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

Nagi OULD TALEB[1], Mohamed Larbi BEN MAATI[2]
Mohamedade Farouk NANNE[3], Aicha Mint Aboubekrine[4], Adil CHERGUI[5]
Department of Computer Sciences, CSSE Laboratory, University of Abdelmalek Essaadi, Tetouan, Morocco[1, 2]
Department of Computer Sciences, University of Nouakchot Al Asriya, Nouakchot, Mauritania[3]
Department of Computer Sciences, LIROSA Laboratory, University of Abdelmalek Essaadi, Tetouan, Morocco[4]
Department of Computer Sciences, CSSE Laboratory, ENSAM-Casablanca, Tetouan, Morocco[5]

***Abstract*—Object detection 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. The detection of people is a particularly difficult subject because of the great variability of appearances and the situations in which a person may find themselves (a person is not a rigid object; it is articulate and unpredictable; its shape changes during its movement). The situations in which a person may find themselves are very varied: They are alone, near a group of people or in a crowd, obscured by an object. In addition, the characteristics vary from one person to another (color of the skin, hair, clothes, etc.), the background simple, clear or complex, the lighting or weather conditions, the shadow caused by different light sources, etc. greatly complicate the problem. In this article, we will present a comparative study of the performance of the two detectors Haar-AdaBoost and HOG-AdaBoost in detecting people in the INRIA images database of persons. An evaluation of the experiments will be presented after making certain modifications to the detection parameters.**

***Keywords*—Pedestrian detection; learned-based methods; Haar-like features; HOG descriptor; AdaBoost; behavior analysis**

I. Introduction

The detecting of people in images is a very important subject in the field of computer vision. The pedestrians' detection 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 [4] and Pfinder [5]. Other methods have been conducted [6, 7, 8, 9] 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 [10, 11].

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 [10, 12, 13]. Viola and Jones [14, 15] [16] 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 [17] that proposed a detector based on the Haar wavelet.

The latest so-called hybrid approach combines local and global characteristics to improve recognition performance [18].

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

II. Learning Methods for Human Detection

The main approaches based on discriminate 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 (misses the object) or mistakenly detects the absent 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.

The Fig. 1 shows an illustration of the data formation process that is common basic in all detectors.

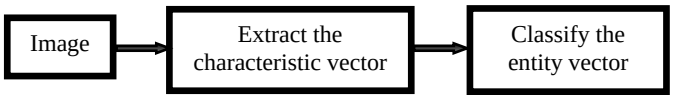


Fig. 1. Common Learning Process.

III. Study of Haar-AdaBoost and HOG-Adaboost Detectors for the People Detection

The study which we carried out in the paper [19] 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, Viola and Jones) [14, 15] and HOG-AdaBoost (PoseInv, Pose-Invariant) [20] which constitutes a variant of HOG-SVM [12] (SVM design a Support Vector Machine classifier [21]) to study their feasibility for detecting people.

Note that the detectors Haar-AdaBoost and HOG-AdaBoost uses respectively the Haar-like features (or Haar wavelets) [14, 15, 17] and HOG (Histograms of Oriented Gradients) descriptor [12] to extract the characteristic vector from an image of person and they are based on the same classifier AdaBoost [21] to classify this vector. AdaBoost is one of the most powerful binary classification methods in supervised learning, its uses an iterative selection of weak classifiers based on a distribution of learning examples. Each example is weighted according to its difficulty with the current classifier. The main motivation for boosting was to form a procedure which combines the output of several weak binary classifiers to produce a powerful binary classifier [22].

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 evaluing people detection in images using the two detectors Haar-AdaBoost and HOG-AdaBoost.

The performance analysis of these two detectors was carried out on the people images database 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 TP-IoU, FP-IoU, FN-IoU, IoU-Recall curves and the evaluation of the AR (Average Recall) metric. Plotting the Precision-Recall 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.

Here is the meaning of the TP, FP, FN, Precision, Recall and IoU metrics:

- TP: True Positive, also called detection, is a correctly detected person (or object).
- FP: False Positive, also called false detection, occurs when the predicted box provided by the classifier does not contain any person to be detected.
- FN: False Negative, denotes a case where a person is missed.
- Recall: is the number of true positives divided by the sum of true positives and false negatives, this last sum is just the number of ground-truths boxes: $\text{Recall} = \frac{\text{TP}}{\text{TP+FN}}$. This metric measures the rate of true positives detected among all positives, so it is a measure of detector performance in finding positives.
- Precision: is the number of true positives divided by the sum of true positives and false positives: $\text{Precision} = \frac{\text{TP}}{\text{TP+FP}}$. This metric measures the rate that the detection is correct, so it is used to measure the accuracy of the detection.
- IoU: Intersection over Union, is the ratio of the area of the intersection between the predicted bounding box and the ground-truth bounding box on the area of their union.

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 Dalal and Triggs detector [12]).

IV. INRIA Person Dataset Image Database and Manual Images Labeling

Among the 460 images in the INRIA Person Dataset, we have manually annotated only 187 images from the first images of this dataset using the objectmarker annotation program, resulting in a total of 481 ground-truth bounding boxes. But, it was better to annotate all the images in the database, which requires more effort and time. Also note that during the annotation, we ignored some images containing a single person very close (or on a very large scale) that we did not consider interesting, the number of this last images was very little.

The Fig. 2 shows four images from the INRIA Person Dataset labeled using the objectmarker program. Each person presented in these images is manually framed using a rectangle, called a ground-truth bounding box. 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.

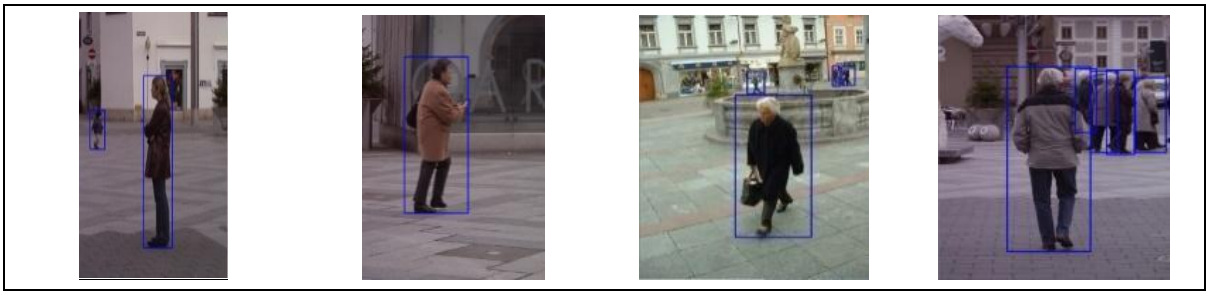


Fig. 2. People Labeled Manually in the INRIA Person Dataset Images using the Objectmarker Program and they are Framed with a Blue Ground Truth Bounding Box. The Four Images show Respectively the Labeling of One Person, Two Persons and Four Persons at different Scales and then One Person with a Crowd of People.

V. Comparison of the Two Detectors Haar-AdaBoost and HOG-AdaBoost using Examples of People Detection in the INRIA Person Dataset

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 first 187 images that we were precedently annotated in the INRIA Person Dataset images. We will discuss the strengths of each of these two detectors as well as their failing.

Table I shows some examples of people detection obtained respectively by the application of the two detectors Haar-AdaBoost and HOG-AdaBoost on the 187 images tagged in the INRIA Person Dataset. 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 bounding 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.

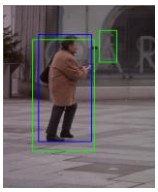
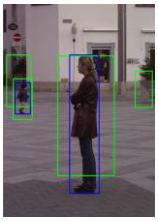
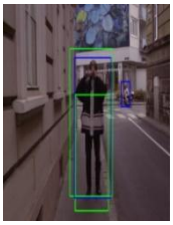
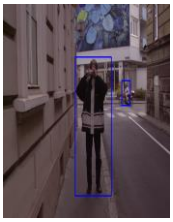
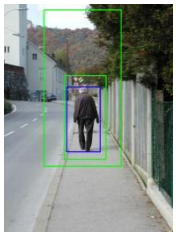
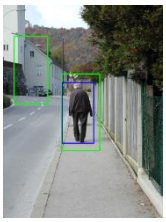
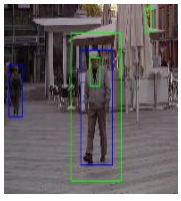
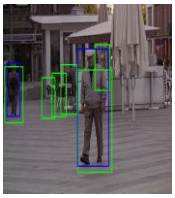
Experimentation with Haar-AdaBoost and HOG-AdaBoost detectors on images from the INRIA Person Dataset allowed the following conclusions to be drawn:

- 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, small predicted framing boxes whose IoU with their associated ground truth boxes is of small value.
- Sometimes the shape of the clothes (especially if a person is wearing a coat or a dust jacket) 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 bounding 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 lower than 0.5, this leads to an erroneous interpretation of the detections result obtained. This situation rarely happens with the HOG-AdaBoost detector where the IoU of detection is often greater than 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 ground-truth bounding boxes are manually drawn too tight to the person that they frame, the detector can sometimes generate an IoU with the predicted bounding box less than 0.5.

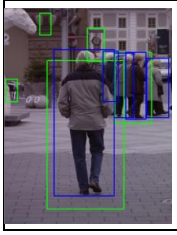
Based on the analysis of the detection results obtained by the Haar-AdaBoost and HOG-AdaBoost detectors, we found that for the minimum IoU threshold value set at 0.4, almost all detections that give rise to true positive are correctly determined. Therefore, to study the performance of the two detectors, 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 (TP-IoU), False Positive as a function of the IoU (FP-IoU), False Negative as a function of the IoU (FN-IoU) and Recall as a function of the IoU (Recall-IoU).

TABLE I. THIS TABLE SHOWS SOME PEOPLE DETECTION RESULTS OBTAINED RESPECTIVELY BY APPLYING THE TWO HAAR-ADABOOST AND HOG-ADABOOST DETECTORS TO THE IMAGES IN THE INRIA PERSON DATASET

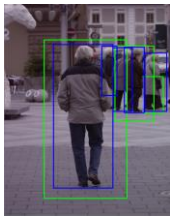
Haar-AdaBoost detector	**HOG-AdaBoost detector**
The person on a large scale (or at a close distance) is not detected. This is therefore a false negative.	 Here, the same large-scale person is indeed detected with an IoU equal to 0.755259, this is a true positive. The image also shows a false detection that matches the green frame on the side of the display case, so this is a false positive.
The medium-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 as 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.	The 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).
 The same person is detected twice, the two predicted bounding boxes in green have respectively for IoU 0.622019 and 0.793165. Only the box with the maximum IoU, that is 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 jacket worn by the person who trained it to go undetected by the HOG-AdaBoost detector.
 Same thing as the previous example, the same person at medium scale is detected twice with two predicted bounding boxes having respectively for IoU of 0.623512 and 0.181492. In this case the box having the IoU of 0, 623512 will be considered as a true positive, however the one with an IoU of 0.181492 will be considered as 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 medium-scale person is detected with a predicted bounding box having an IoU equal to 0.457869. If the IoU threshold is taken equal to 0.5, then the ground-truth bounding box in blue will be considered as a false negative and the predicted bounding box will be considered as 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 with the green boxes predicted having respectively IoU equal to 0.48906 and 0.0806955, the box with IoU 0.0806955 will be rejected and considered as a false positive. Likewise, the box with the IoU equal to 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 with the predicted bounding boxes 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 medium-scale lady is also detected with the predicted bounding box which has the IoU equal to 0.612489.

Both persons are detected, but they have respectively for IoU 0.437131 (person on the left) and 0.518307 (person on the right). 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 bounding 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 persons are well detected with respectively IoU of 0.543696 (person on the left) and 0.5461 (person on the right). 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 bounding box overlaps with multiple ground truth framing boxes, but only the ground truth framing box having the highest IoU with 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. With the IoU threshold set at 0.5, the detection in this example gives 1 true positive, 4 false negatives and 2 false positives.

VI. Simple Version of the TP, FN and FP Metrics Evaluation Algorithm

To evaluate the TP (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 with a predicted bounding box. In other words, there is a single predicted bounding box associated with the ground-truth bounding box framing a detected person.

Algorithm: *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 ground truth bounding boxes and the list of predicted bounding boxes.*
- *The minimum threshold of the IoU.*

Output: *TP, FN et FP.*

We initialize: TP = 0 and FP = 0.
For each image of the database:
    For each detection (or predicted frame box) in the current image:
        Choose from all the ground-truth bounding boxes labeled in the image, the one that has the highest IoU with the predicted bounding box.
        If all the ground-truth bounding 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
        else:
            The detection is a true positive and we increment TP : TP = TP + 1
Since each predicted bounding box corresponds to one and only one ground-truth bounding box (or a person in the image), one can easily calculate FN by:
FN = Total number of ground-truth bounding boxes in the database – TP

This simple algorithm has the advantage of quickly calculating TP, FN, and FP metrics, but unfortunately it is only suitable if the detector effectively returns a single predicted bounding 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 last one sometimes returns two predicted bounding boxes for the same person and therefore this box will be counted twice as a true positive. Whereas normally only one predicted bounding 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 TP, 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 bounding boxes returned by a detector for the same object (or person) labeled in an image using a ground truth-bounding box.

VII. General Version of the TP, FN and FP Metrics Evaluation Algorithm

It is assumed that the same person can be detected more than once, that is, there are several predicted bounding boxes which may correspond to the ground-truth bounding box framing that person.

Here are two problems that can arise when it comes to associate predicted bounding boxes with ground-truth bounding boxes (or detected person):

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

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

To overcome these two difficulties and correctly evaluate the TP, FN and FP metrics, that is to say, to avoid repeatedly counting the same person detected with several predicted bounding boxes, which will distort the calculation of TP and FP, we propose a general algorithm whose idea is based on the principle of Non Maximum Suppression (NMS) [23]. We have used this late one to associate each ground-truth bounding box (or detected person) with the predicted bounding box that maximizes the IoU with it and eliminate other predicted bounding boxes that do not maximize the 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, see Fig. 3.

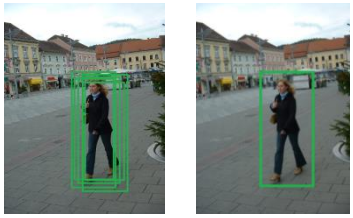


Fig. 3. For the Same Person in the Image, a Multitude of Windows are Detected (the Image on the Left). We must Determine which one Best Frames the Person (the Image on the Right). The Confidence Score is used by the non-maximum Elimination Technique to Find the Window that Maximizes it and to Eliminate the others that do not.

The Non Maximum Suppression 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 people detection, we based it on the comparison of the IoU between the predicted bounding 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 bounding boxes that correspond to it.
- The Haar-AdaBoost and-HOG-AdaBoost detectors are respectively based on the binary classifier AdaBoost which do not return a confidence score, but rather the values -1 (or a negative value) for non-membership of the object class to be detected or 1 (or a positive value) to indicate membership of the object class.

The principle of the general evaluation algorithm for TP, FN and FP metrics that we have developed is as follows:

Algorithm: *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 bounding boxes and the list of predicted bounding boxes.*
- *The minimum threshold of the IoU.*

Output: *TP, FN and FP.*
We initialize TP = 0, FN = 0 and FP = 0.
For each image of the database perform the following processing:

- *Mark all bounding boxes predicted as not being assigned to a ground-truth bounding box.*
- *Associate with each ground-truth bounding box (or a labeled person) in the current image the list of predicted bounding boxes that have an IoU with it that exceeds a certain threshold (typically 0.5). The list of predicted bounding boxes is sorted in descending order of IoUs and all predicted bounding boxes in the list are marked as affected.*
  If the list of predicted bounding boxes is empty, that is, there is no predicted bounding box associated with the ground-truth bounding box, and then this last one is a false negative or a missed person. In this case, we increment the FN metric.
  The final goal of the algorithm is to determine for a ground-truth bounding box framing a person detected in the image one predicted bounding box that maximizes the IoU with it. In this case, only this predicted box will be counted as a true positive, the other predicted bounding 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 bounding boxes that are not marked as assigned to a ground-truth bounding box.*
- *If several detected bounding boxes correspond to the same ground-truth bounding box framing a person (or object in general), the Non Maximum Suppression principle is applied to keep only the detected bounding*

box having a maximum IoU with the ground-truth bounding box. This operation is necessary to properly calculate the TP number, as it avoids counting the predicted boxes for a detected person several times.
For each ground-truth bounding box b1 in the current image:
If the list of predicted bounding boxes associated with box b1 is not empty, then:
The predicted bounding box p1 at the beginning of the list has the maximum IoU. We then take this box p1.
For each ground-truth bounding box b2 in the current image:
If the box p2 at the beginning of the list of predicted bounding boxes associated with box b2 is the same as p1:
If the IoU of p2 with b2 is greater than that of p1 with b1 then it can be confirmed that the predicted bounding 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 list of the predicted bounding boxes associated with the box b2.
If in the previous iteration it was determined that the predicted bounding box p1 was not associated with b1, then in this case p1 is removed from the beginning of the predicted bounding boxes list associated with box b1.

- *Evaluate the TP metric: it corresponds to the number of ground-truth bounding boxes with a list of predicted bounding boxes associated with them non-empty (these ground-truth bounding boxes therefore correspond to detected people).*

VIII. Analysis of the Two Detectors Haar-AdaBoost and HOG-AdaBoost Performance on the Images of the INRIA Person Dataset

After having implemented the general algorithm for evaluating TP, 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.

The following two tables show the results of analyzes obtained by applying respectively the two detectors on 187 first images taken from the 460 bitmap color images of people in the INRIA Person Dataset. Manual labeling of the people in the 187 images resulted in a total of 481 ground-truth bounding boxes framing these people.

TABLE II. Result Obtained by Applying the Haar AdaBoost-Detector on 187 Images from INRIA Person Dataset Containing 481 People Labeled with Ground-Truth Bounding Boxes

Haar-AdaBoost					
IoU threshold	TP	FN	FP	Precision	Recall
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 III. Result Obtained by Applying the HOG-AdaBoost Detector on 187 Images from the INRIA Person Dataset Containing 481 People Labeled with the Ground-Truth Bounding Boxes

HOG-AdaBoost					
IoU threshold	TP	FN	FP	Precision	Recall
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

Since the two detectors are applied to the same images in the INRIA Person Dataset, we will start by making a simple comparison by plotting the curves of the True Positives as a function of the Intersection over Union (TP-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. 4).

It can be seen from the TP-IoU curve in Fig. 4 that the HOG-AdaBoost detector (curve in red) is more efficient than the Haar-AdaBoost detector (curve in blue), since it allows to detect more positives (or the persons labeled) than Haar-AdaBoost.

Likewise, the FP-IoU curve also shows that there are fewer false negatives or misses' people 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.

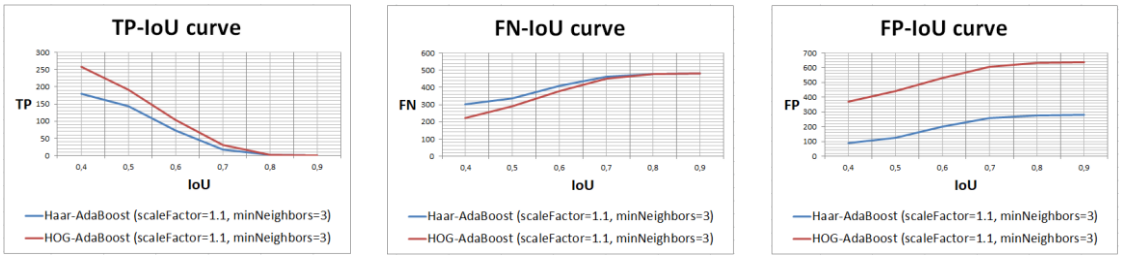


Fig. 4. These Curves show that the HOG-AdaBoost Detector (in Red) is more efficient to Detect People than Haar-AdaBoost (in Blue), but it Generates more False Detections than the Latter. These Results were Obtained for the Values of the Detection Parameters ScaleFactor and MinNeighbors respectively Equal to 1.1 and 3.

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 required class or not, it will therefore not be possible to use the curve Precision-Recall that can be used to calculate the AP (Average Precision) metric. We will therefore use in its place the Recall-IoU curve which makes it possible to calculate the AR (Average Recall) metric.

Subsequently, we will complete the comparisons made by the curves in Fig. 4 by plotting the Recall-IoU curve (Fig. 5). This is more general than the previous curves, it is often used to study the efficiency in detecting true positives; in addition it allows evaluating the average recall metric AR (Average Recall) which is used to compare detectors even if they are applied to different image databases.

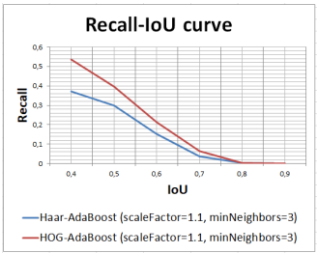


Fig. 5. This Curve shows that the HOG-AdaBoost Detector (in Red) is more Efficient at Detecting People than the Haar-AdaBoost Detector (in Blue).

Knowing that the AR metric corresponds to the area of the region below the Recall-IoU curve between the values of IoU 0.5 and 1 and, it is given by equation (1). To evaluate this metric, we will approximate the integral (1) using the rectangle method that is given by the equation (2):

$$AR = 2\int_{0,5}^{1} Recall(IoU) \quad (1)$$

$$AR = 2\sum_{i=1}^{n-1}(IoU_{i+1} - IoU_i)\, Recall(IoU_{i+1}) \quad (2)$$

Here, IoU_1 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 $IoU_{i+1} - IoU_i = \frac{1-0,5}{n} = \frac{0,5}{n}$, $1 \le i \le n - 1$.

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 $\frac{1-0,5}{n} = \frac{0,5}{5} = 0,1$, equation (2) will become:

$$AR = 2 \times 0,1 \times \sum_{i=1}^{n-1} sensibility(IoU_{i+1}) \quad (2)$$

With IoU1 = 0,5 and IoUi+1 = IoUi + 0,1, $1 \le i \le n - 1$.

Based on the detections results obtained by the Haar-AdaBoost and HOG-AdaBoost detectors and which are presented in Tables II and III, 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 Recall-IoU curve and the evaluation of the AR metric for both detectors, the HOG-AdaBoost detector is more efficient to detect people than Haar-AdaBoost, because the HOG-AdaBoost AR is larger than the Haar-AdaBoost AR. But unfortunately, according to the FP-IoU curve in Fig. 4, the HOG-AdaBoost detector has the disadvantage of generating a lot of false detections than the Haar-AdaBoost detector.

IX. Experimenting by Changing Certain Detection Parameters

To perform the detection of people, we used the detectMultiScale method of the CascadeClassifier 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 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 other parameters are: the image matrix, the flags (not used in detection), minSize (minimum size of the object, the default value is size 0×0) and maxSize (maximum size of the object, the value by default is the full size of the image) are not important for detection.

The results of the analyses presented in paragraph 8 above were obtained using the values of scaleFactor and minNeighbors parameters respectively equal to the default values 1.1 and 3.

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-AdaBoost are respectively shown in Fig. 6 and Fig. 7.

The TP-IoU, FN-IoU and Recall-IoU curves show that when the value of the minNeighbors parameter decrease from 5 to 2, 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 in Fig. 8 (Haar-AdaBoost) and Fig. 9 (HOG-AdaBoost) were obtained for the values of (scaleFactor,minNeighbors) equal to (1.01, 3), (1.05,3), (1.1,3), (1.15,3) and (1.2,3), they give an idea for comparing the

detections that we obtained by varying the values of the scaleFactor and minNeighbors parameters as shown above.

The analyses we performed for the scaleFactor parameter value varying from 1.01 to 1.2 and the minNeighbors parameter value fixed at 3 and which are illustrated by Fig. 8 and 9 allowed us to deduce the following conclusions:

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

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

Tables IV and V 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 scaleFactor parameter value which successively takes the values 1.01, 1.05, 1.1, 1, 15 and 1.2.

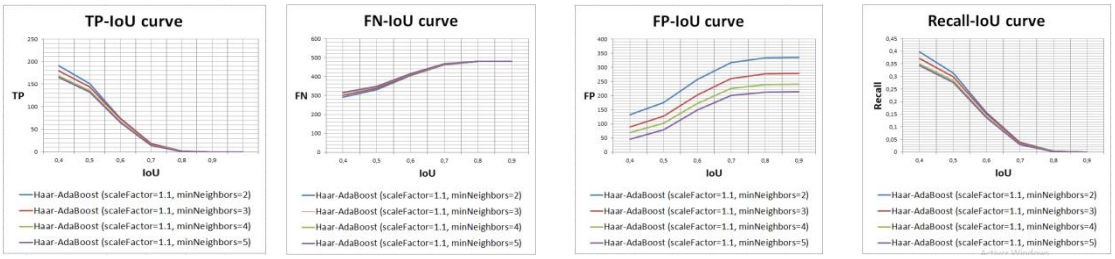


Fig. 6. Haar-AdaBoost Detector.These Curves show the Results of the Detections Analysis Obtained by the Haar-AdaBoost Detector by Setting the Value of the ScaleFactor Parameter to 1.1 and Varying the Value of the minNeighborsen Parameter Successively Assigning it the Values 2, 3, 4 and 5.

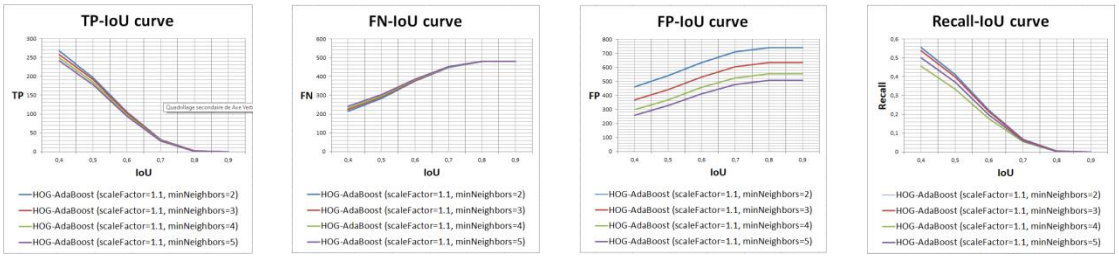


Fig. 7. HOG-AdaBoost Detector. These Curves show the Results of the Detection Analysis Obtained by the HOG-AdaBoost Detector by Setting the Value of the ScaleFactor Parameter to 1.1 and Varying the Value of the minNeighbors Parameter by Successively Assigning it the Values 2, 3, 4 and 5.

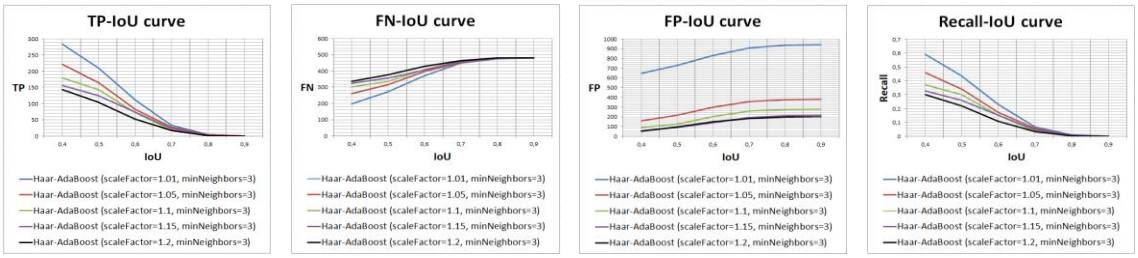


Fig. 8. Haar-AdaBoost Detector. These Curves show the Results of Detections Analysis Obtained by the Haar-AdaBoost Detector by Varying the Value of the ScaleFactor Parameter Assigning it the Successive Values 1.01, 1.05, 1.1, 1.15 and 1.2 and Setting the Value of the minNeighbors Parameter to 3.

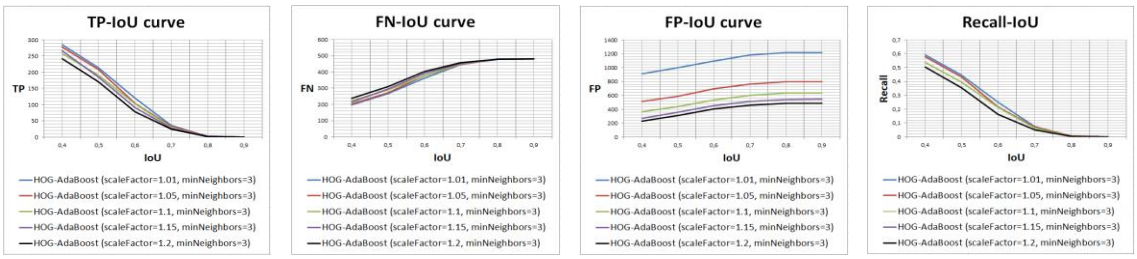


Fig. 9. HOG-AdaBoost Detector. These Curves show the Results of Detections Analysis Obtained by the HOG-AdaBoost Detector by Varying the Value of the ScaleFactor Parameter Assigning it the Successive Values 1.01, 1.05, 1.1, 1.15 and 1.2 and Setting the Value of the minNeighbors Parameter to 3.

Knowing that in the first 187 images of the INRIA Person Dataset, we have labeled 481 people by ground-truth bounding boxes, in this case, the detection rate (equal to the recall metric) will therefore be equal to the number of true positives detected in all 187 images divided by 481, that is, equal to $\text{Recall} = \frac{\text{TP}}{481}$.

It can be seen from Tables IV and V 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.

TABLE IV. 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

Haar-AdaBoost detector					
scaleFactor	1.01	1.05	1.1	1.15	1.2
TP	210	165	144	125	105
Detection rate in %	43,66	34,30	29,94	25,99	21,83

TABLE V. DETECTION RATES OBTAINED BY THE HOG-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

HOG-AdaBoost detector					
scaleFactor	1.01	1.05	1.1	1.15	1.2
TP	215	208	191	186	171
Detction rate in %	44,70	43,24	39,71	38,67	35,55

- The respective default values 1.1 and 3 of the two parameters scaleFactor and minNeighbors are central; they allow obtaining a suitable detection result that provides a compromise between true detections and false detections and a reasonable calculation time.
- The parameters (scaleFactor,minNeighbors) pairs of values (1.15,3) and (1.2,3) also provide a suitable detection result, since the TP-IoU, FN-IoU and FP-IoU and Recall-IoU curves obtained for these two pairs of values are very close to those obtained for the pair value (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).

X. TRAINING THE HAAR-ADABOOST DETECTOR BY SUPERVISED LEARNING ON THE INRIA PERSON DATASET IMAGES

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 first 187 images taken from 460 images in the INRIA Person Dataset. The manual people labeling in these 187 images resulted in 481 ground-truth bounding 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. The positive images are labeled in 481 people who will be used jointly 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 learning database examples of people images on medium and large scales. During the learning phase, the training of the detector with the opencv_traincascade program 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-AdaBoost detector increases according to the used size w×h.

For example, this time takes 1 hour and 50 minutes for the size 24×24, 3 days and 21 hours for the size of 32×32 and more than 5 days for the sizes 64×64, 24×60 and 32×80 on the Intel (R) Core (TM) microprocessor having the frequency of 1.8 GHz and a RAM memory of 4 GB.

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

Fig. 10 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 ground-truth bounding boxes, while the green frames correspond to the predicted bounding boxes.

These detections were obtained with the values 1.1 and 3 assigned respectively to the scaleFactor and minNeighbors parameters of the detectMultiScale method defined in CascadeClassifier class provided by OpenCV.

The detection results obtained by 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. 11).

On the other hand, the results obtained by the detectors formed with sizes 64×64, 24×60 and 32×80 are suitable, they resemble practically to the detection results obtained with the detector provided by OpenCV.

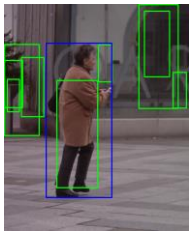
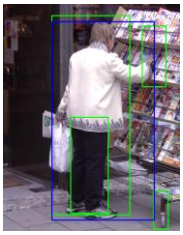
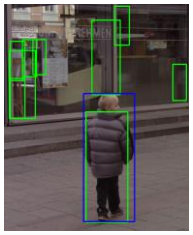
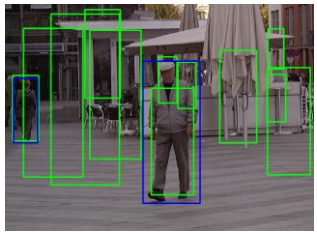
In addition, since the learning examples of these detectors contained many people on a large scale, they thus made it possible to slightly exceed the OpenCV detector in terms of detecting people on a large and medium scale (see Fig. 10 and 11).

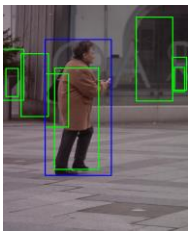
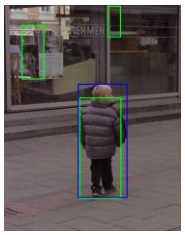
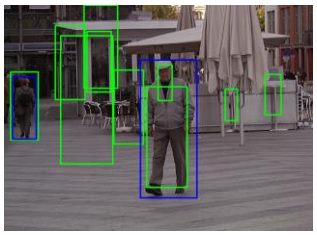
In addition, the TP-IoU, FN-IoU and Recall-IoU curves in Fig. 11 also show that detectors trained for sizes 24×60 and 32×80 provide a better detection result with regard to the number of true positives that is larger and the number of false detections that are smaller than those provided by the detector 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 ($\frac{24}{60} = \frac{32}{80} = 0.4$) which generally corresponds to the aspect ratio of people standing.

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 (see Fig. 10 and 11), this is most likely due to the number of negative (273 images) and positive (481 positive images of persons) examples of learning which is very low.

Normally, to train correctly 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.

64×64			
IoU = 0.43435	IoU = 0.725217	IoU = 0.700581	(man) IoU = 0.558974 (lady) IoU = 0.860678
			

24×60			
IoU = 0.496905	IoU = 0.694407	IoU = 0.737387	(man) IoU = 0.524845 (lady) IoU = 0.79023
			

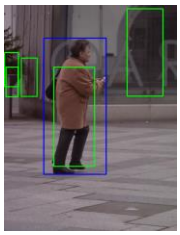
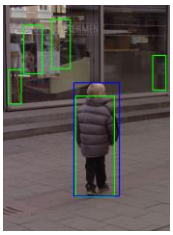
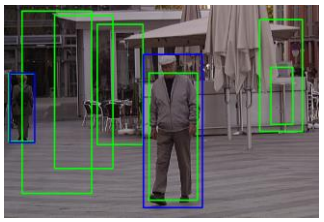
32×80			
IoU = 0.477611	IoU = 0.711113	IoU = 0.748899	(man) IoU = 0.681874 (lady) IoU = 0.940507
			

Fig. 10. People Detection Obtained by Haar-AdaBoost Detectors Formed respectively with Sizes 64×64, 24×60 and 32×80.

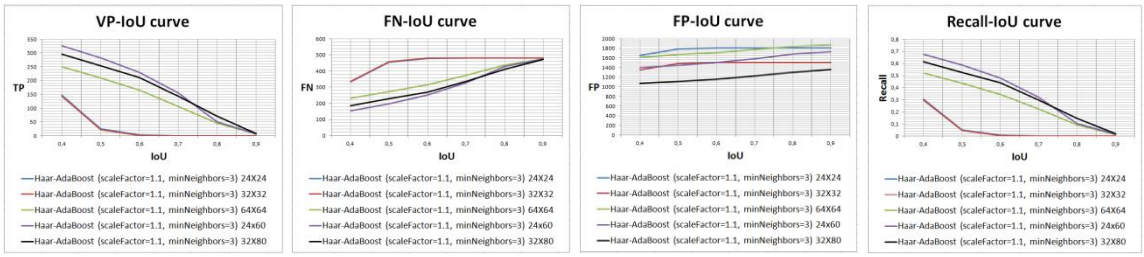


Fig. 11. These Curves show the Results of Detections Analysis Obtained by the Haar-AdaBoost Detector Formed for Sizes 24×24, 32×32, 24×60 and 32×80.The ScaleFactor and minNeighbors Parameters have Values of 1.1 and 3, respectively.

XI. CONCLUSION

In this article, we first studied the two detectors Haar-AdaBoost (VJ) and HOG-AdaBoost (PoseInv) following the study that we did before in the paper [19].

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 [12]. 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 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 much more false detection 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 on a 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 less than 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 on an image database containing many examples of medium and large scale people.

There are practically no detection results provided by 24×24 and 32×32 detectors. 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 of false detections. This disadvantage is probably due to the reduced number of positive and negative images that we used to train these detectors.

Unfortunately, it is not possible to apply the finetuning operation to the Haar-AdaBoost and HOG-AdaBoost 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 large scale people.

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 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 [24, 25, 26] and in particular of people [27, 28, 29, 30, 31, 32].

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

[1] Ming-ShiWang and and Zhe-Rong Zhang, “FPGA implementation of HOG based multi-scale pedestrian detection”, Proceedings of IEEE International Conference on Applied System Innovation,ISBN={978-1-4503-5614-5}, 2018.

[2] Jia Xiang Zhao and Jun Li, “RPN+ Fast Boosted Tree: Combining deep neural network with traditional classifier for pedestrian detection”, ACM Digital Library, Volume 47 Issue C, Sep 2016.

[3] Xiaowei Zhang Li Cheng, Hai-Miao Hu, “Too far to see? Not really! pedestrian detection with scale–aware localization policy”, University Library, Volume {abs/1709.00235}, 2017.

[4] J. Leskovec, Sentjost, and Slovenia, “Detection of human bodies using computer analysis of a sequence of stereo images”. 11th European Union Contest for Young Scientists, 1999.

[5] B. Wu and R. Nevatia, “Tracking of multiple, partially occluded humans based on static body part detection” in IEE Proc. of the IEEE International Conference on Computer Vision, pp. 886–893, 2005.

[6] P. Viola, M. J. Jones and D. Snow, “Detecting pedestrians using patterns of motion and appearance”, in International Journal of Computer Vision, pp. 153–161, 2005.

[7] Y. Benezeth, H. Laurent, B. Emile, C. Rosenberger, “Humain presence detection and charactezation of activity”, In XXIIe conference GRETSI (signal and image processing), Dijon (FRA), 2009.

[8] L. Biancardini, S. Beucher, L. Letelllier, "Object extraction in motion: a mixed approach contours-regions" ORASIS 2005 9th Congress Young Researches in Computers Vision, Clermont-Ferrand, France, May 2005

[9] S. Zhang, R. Benenson, M. Omran, J. Hosang and B. Schiele, “How far are we from solving pedestrian detection”, Cornell University Library, (2016).

[10] Qiang Zhu; Mei-Chen Yeh; Kwang-Ting Cheng; S. Avidan, "Fast human detection using a cascade of histograms of oriented gradients", Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on. Vol. 2. IEEE, 2006.

[11] S. Zhang, R. Benenson, B. Schiele, “Filtered feature channels for pedestrian detection”, in ‘CVPR’, IEEE Conference on Computer Vision and Pattern Recognition, pp. 1751-1760, (2015).

[12] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, , pp. 886-893 vol. 1. (2005).

[13] S. Singh, R. Mittal, “Image segmentation using edge detection and poincare mapping method”. European Journal of Advances in Engineering and Technology, Volume2, Issue4, p. 81-83. (2015).

[14] P. Viola and M. Jones. “Rapid object detection using a boosted cascade of simple features”, in IEE Proc. Of the conference on Computer Vision and Pattern Recognition, pp. 511-518, 2001.

[15] P. Viola and M. Jones "Robust real-time face detection". International Journal of Computer Vision, Spring Journal, pages 137–154, May 2004.

[16] P. Viola, M. J. Jones and D. Snow, "Detecting Pedestrians Using Patterns of Motion and Appearance", in International Journal of Computer Vision, pp. 153–161, 2005.

[17] C. Papageorgiou, M. Oren and T. Poggio, "A general framework for object detection", in Proc. of the IEEE International Conference on Computer Vision, pp. 555-562,1999.

[18] B. Wu, R. Nevatia, "Detection and tracking of multiple, partially occluded humans by bayesian combination of edgelet based part detectors", International Journal of Computer Vision 75, no. 2 (2007): 247-266. CA 90089- 0273, 2007.

[19] N. Ould Taleb, A. Chergui, M. L. Ben Maâti, M. F. Nanne, "Overview on automatic detection of human body", IEEE Xplore, 24 April 2017, ISSN: 2472-7652. DOI: 10.1109/ICMCS.2016.7905638.

[20] Z. Lin, L. S. Davis, A Pose-Invariant descriptor for human detection and segmentation. In Computer Vision – ECCV 2008. Lecture Notes in Computer Science, vol 5305. Springer, pp. 423-436. Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-88693-8_310.

[21] V. Vapnik, The nature of statistical learning theory. Springer-Verlag, 1995.

[22] Yoav Freund and Robert Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, Journal of Computer and System Sciences, vol. 55, no 1, 1997, p. 119-139.

[23] Navaneeth Bodla, Bharat Singh, Rama Chellappa, Larry S. Davis, "Soft-NMS – Improving object detection with one line of code" ,Computer Vision (ICCV), IEEE International Conference, 2017

[24] Y. LeCun, B. Boser, J.S. Denker, D. Henderson R.E. Howard, W. Hubbard, and L.D Jackel (1989). Backpropagation applied to handwritten zip code recognition. Neural computation, MIT Press.1(4):541–551.

[25] S. Ren, K. He, R. Girshick. and J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, IEEE Transactions on Pattern Analysis and Machine Intelligence, June 2016.

[26] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed4, Cheng-Yang Fu1, Alexander C. Berg (2016). SSD: Single shot multiBox detector, European Conference on Computer Vision, pages 21-37, Oct 2016.

[27] L. Zhang, L. Lin, X. Liang., K. He, Is faster R-CNN doing well for pedestrian detection?. In: Leibe B., Matas J., Sebe N., Welling M. (eds) Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science, vol 9906. Springer, Cham. https://doi.org/10.1007/978-3-319-46475-6_28.

[28] Shanshan Zhang, Rodrigo Benenson, Mohamed Omran, Jan Hosang and Bernt Schiele. How far are we from solving pedestrian detection. Cornell University Library, (2016).

[29] Nagi Ould Taleb, Adil Chergui, Mohamed Larbi Ben Maâti, Mohamedade Farouk Nanne, Mohamed O. M. Khelifa, Aicha Mint Aboubekrine, Pedestrian detection and the effect of diverse benchmarks. International Journal of Scientific & Engineering Research, Volume 9, Issue 9, pages 922-927. September 2018.

[30] Ujjwal, Aziz Dziri, Bertrand Leroy, Francois Bremond (2018). Late fusion of multiple convolutional layers for pedestrian detection. 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). DOI: 10.1109/AVSS.2018.8639083.

[31] ZHANG, X., CHENG, L., LI, B., AND HU, H.-M (2018). Too far to see? not really! A pedestrian detection with scale-aware localization policy. IEEE transactions on image processing 27, 8 (2018), 3703–3715.

[32] Farzin Ghorban, Javier Marín, Yu Su, Alessandro Colombo, and Anton Kummert, "Aggregated channels network for real-time pedestrian detection", Proc. SPIE 10696, Tenth International Conference on Machine Vision (ICMV 2017), 106960I (13 April 2018); https://doi.org/10.1117/12.2309864.