that person.

Here are two problems that can arise when it comes to pairing predicted boxes with ground truth boxes (or detected people):

Several predicted framing boxes can correspond to the same person if they have, together with the ground truth framing box framing this person, an IoU greater than a certain minimum threshold of the IoU (typically 0.5). In this case, only a single predicted box should be counted as a true positive, others if not associated with other nearby people should be ignored.

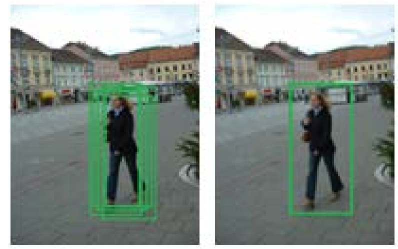


Fig. 4: For the same person in the image, a multitude of windows are detected (the image on the left). You have to determine which one best frames the person (the image on the right). The confidence score is used by the non-maximum elimination technique to find the window that maximizes it and to eliminate the others that do not

For people located side-by-side in an image, the ground truth boxes can usually overlap with each other. In this case, the predicted framing boxes may also overlap with each other and with several ground truth framing boxes. These predicted boxes can therefore have an IoU greater than the minimum threshold with several ground truth framing boxes (or several labeled people). We must therefore, determine how to correctly associate each predicted box with the ground truth box it represents (or the person detected).

To overcome these two difficulties and correctly evaluate the VP, FN and FP metrics, that is to say, to avoid repeatedly counting the same person detected with several predicted boxes which will distort the calculation of VP and FP, we propose a general algorithm whose idea is based on the principle of Non Maximum Suppression[16]. We have used the latter to associate each field truth framing box (or detected person) with the predicted framing box that maximizes the IoU with it and eliminating other predicted framing boxes that do not maximize the IoU. IoU provided that they do not match other people in the vicinity.

Typically on a sliding detection window, the exhaustive search for a person (or an object in general) in an image carried out by certain detectors such as Haar AdaBoost and HOG AdaBoost, for example, test all the possible detection windows at all scales. and locations. For each of these detection windows, a decision on whether or not it belongs to the desired class is obtained by the detector.

For a person in the initial image, there is a window framing it in the most precise way. However, windows that are spatially close or in scale may also give a positive classification. We then obtain a constellation of positive detection windows around the same detected person (Fig. 4).

The Non Maximum Suppression (NMS) technique is one of the methods used during the object detection phase to eliminate neighboring windows that do not maximize the confidence score for a detected object. The confidence score is a value between 0 and 1 generally predicted by a classifier, it represents the probability that a detection window contains an object. The confidence score is used as a comparison value between neighboring detection windows. The principle consists in keeping for a detected object only the detection window which maximizes the confidence score and to eliminate the others which do not maximize it.

In our case, we use the principle of no maximum suppression after the phase of the detection of people, we based it on the comparison of the IoU between the predicted framing boxes and those of ground truth and not on the confidence score. This choice to use the NMS with the IoU was made for the following two reasons:

The Haar AdaBoost detector can sometimes generate for the same person detected two predicted boxes that correspond to it.

The Haar AdaBoost and HOG AdaBoost detectors are respectively based on the binary classifier AdaBoost (or SVM for the Dalal and Triggs detector[9] which do not return a confidence score but rather the values-1 (or a negative value) for non-membership of the class of the object to be detected or 1 (or a positive value) to indicate membership of the class of the object.

The principle of the general evaluation algorithm for VP, FN and FP metrics that we have developed is as follows (Algorithm 2):

Algorithm 2; Evaluate the number of true positives, false negatives and false positives:

Input:

- Database of labeled images
- For each image in the database, we have the list of field truth framing boxes and the list of predicted frame boxes
- The minimum threshold of the IoU

Output : VP, FN et FP

We initialize VP = 0, FN = 0 et FP = 0

For each image of the data bass perform the following processing:

- Mark all frame boxes predicted as not being assigned to a field truth framing box
- Associate with each field truth framing box (or a labeled person) in the current image the list of predicted framing boxes that have an IoU with it that exceeds a certain threshold (typically 0.5). The list of predicted frame boxes is sorted in descending order of IoUs and all predicted boxes in the list are marked as affected. If the list of predicted frame boxes is empty, that is, there is no predicted box associated with the field truth box, then the latter is a false negative or a missed person. In this case, we increment the FN metric. The ultimate goal of the algorithm is to determine for each field truth framing box framing a person detected in the image a single predicted framing box that maximizes the IoU with it. In this case, only this predicted box will be counted as a true positive, the other predicted boxes on the list if they are not associated with other people located side by side will be ignored
- Evaluate the FP metric: it corresponds to the number of predicted frame boxes that are not marked as assigned to a field truth frame box
- If several detected framing boxes correspond to the same field truth framing box framing a person (or object in general), the Non Maximum Deletion principle is applied to keep only the detected framing box having a maximum IoU with the field truth framing box. This operation is necessary to properly calculate the VP number, as it avoids counting the predicted boxes for a detected person several times

For each field truth framing box b1 in the current image:

If the list of predicted framing boxes associated with box b1 is not empty, then::

The p1 box of framing predicted at the beginning of the list has the maximum IoU. We then take this box p1

For each field truth framing box b2 in the current image:

If the box p2 at the beginning of the list of predicted framing boxes associated with box b2 is the same as p1::

If the IoU of p2with b2 is greater than that of p1with b1 then it can be confirmed that the predicted box p1 is not associated with the box b1.

Otherwise (the IoU of p2 is smaller than that of p1), we remove p2 from the beginning of the lite of the predicted framing boxes associated with the box b2

If in the previous iteration it was determined that the predicted box p1 was not associated with b1, then in this case p1 is removed from the beginning of the lite predicted framing boxes associated with box b1

- Evaluate the VP metric: it corresponds to the number of field truth framing boxes with a list of predicted framing boxes associated with them non-empty (these field truth framing boxes therefore correspond to detected people)

RESULTS AND DISCUSSION

Analysis of the performance of the two detectors Haar AdaBoost and HOG AdaBoost on the images of the INRIA person dataset database: After having implemented the general algorithm for evaluating VP, FN and FP metrics in C++ language using the OpenCV library, we used it to evaluate the performance of the two detectors Haar AdaBoost and HOG AdaBoost.

Table 1 and 2 show the results of the analyzes obtained by the respective application of the two detectors on 187 images chosen from 460 bitmap color images of people from the INRIA Person Dataset database. Manual tagging of the people in the 187 images resulted in a total of 481 ground truth framing boxes framing these people.

Since, the two detectors are applied to the same images in the INRIA Person Dataset database, we will start by making a simple comparison by plotting the True Positives curves as a function of the Intersection on the union (VP IoU), False Positives as a function of the Intersection on the union (FP IoU) and false negatives as a function of the Intersection on the union (FN IoU) (see these curves in Fig. 5, they were plotted under Microsoft Excel).

It can be seen from the VP IoU curve in Fig. 5 that the HOG AdaBoost detector (curve in red) is more efficient than the Haar AdaBoost detector, since, it allows more positives to be detected, i.e., say, of people labeled as Haar AdaBoost (curve in blue).

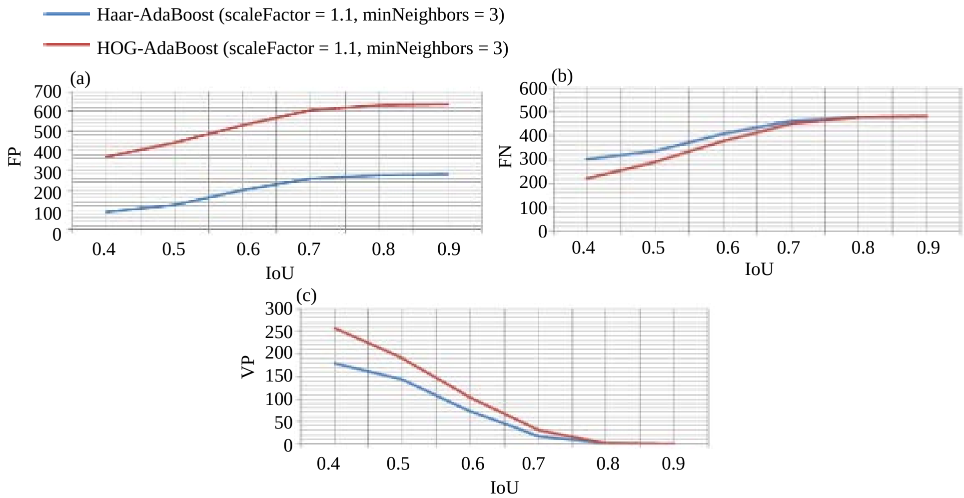


Fig. 5(a-c): These curves show that the HOG AdaBoost detector (in red) is more efficient at detecting people than Haar AdaBoost (in blue) but it generates more false detections than the latter. These results were obtained for the values of the detection parameters scaleFactor and minNeighbors respectively equal to 1.1 and 3; (a) Courbe FP-IoU, (b) Courbe FN-IoU, (c) Courbe VP-IoU

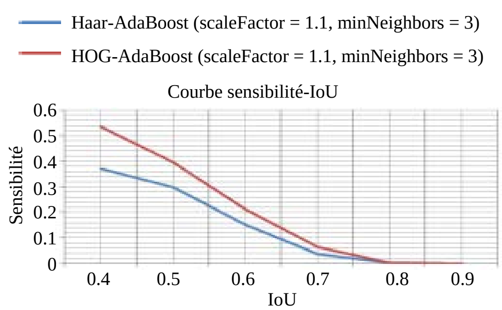


Fig. 6: This curve shows that the HOG-AdaBoost detector (red) is more effective at detecting people than the Haar AdaBoost detector (blue)

Table 1: Result obtained by applying the Haar AdaBoost detector on 187 images from INRIA Person Dataset containing 481 people labeled using field truth framing boxes

Haar-AdaBoost

Seuil IoU	VP	FN	FP	Précision	Sensibilité
0.4	179	302	90	0.665428	0.372141
0.5	144	337	127	0.531365	0.299376
0.6	73	408	203	0.264493	0.151767
0.7	18	463	261	0.064516	0.037422
0.8	2	479	277	0.007168	0.004158
0.9	0	481	279	0	0

Table 2: Result obtained by applying the HOG AdaBoost detector on 187 images from the INRIA Person Dataset containing 481 people labeled using the field truth framing boxes

HOG-AdaBoost

Seuil IoU	VP	FN	FP	Précision	Sensibilité
0.4	258	223	368	0.412141	0.536383
0.5	191	290	441	0.302215	0.397089
0.6	103	378	532	0.162205	0.214137
0.7	31	450	604	0.048819	0.064449
0.8	2	479	633	0.00315	0.004158
0.9	0	481	635	0	0

Likewise, the FP IoU curve also shows that there are also fewer false negatives or misses with HOG AdaBoost than with Haar AdaBoost. On the other hand, the HOG AdaBoost detector is less efficient than Haar AdaBoost with regard to false detections, since in return for its efficiency in detecting positives, it has the disadvantage of generating a lot of false detections or false positives.

Since, the two detectors are based respectively on the AdaBoost binary classifier which does not return a confidence score but rather a response indicating whether the detected object is part of the sought class or not, it will therefore not be possible to use the curve. Precision sensitivity used to calculate the AP (Average Precision) metric. We will therefore use in its place the IoU Sensitivity curve which makes it possible to calculate the AR (Average Recall) metric.

Subsequently, we will complete the comparisons made by the curves in Fig. 5 by plotting the IoU Sensitivity curve (Fig. 6). This is more general than the previous curves, it is often used to study the efficiency in detecting true positives, in addition it allows to evaluate the average sensitivity metric AR (Average Recall) which is used to compare detectors even if they are applied to different image databases.

Knowing that the AR metric corresponds to the area of the region below the IoU sensitivity curve between IoU 0.5 and 1 and it is given by equation B.6. To evaluate this metric, we will approximate the integral B.6 using the rectangle method that is given by Eq. 1:

$$AR = 2\sum_{i=1}^{n=1}\left(IoU_{i+1} - IoU_{i}\right) \text{ sensibility } \left(IoU_{i+1}\right) \quad (1)$$

Here, IoU1 is equal to 0.5 and IoU n is equal to 1. The interval [0, 5, 1] is divided into n intervals of the same length equal to IoUi+1- IoUi = 1-0.5/n = 0.5/n, 1≤ i≤n-1.

Since, in our case, we have taken n = 5 and the IoU variable between 0.5 and 1, we can deduce that the step of the variation will be fixed at 1-0.5/n = 0.5/5 = 0.1 Eq. 2 will become:

$$AR = 2\times 0.1\times \sum_{i=1}^{n=1} \text{sensibility } (IoU_{i+1}) \tag{2}$$

With IoU1 = 0.5 and IoUi+1 = IoUi+0, 1, 1≤i≤n-1. Based on the results of the detections obtained by the Haar AdaBoost and HOG AdaBoost detectors and which are presented in Table 1 and 2, respectively, we evaluated the AR metric for each of the two detectors which made it possible to obtain the following result:

- Haar-AdaBoost: AR = 0,0985446
- HOG-AdaBoost: AR = 0,1359666

From the plot of the IoU sensitivity curve and the evaluation of the AR metric for both detectors, the HOG AdaBoost detector is more efficient at detecting people than Haar AdaBoost because the AR HOG AdaBoost is larger than the AR Haar AdaBoost. But unfortunately, according to the FP IoU curve in Fig. 5, the AdaBoost HOG detector has the disadvantage of generating a lot of false detections than the Haar AdaBoost detector.

Experimenting by changing certain detection parameters: To perform the detection of people, we used the detect multi scale method of the Cascade classifier class. It admits seven parameters, the most important that can be varied to study the detection of people or objects in general are the following two parameters:

- scaleFactor: allows you to define how much the size of the detection window will be reduced with each iteration. The default value for this parameter is 1.1
- minNeighbors: defines the minimum number of neighboring detections that a candidate area must have to be retained. The default value for this parameter is 3

The results of the analyses presented in paragraph 2.4.5 above were obtained using scaleFactor and minNeighbors parameters 1.1 and 3, respectively as values.

We repeated these experiments by assigning to the scaleFactor parameter the fixed value 1.1 and by varying the value of the minNeighbors parameter by assigning it the successive values 2, 3, 4 and 5.

The results of the analyses obtained by the two detectors Haar AdaBoost and HOG AdaBoosts on, respectively shown in Fig. 7 and 8.

The VP-IoU, FN-IoU and Sensitivity-IoU curves show that when the value of the minNeighbors parameter increases from 2-5, the detection of people (or true positives) improves by both detectors but in return, the number of false detections (or false positives) increases (see the FP-IoU curve).

It can be seen that the lower the value of the minNeighbors parameter, the better the detection of people and the higher the number of false detections. A compromise between good detection and false detections can be achieved by the intermediate value of minNeighbors equal to 3 (values 4 and 5 also give a suitable result).

To complete this study, we also assigned other values to the parameters (scaleFactor, minNeighbors) such as (1.01, 5), (1.01, 4), (1.01, 3), (1.01, 2), (1.05, 5), (1.05, 4), (1.05, 3), (1.05, 2), (1.1, 5), (1.1, 4), (1.1, 3), (1.1, 2), (1,15,5),(1.15, 4),(1.15, 3), (1.15, 2), (1.2, 5), (1.2, 4), (1.2, 3),(1.2, 3).

The curves of Fig. 9 (Haar AdaBoost) and 2.6.9 (HOG AdaBoost) were obtained for the values (scaleFactor, minNeighbors) equal to (1.01, 3), (1.05,3), (1.1,3), (1.15,3) and (1.2,3), they give an idea of the comparison of the detections that we obtained by varying the values of the scaleFactor and minNeighbors parameters as shown above.

The analyses, we performed for the value of the scaleFactor parameter varying from 1.01 to 1.2 and the value of the minNeighbors parameter fixed at 3 and which are illustrated by Fig. 10 allowed us to deduce the following conclusions:

When the value of the scaleFactor parameter decreases, there is an overall improvement in the detection of people due to an increase in the number of true positives (see the VP-IoU, FN-IoU and Sensitivity-IoU curves).

Unfortunately, this improvement is achieved at the expense of an increase in false detections (or the number of false positives, see the FP-IoU curve) and also in the calculation time of detections.

Table 3 and 4 also confirm the previous results, they give an overview of the detection rates obtained by the two detectors when the IoU value is set at 0.5, that of the minNeighbors parameter is set at 3 and by varying the value of the scaleFactor parameter which successively takes the values 1.01, 1.05, 1.1, 1, 15 and 1.2.

Knowing that in 187 images of the INRIA person dataset database, we have labeled 481 people using the field truth framing boxes (see paragraph 2.6.1), in this case, the detection rate will therefore be equal to the number of true positives detected in all 187 images divided by 481, that is, equal to.

Table 3: Detection rates obtained by the Haar-AdaBoost detector by setting the value of the IoU to 0.5, that of the minNeighbors parameter to 3 and by varying the value of the scaleFactor parameter

Détecteur Haar-AdaBoost					
	Values				
Parameters	1	2	3	4	5
scaleFactor	1.01	1.05	1.1	1.15	1.2
VP	210	165	144	125	105
Taux de détection en (%)	43.66	34.30	29.94	25.99	21.83

Table 4: Detection rates obtained by the Haar-AdaBoost detector by setting the IoU value to 0.5, that of the minNeighbors parameter to 3 and varying the value of the scaleFactor parameter

Détecteur HOG-AdaBoost					
	Values				
Parameters	1	2	3	4	5
scaleFactor	1.01	1.05	1.1	1.15	1.2
VP	215	208	191	186	171
Taux de détection en (%)	44.70	43.24	39.71	38.67	35.55

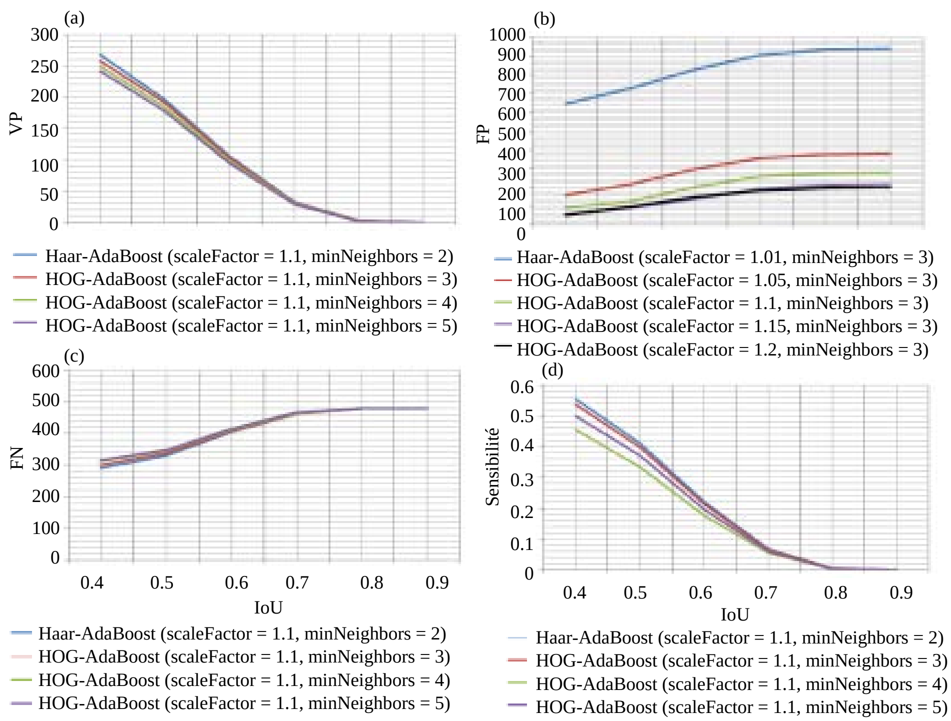


Fig. 7(a-d): Haar AdaBoost detector. These curves show the results of analysis of the detections obtained by the Haar AdaBoost detector by setting the value of the scaleFactor parameter to 1.1 and varying the value of the minNeighborsen parameter successively assigning it the values 2, 3, 4 and 5; (a) Courbe VP-IoU, (b) Courbe FP-IoU, (c) Courbe FN-IoU and (d) Courbe Sensibilité-IoU

It can be seen from Table 3 and 4 that the detection rate obtained by the two detectors is overall less than 50%, it increases when the scaleFactor parameter decreases from 1.2 to 1.01. The respective default values 1.1 and 3 of the two parameters scaleFactor and minNeighbors are central, they allow to obtain a suitable detection result that provides a compromise between real detections and false detections and a reasonable calculation time. The pairs of values (1.15,3) and (1.2,3) of the parameters (scaleFactor, minNeighbors) also provide a suitable detection result, since, the VP IoU, FN IoU and FP IoU and IoU sensitivity curves obtained for these two pairs of values are very close to those obtained for the value pair (1.1,3). In addition, these two pairs of values make it possible to carry out detections with a lower calculation time than that obtained for (1.1, 3).

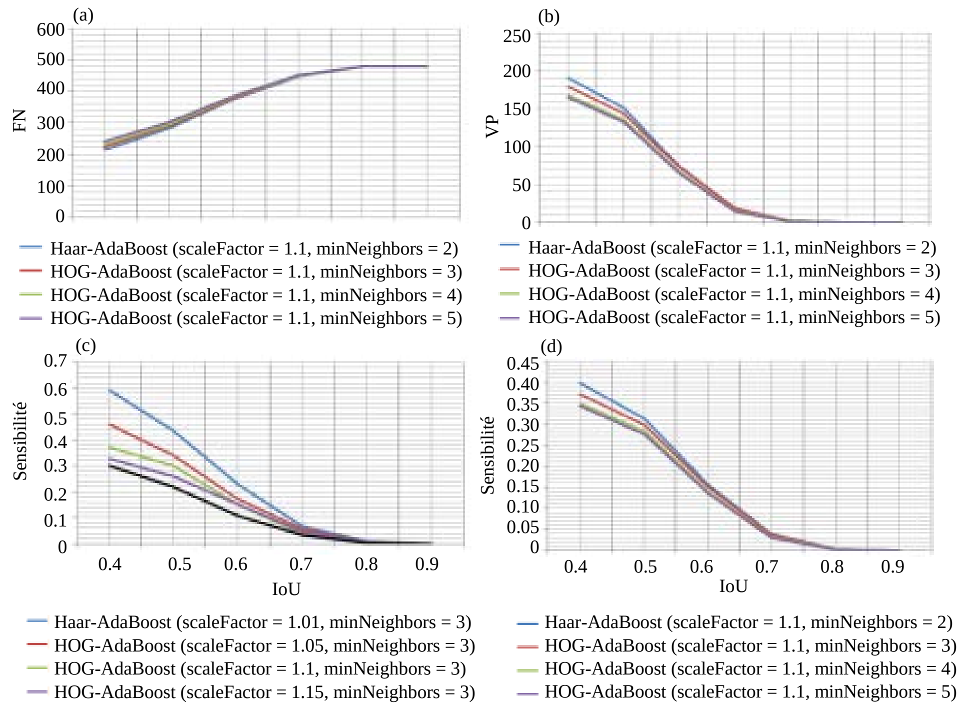


Fig. 8(a-d): HOG AdaBoost detectors. These curves show the results of analysis of the detections obtained by the HOG AdaBoost detector by setting the value of the scaleFactor parameter to 1.1 and varying the value of the minNeighbors parameter by successively assigning it the values 2, 3, 4 and 5; (a) Courbe FN-IoU, (b) Courbe VP-IoU, (c) Courbe sensibilité-IoU and (d) Courbe sensibilité-IoU

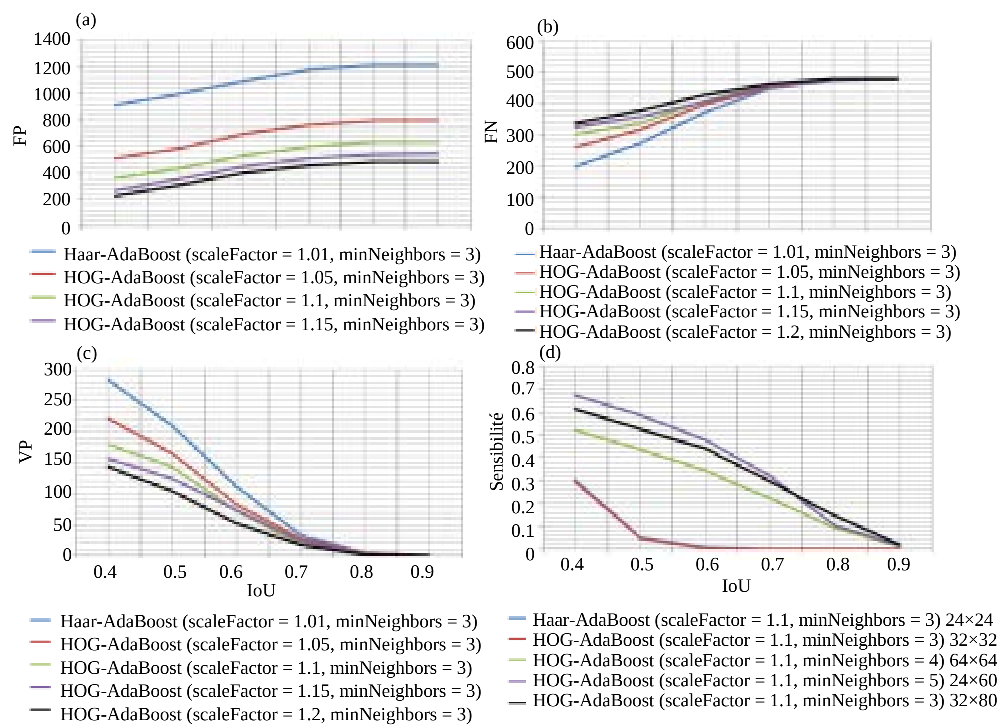


Fig. 9(a-d): Détecteur Haar AdaBoost. These curves show the results of analysis of the detections obtained by the Haar AdaBoost detector by varying the value of the scaleFactoren parameter assigning it the successive values 1.01, 1.15, 1.1, 1.15 and 1.2 and setting the value of the minNeighbors parameter to 3; (a) Courbe FP-IoU, (b) Courbe FN-IoU, (c) Courbe VP-IoU and (d) Courbe sensibilité-IoU

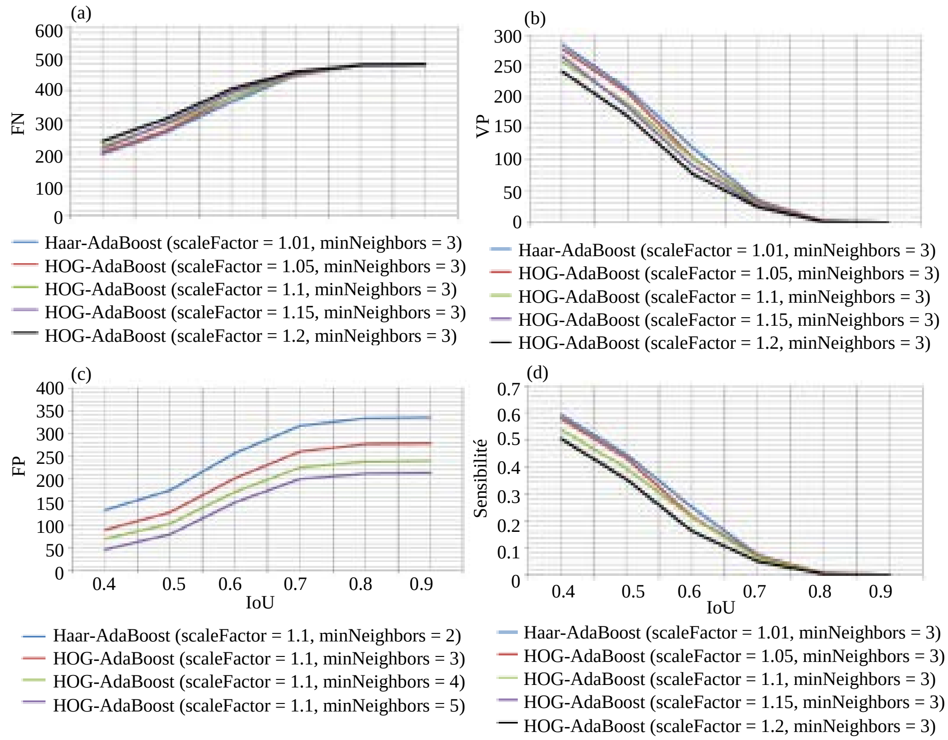


Fig. 10(a-d): Détecteur HOG AdaBoost. These curves show the results of analysis of the detections obtained by the HOG AdaBoost detector by varying the value of the scaleFactor parameter by assigning it the successive values 1.01, 1.15, 1.1, 1.15 and 1.2 and setting the value of the minNeighbors parameter to 3; (a) Courbe FN-IoU, (b) Courbe VP-IoU, (c) Courbe FP-IoU and (d) Courbe sensibilité-IoU

Training of a Haar-AdaBoost detector by supervised learning on the images of INRIA person dataset: Training a classifier is a long step. It requires gathering and annotating a large number of images containing the object to be detected (positive images) and images not containing the object to be detected (negative images).

In our case, we used the 187 images taken from 460 images from the INRIA Person Dataset database (see subsection 2.6.1). Manual labeling of people in these 187 images resulted in 481 field truth framing boxes that we will use to train the Haar AdaBoost detector by supervised learning.

The training database therefore, consists of 187 positive images and 273 negative images (also called background images). All these images are taken from the INRIA person dataset database. The positive images are labeled in 481 people who will be used in conjunction with the negative images during the learning process as training examples of the Haar AdaBoost detector.

The aim of this experiment is to test whether, we can improve the detection of people on a medium and large scale by injecting into the database of learning examples images of people on medium and large scales. During the learning phase, the training of the detector takes a lot of time depending on the number of positive and negative images and the size w×h. In our case, the number of positive and negative images was set at 481 and 273, respectively. The training time of the Haar-AdaBost detector for each VEC file that we generated in step 3 increases according to the size w×h used.

Par example, ce temps prend 1 heure et 50 min pour la taille 24×24, 3 jours et 21 heures pour la taille de 32×32 et plus de 5 jours pour les tailles 64×64, 24×60et 32×80.

After training the detectors we trained for sizes 24×24, 32×32, 64×64, 24×60 and 32×80, we applied them to the INRIA images to analyze the results obtained.

Figure 11 shows some images of people detections obtained by Haar AdaBoost detectors formed with sizes 64×64, 24×60 and 32×80. The frames in blue are the terrai truth framing boxes, while the green frames correspond to the predicted (or detected) framing boxes. These detections were obtained with the values 1.1 and 3 assigned, respectively to the scaleFactor and minNeighbors parameters of the detect MultiScale method of the CascadeClassifier class provided by OpenCV (see subsection 2.6.6).

The detection results obtained with Haar AdaBoost detectors formed with sizes 24×24, 32×32 are very bad, there is practically no detection of people and generate a very high number of false detections (see the curves in blue and red that are often confused in Fig. 12).

On the other hand, the results obtained by the detectors formed with sizes 64×64, 24×60 and 32×80 are

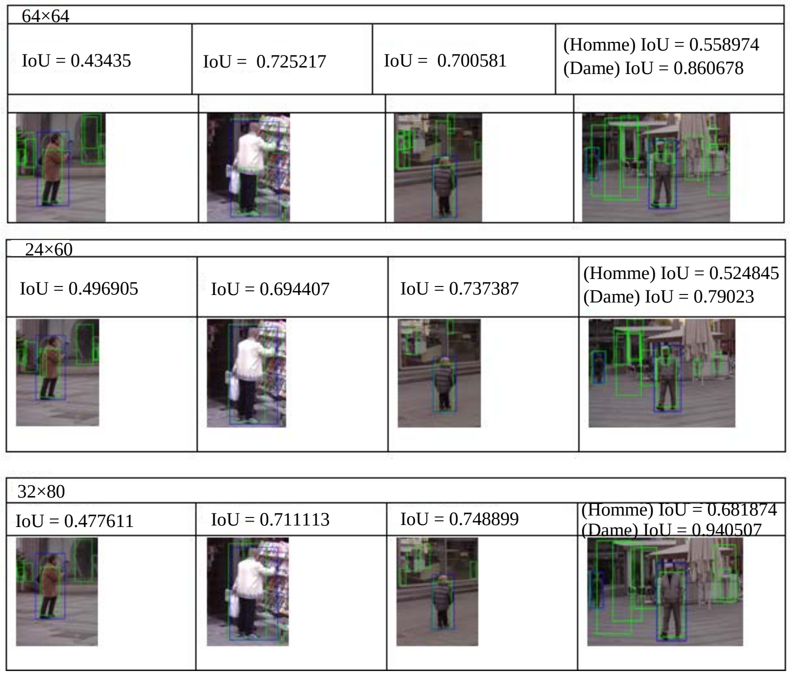


Fig. 11: Détection de personnes obtenue par les détecteurs Haar AdaBoost formés respectivement avec les tailles 64×64, 24×60, 32×80

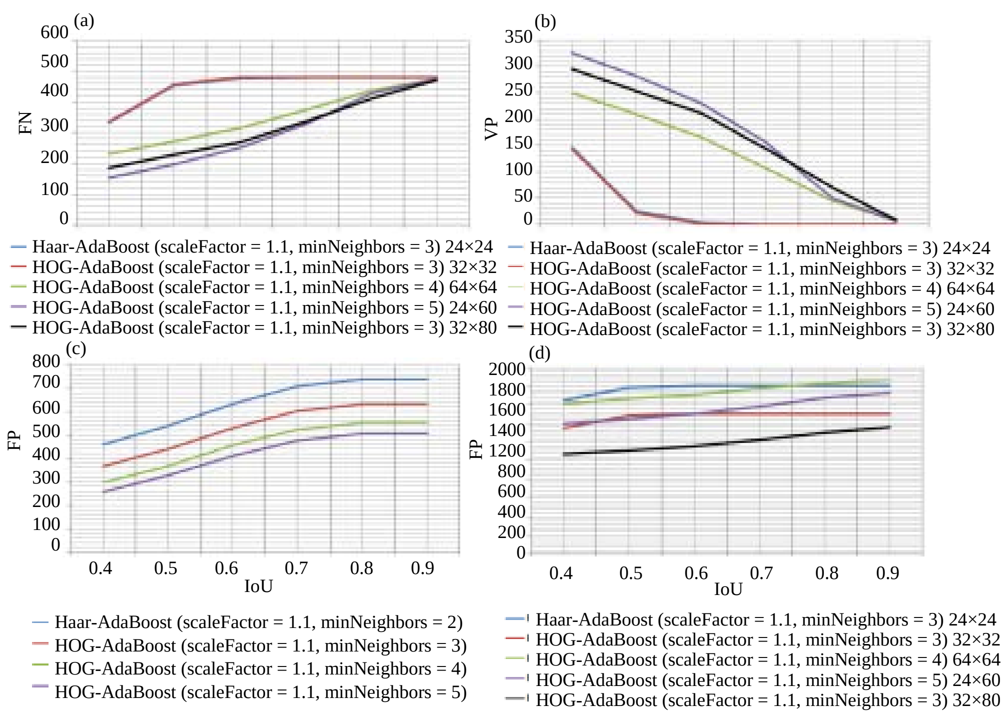


Fig. 12(a-d): These curves show the results of analysis of the detections obtained by the Haar AdaBoost detector formed for sizes 24×24, 32×32, 24×60, 32×80.The scaleFactor and minNeighbors parameters have values of 1.1 and 3, respectively; (a) Courbe FN-IoU, (b) Courbe VP-IoU, (c) Courbe FP-IoU and (d) Courbe FP-IoU

suitable, they practically resemble the detection results obtained with the detector provided by OpenCV. In addition, since, the learning examples of these detectors contained many people on a large scale, they thus made it possible to slightly exceed the OpenCV detector in terms of detecting people on a large and medium scale (Fig. 11 and 12). In addition, the VP IoU, FN IoU and IoU sensitivity curves in Fig. 11 also show that detectors trained for sizes 24×60 and 32×80 provide a better detection result with regard to the number of true positives that is larger and the number of false detections that are smaller than those provided by the detector formed for size 64×64. This result comes from the fact that the aspect ratio, that is, the ratio of width to height, chosen for the detectors 24×60 and 32×80 is equal to 0.4

Unfortunately, the disadvantage of these detectors thus formed is that they generate a very high number of false detections compared to those generated by the OpenCV detector (Fig. 11 and 12), this is most likely due to the number of negative (273 images) and positive (481 positive images of people) examples of learning which is very low.

Normally, to properly train a detector, it actually takes thousands of positive and negative examples which requires gathering a very large number of positive images containing people to be labeled and negative images not containing people. In this case, the training of the detector will require a very high learning time.

CONCLUSION

In this article, we first studied the two descriptors Haar AdaBoost (VJ) and HOG AdaBoost (PoseInv) following a study that we did before in the paper[14].

After having studied the two methods we made a comparison of two approaches Haar-AdaBoost and HOG-Adaboost which constitutes a variant of HOG SVM[9]. Secondly and after modifying certain detection parameters, we carried out an evaluation of the experiments found to have more performance.

The application of these two detectors on the images taken from the INRIA Person Dataset database enabled us to draw the following conclusions:

The HOG AdaBoost detector is more efficient at detecting people on a medium scale (or nearby) than the Haar AdaBoost detector but on the other hand, it generates many more false detections than the latter.

Generally, the two detectors studied do not correctly detect people on a small scale (or distant) and on a very large scale (or very close). This is most likely due to the training examples that were used to train these two detectors which contained very few examples of people on a small and very large scale.

Sometimes the shape of the clothing, people close together, crowds, etc. can prevent these detectors from properly detecting people in images. The detection rate of people obtained by the two detectors Haar AdaBoost and HOG AdaBoost is <50%. In an attempt to improve the detection of people at medium and large scale, five Haar AdaBoost detectors were formed for the respective image sizes 24×24, 32× 32, 64×64, 24×60 and 32×80 and on an image database containing many examples of medium and large scale people (see sub-section 2.6.6).

Detection results provided by 24×24 and 32×32 detectors are virtually zero. In contrast, 64×64, 24×60 and 32×80 detectors have improved the performance of detecting people at medium and large scale compared to the detector provided by OpenCV but on the other hand, they generate a very high number. high false detections. This disadvantage is probably due to the reduced number of positive and negative images used to train these detectors.

Unfortunately, it is not possible to apply the fine tuning operation to the Haar AdaBoost and HOG AdaBoostqui detectors. This operation consists of re-training a detector already trained on new examples in order to readjust it so that they can adapt to the recognition of these new examples such as for example in our case, the detection of small, medium and small people. in large scale.

In practice, the fine-tuning operation is preferable to training a new detector on a new sample database which is a very computationally expensive operation. The fact that this operation is not supported by Haar AdaBoost and HOG AdaBoost, this is a disadvantage of these detectors, as it will be difficult to expand the capacity of these detectors to new examples.

Another disadvantage of the Haar and HOG descriptors is that they only allow you to process grayscale images and only take into account the shape of the objects.

An alternative to the Haar AdaBoost and HOG AdaBoost detectors is to use deep convolutional neural network models. Indeed, the latter have made it possible to obtain great performances by their training for the detection of objects and in particular of people[6].

In addition, it is very easy to expand the capacity of an already trained deep convolutional neural network to new learning examples through the fine-tuning operation. Deep convolutional neural networks also have the advantage of being applied to color images which gives them the ability to take into account not only the shape of objects but their texture and color as well.

REFERENCES

01. Wang, M.S. and Z.R. Zhang, 2018. FPGA implementation of HOG based multi-scale pedestrian detection. Proceedings of the 2018 IEEE International Conference on Applied System Invention (ICASI), April 13-17, 2018, IEEE, Chiba, Japan, pp: 1099-1102.

02. Zhao, J., J. Li and Y. Ma, 2018. RPN+ fast boosted tree: Combining deep neural network with traditional classifier for pedestrian detection. Proceedings of the 2018 4th International Conference on Computer and Technology Applications (ICCTA), May 3-5, 2018, IEEE, Istanbul, Turkey, pp: 141-150.
03. Zhang, X., L. Cheng, B. Li and H.M. Hu, 2018. Too far to see? Not really!-Pedestrian detection with scale-aware localization policy. IEEE. Trans. Image Process., 27: 3703-3715.
04. Wu, B. and R. Nevatia, 2006. Tracking of multiple, partially occluded humans based on static body part detection. Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), June 17-22, 2006, IEEE, New York, USA., pp: 951-958.
05. Viola, P., M.J. Jones and D. Snow, 2005. Detecting pedestrians using patterns of motion and appearance. Int. J. Comput. Vision, 63: 153-161.
06. Zhang, S., R. Benenson, M. Omran, J. Hosang and B. Schiele, 2016. How far are we from solving pedestrian detection?. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, IEEE, Las Vegas, Nevada, pp: 1259-1267.
07. Zhu, Q., M.C. Yeh, K.T. Cheng and S. Avidan, 2006. Fast human detection using a cascade of histograms of oriented gradients. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Volume 2, June 17-22, 2006, New York, USA., pp: 1491-1498.
08. Zhang, S., R. Benenson and B. Schiele, 2015. Filtered feature channels for pedestrian detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 7-12, 2015, IEEE, Boston, Massachusetts, pp: 1751-1760.
09. Dalal, N. and B. Triggs, 2005. Histograms of oriented gradients for human detection. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognition, 1: 886-893.
10. Singh, S. and R. Mittal, 2015. Image segmentation using edge detection and poincare mapping method. Eur. J. Adv. Eng. Technol., 2: 81-83.
11. Viola, P. and M. Jones, 2001. Rapid object detection using a boosted cascade of simple feature. IEEE Conf. Comput. Vision Pattern Recogn., 1: 511-518.
12. Papageorgiou, C.P., M. Oren and T. Poggio, 1998. A general framework for object detection. Proceedings of the 6th International Conference on Computer Vision (IEEE Cat. No. 98CH36271), January 7, 1998, IEEE, Bombay, India, pp: 555-562.
13. Wu, B. and R. Nevatia, 2007. Detection and tracking of multiple, partially occluded humans by bayesian combination of edgelet based part detectors. Int. J. Comput. Vision, 75: 247-266.
14. Taleb, N.O., A. Chergui, M.L.B. Maati and M.F. Nanne, 2016. Overview on automatic detection of human body. Proceedings of the 2016 5th International Conference on Multimedia Computing and Systems (ICMCS), September 29-October 1, 2016, IEEE, Marrakech, Morocco, pp: 671-675.
15. Viola, P. and M.J. Jones, 2004. Robust real-time face detection. Int. J. Comput. Vision, 57: 137-154.
16. Bodla, N., B. Singh, R. Chellappa and L.S. Davis, 2017. Soft-NMS-improving object detection with one line of code. Proceedings of the IEEE International Conference on Computer Vision, October 22-29, 2017, IEEE, Venice, Italy, pp: 5561-5569.