Real time human detection in video streams

FATMA SAYADI*, YAHIA SAID, MOHAMED ATRI AND RACHED TOURKI
Electronics and Microelectronics Laboratory
Faculty of Sciences Monastir, 5000 Tunisia
Address (12pt Times New Roman, centered)
TUNISIA
sayadi_fatma@yahoo.fr, said.yahia1@gmail.com, mohamed.atri@fsm.rnu.tn,
Rached.tourki@fsm.rnu.tn

Abstract: - Detecting humans in films and videos is a challenging problem owing to the motion of the subjects, the camera and the background and to variations in pose, appearance, clothing, illumination and background clutter. We develop a detector for standing and moving people in videos, testing several different motions coding schemes and showing empirically that orientated histograms give the best overall performance. The resulting detector is tested on several databases including a challenging test set taken from feature films and containing wide ranges of pose, motion and background variations, including moving cameras and backgrounds. Experimental results from real time video are provided. These results are validated on a challenging test set containing 1000 human examples.

Key-Words: - Human detection, video streams, HOG, SVM, Haar fatures , ROC

1 Introduction

Finding people in images has attracted much attention in recent years for practical applications such as visual surveillance [1,2].

In this context we concentrate on the most successful and popular vector-form feature: histograms of oriented gradients (HOG) [3,4,5]. It is shown that the HOG features are based on the contrast of silhouette contours against the background. The second step of human detection is designing classifier. Large generalization ability and less classifying complexity are two important criteria for selecting classifiers. Linear support vector machine (SVM) [6] and AdaBoost [7] are two widely-used classifiers satisfying the criteria. In this paper, we place emphases on the HOG feature and the SVM classifier and our contributions lie in efficiently computing of HOG features for human detection from video. In fact Histograms of Oriented Gradients (HOG) plus Support Vector Machine (SVM) [6] is one of the most successful human detection algorithms [3,4,7,8]. The HOG+SVM algorithm [3,4] employs sliding window principle to detect humans in an image. It scans the image at different scales and at each scale examines all the sub images. In each sub image, a 3780-dimensional HOG feature vector is extracted and SVM classifier [6,9] is then used to make a binary decision:

human or non-human. In this paper we propose to use HOG in video streams. Such a detection process is very slow. The computational costs are often too high to real-time applications. To overcome the drawback, we develop a sub-cell algorithm to accelerate the calculation of the HOG features in one block, which removes unnecessary (at least unimportant) regions while the necessary regions are remained. We also proposed to use an appearance-based approach to facial expression recognition which is based on Haar features and the Adaboost boosting algorithm.

This combined approach was employed by Viola and Jones in [10] for face detection and has demonstrated both high recognition accuracy and fast run-time performance. However, in most cases the classification accuracy is lower than that of the first proposed algorithm based on HOG+ SVM.

The rest of the paper is organized as follows: in Section 2, we describe traditional human detector algorithm. In Section 3, we describe the proposed method. The experimental results are presented in Section 4. Finally, we provide a brief summary in Section 5.

2 Human Detector Algorithm

Despite all the difficulties on human detection, a lot of work has been done recent years.

First, we may use different features such as edge, Haar features and gradient orientation features; second, we may use different classifiers such as Nearest Neighbor, Neutral Network [11,12], SVM and Adaboost.

The HOG+SVM algorithm concentrates on the contrast of silhouette contours against the background [3,4]. Different humans may have different appearances of wears but their contours are

ISBN: 978-1-61804-103-6

similar. Therefore the contours are discriminative for distinguishing humans from non-humans. It is worth noting that the contours are not directly detected. It is the normal vector of the separating hyper plane obtained by SVM that places large weights on the HOG features along the human contours. The HOG+SVM algorithm is outlined as follows: Input: The scaled input image at the current scale. The size of sliding detection window is 64x128. The sliding step d (d=8forexample). Output: The locations of the sub_images of size 64x128, which are declared to contain humans.

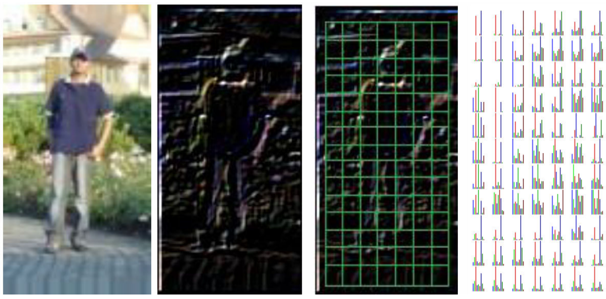


Figure 1: Histograms of Oriented Gradients principle.

As shown in figure 1, The first step is to compute for each pixel of the whole image, the magnitude $|\nabla f(x,y)|$ and the orientation $\Theta(x,y)$ of the gradient $\nabla f(x,y)$. In Step 2: From top to bottom and left to right, scan the whole image with a 64x128 window. Extract the 3780 HOG features from the sub image covered by the detection (scanning) window and then apply the leaned SVM classifier on the high-dimensional HOG feature vector to classify the sub image as human or non-human. In fact SVM is one of the margin-based classifiers [6]. It aims at finding an optimal hyper- plane, which leads to maximal geometric margin.

While Haar-like features encode the existence of oriented contrasts between regions in the image. A set of these features can be used to encode the contrasts exhibited by a human face and their spacial relationships. The training dataset consists of images and videos of human and non-human examples. The detector uses images as inputs and efficiently extracts simple rectangular features using integral images. A cascade of classifiers is created to achieve superior detection and low false positives. Each stage of the classifier is trained on true and false positives from the previous stage using Adaboost to select weak classifiers (which are the simple rectangular features mentioned earlier). The detector is fast and provides good detection results on a large pedestrian dataset.

In this paper, we concentrate on the discriminative power of HOG features for human detection and the real-time property of Viola's face detection framework. In fact methods based on Haar features have been used successfully in face detection and also adopted by a lot of researchers for human detection. Nevertheless HOG descriptors [4] have attracted more attention.

3 Proposed Approach

Initially the program carried out consists in reading an image and detecting the existing people in this image, using HOG, HOG optimized descriptor or Haar fatures. Secondly the program must read a video streams recorded in the hard drive or collected by the webcam or by an external camera; then make the same treatment for each video frame. To be done we create a human –machine interface according to figure 2.

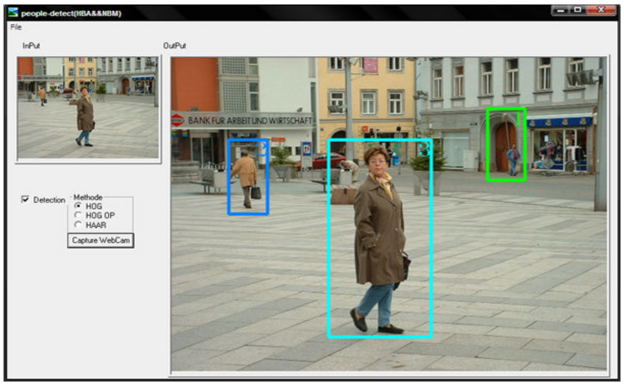


Figure 2: Homepage of the application

Huns we must test if the passed arguments are a static image or a recorded video stream. If there are no arguments the function directly will seek the existence of a webcam or a camera installed on the computer.

3.1 HOG for person detection in video:

In the case of a video stream, we treat the flow components as independent images, we take their local gradients separately and we compute HOGs as in static image (figure 3).

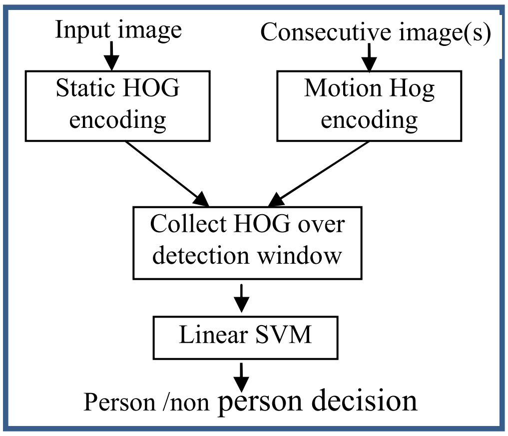


Figure 3: Motion HOG processing

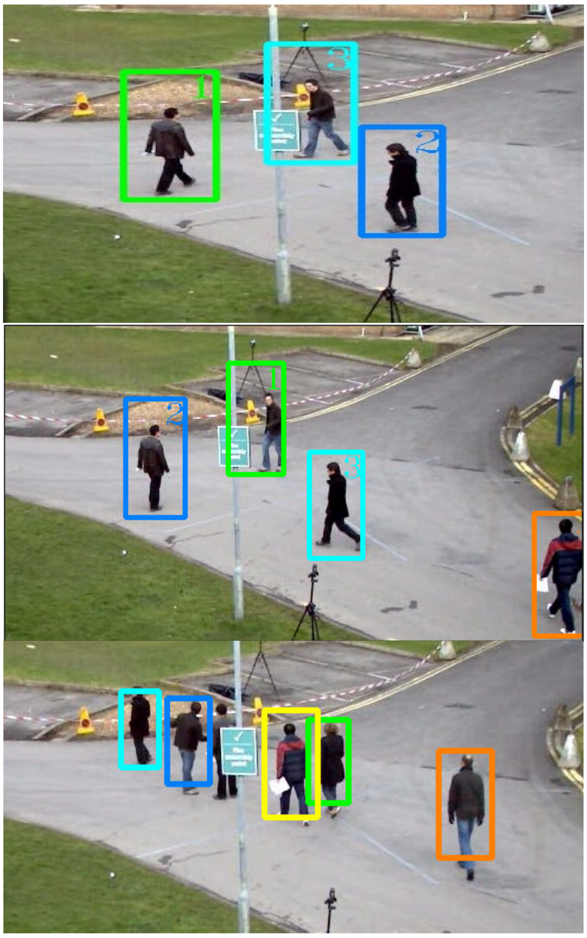


Figure 4: Examples of detection results using HOG. Detected humans are enclosed in rectangles.

The obtained results figure 4 confirm the reports of Yanwei Pang [13] while Histograms of Oriented Gradients plus Support Vector Machine (HOG+SVM) is the most successful human detection algorithm, it is time-consuming. In fact table 1 shows that's our system is slower when using big size images.

Image size (pixels)	Detection time (ms)
70*134	32
512*512	3000
480*640	4000
768*576	7095

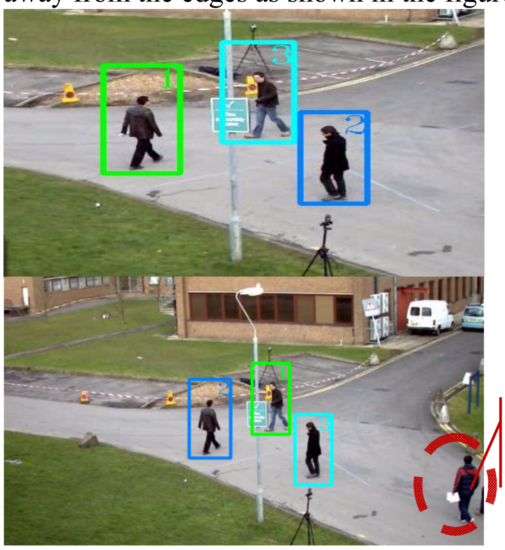
Table 1: Speed comparisons using HOG-SVM

To deal with this problem we have proposed an optimized HOG algorithm. We used the implementation in OpenCV and we noticed that "detectMultiscale" is the most critical function compared to the total time of detection. So we have suggested resizing the image by eliminating a certain number of rows. The number of lines and columns eliminated is obtained after having made several tests. For example for 70x134 image we eliminate the 9*2 lines and 7*2 columns from the edge and for a 576x768 image we eliminate 41*2 lines and 85*2 columns from the edge.
It is clear that the optimized HoG algorithm results in a significant speedup over the pure Hog as shown in table 2. The optimization achieved reduced the total detection time by about 50%.

Image size	HOG	Optimized
768*576	7095ms	3566ms

Table 2: Execution time of detection for the HoG descriptor versus for the optimized HOG.

The experimental results proved the effectiveness of this optimization in the case of persons being far away from the edges as shown in the figure 5.



None detected person in the border

Figure 5: Human detection using optimized HOG
The major disadvantage of this method is that if a person or part of a person is in the borders; he

ISBN: 978-1-61804-103-6

cannot be detected, and the probability of the detection decreases especially if the person size is very small in the image.

3.2 Haar for person detection in video

HAAR features have been used successfully in face detection and also adopted by a lot of researchers for human detection In fact Viola et al. [14] extend the haar features to capture spatial-temporal information for moving-human detection in surveillance system but adopt the boosted cascade classifier.

In this section we try to profit from real-time properties of Viola's face detection to detect persons in video stream.

For the HAAR implementation we used a program developed with OPEN_CV2.2 "facedetect.cpp" which is used for the face detection. Thus we introduce some modifications aiming to adapt this program to the human detection.

It is clear that the HAAR method is much faster than HoG, as shown in table 3 HAAR method is running in 1034ms on DVD resolution 768 × 576 images against 7095ms for HOG.

Image size	HOG	Haar
768*576	7095ms	1034ms

Table 3: Execution time of detection for the HoG descriptor versus HAAR features

Unfortunately the detection results were not competitive. It was noticed that this method presents much false positives as shown in the figure 6.

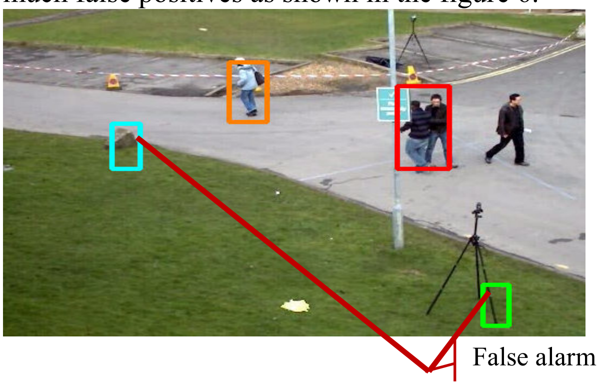
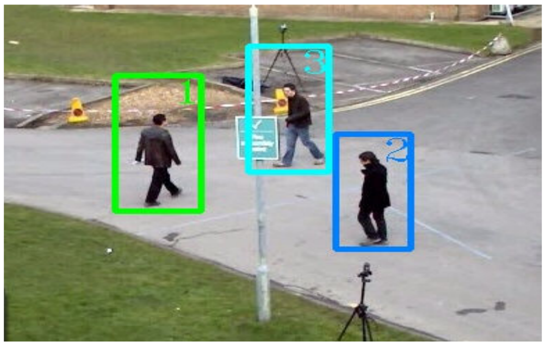


Figure 6: False Alarm using HAAR features

We can see also From figure 7a and 7b that Haar decreases significantly the detector performances compared to the optimized HOG performances



a). Detection using optimized HOG

b) Detection using HAAR

Figure 7: Person 2 and 3are detected only by HOG descriptor

4 Interpretation Of The Experimental Results

There are some striking differences in the classification and detection performance of different systems reported in previous sections. The variations come from different test criteria. In order to make objective and fair comparisons, we use the INRIA publicly and available database to compare the classification and detection performance of different systems. To evaluate the proposed methods, we have established a top-view human database.

To quantify the performance of the various detectors we plot the Receiver Operating Characteristic (ROC) curve. We start by calculating the different points of the curve by using Visual Studio 2008 and the Opencv2.2 library. In fact each point on the ROC curve represents a sensitivity/specificity pair corresponding to a particular decision threshold.

We tested the different detectors with the INRIA person database. The total training session is composed by 1000 positive images (humans) and 1000 negatives (not humans) taken from random windows cropped from no human images.

The true positive rate (Sensitivity) is plotted in function of the false positive rate (100-Specificity) for different cut-off points of a parameter. To be

ISBN: 978-1-61804-103-6

done for the three methods (HOG, HOG optimized and Haar) we used Matlab.

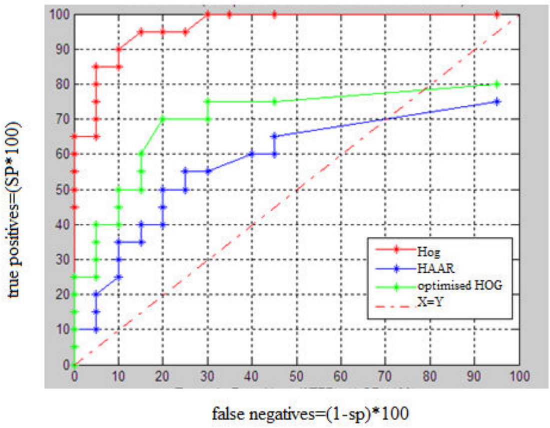


Figure 8 :ROC curve of various detectors

The ROC curves (figure8) show the performances achieved by the three methods. As expected the features with HOG+SVM are performing better than the others. However, and depending on the applications we can merely use the optimized HOG which gives a good ratio accuracy/time execution.

5 Conclusion

We have presented three ways for human detection in video streams. One way is to use HOG features. This first method seems to be time-consuming nevertheless it is a very successful human detection algorithm. The second way is to optimize the HOG algorithm by resizing the image this method results in more than two times increase in detecting humans in a 768*576 image while scarifying in the detection accuracy mainly when persons are on the edges. The third way consists in adapting Haar feature for human detection this solution significantly reduces computational cost in spite of the precision.
To evaluate the proposed method, we have established a top-view human database. Experimental results have demonstrated the effectiveness and efficiency of the proposed algorithm which gives a good ratio accuracy/time execution.

References:

[1] Roland Perko, Ales Leonardis, “A framework for visual context-aware object detection in still images,” Computer Vision and Image Understanding. 114, 700-711 (2010).

[2] Hou Beiping, Zhu Wen, “Fast Human Detection Using Motion Detection and Histogram of Oriented Gradients”, Journal of computers. 6(8), (2011).

[3] NavneDalal, “Finding people in images and videos”, Ph.D.thesis, INRIA Rhone-Alpes, (2006).

[4] Navneet Dalal, Bill Triggs, “Histograms of oriented gradients for human detection”, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognitio. 2, 886–893 (2005).

[5] YuanYuan, YanweiPang, XuelongLi, “Foot wear for gender recognition”, IEEE Transactions on Circuits and Systems for Video Technology. 20(1), .131–134 (2010).

[6] ChistopherJ. C.Burges, A tutorial on support vector machines for pattern recognition in Data Mining and Knowledge Discovery2. 121–167 (1998).

[7] QiangZhu, Mei-ChenYeh, Kwang-TingCheng, S.Avidan, “Fast human detection using a cascade of histograms of oriented gradients”, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 1491–1498 (2006).

[8] XiaoyuWang, Tony X. Han, ShuichengYan, AnHOG-LBP “human detector with partial occlusion handing”, in: Proceedings of the IEEE International Conference on Computer Vision. (2009).

[9] DachengTao, XiaoouTang, XuelongLi, XingdongWu, “Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval”, IEEE Transactions on Pattern Analysis and Machine Intelligence. 28 (7), 1088–1099 (2006).

[10] P. Viola and M. Jones “Robust real-time object detection”. International Journal of Computer Vision. (2001).

[11] Xin-Zheng Wang, Xiao-chen Duan, “Application of Neural Network in the Cost Estimation of Highway Engineering,” Journal of computers..5, 1755-761 (2010).

[12] Long Wang, Yanheng Liu, Xiaoguang Li, “Analog Circuit Fault Diagnosis Based on Distributed Neural Network,”Journal of computers. 5,1747-1754 (2010).

[13] Yanwei Pang a, YuanYuan b,n, XuelongLi b, JingPan,”Efficient HOG human detection”, Signal Processing 91, 773–781 (2011)).

[14] P. Viola, M. Jones, and D. “Snow Detecting pedestrians using patterns of motion and appearance”.ICCV, (2003).