AIVP ADVANCES IN IMAGE AND VIDEO PROCESSING

SOCIETY FOR SCIENCE AND EDUCATION UNITED KINGDOM

VOLUME 5, ISSUE 3
ISSN 2054-7412

People Detection in Complex Scenes By Using An Improved and Robust HOG Descriptor

[1]Hdioud Boutaina, [2]Oulad Haj Thami Rachid, [3]El Haj Tirari Mohammed
[1,2]RIITM group Research, ENSIAS, Mohamed V Souissi University, Rabat, MOROCCO ;
[3]National Institute of Statistics and Applied Economics, Rabat, MOROCCO
hdioud.boutaina@hotmail.fr; rachid.ouladhajthami@gmail.com; mtirari@hotmail.fr

ABSTRACT

The detection of moving people in a complex scene filmed with a single camera is among the most difficult fields of research in vision by computer. In this work, we suggest improving the quality of detection methods based on the histogram of oriented gradients (HOG) descriptor. For that, we purpose to use a combination of type detector/descriptor to minimize the rate of the false detection produced by the descriptor HOG. The implementation of this combination as well as its evaluation on public bases show clearly that the technique which we propose produces many good results at the level of the detection of the people in movement compared with the descriptor HOG

Keywords: HOG ; SVM; Harris detector.

1 Introduction

The detection of people in a scene is a difficult task due to several factors, including the change in appearance, the wide variation of poses, colors and textures in the class of people as well as the complexity of the environment in which the person evolves. This is why the recognition of people in scenes requires the use of specific tools.

Several methods have been developed to solve the problem of detection of persons. A large part of these methods are based on a descriptor/classifier structure whose principle of operation consists in associating a descriptor with a classifier. The descriptor isolates the most discriminating information of the class sought and the classifier compares this information with examples of an elements of database the class sought in order to determine whether the tested element resembles these examples or not (whether the element is a person or not).

In this work, we have been interested in descriptor/ classifier detection methods, in particular the combination of the histogram of oriented gradients (HOG) descriptor with the SVM classifier. This choice is justified by the fact that HOG descriptors are currently the most used in object detection because they are the most discriminating of these objects and the best results are obtained by the approaches that use them.

Let us note that the challenge of all the systems of detection of people is to be able to recognize the people by avoiding the false negatives and the false positive. Through a series of tests which we realized with the combination of HOG descriptor with the classifier SVM, the obtained results reveal the existence of false detection. Thus, to improve the performances of this descriptor/classifier (HOG/SVM) by reducing the rate of false detection, we propose a new technique of detection of moving people.

DOI: 10.14738/aivp.53.3280
Publication Date: 20th June, 2017
URL: http://dx.doi.org/10.14738/aivp.53.3280

Our technique consists in replacing the HOG descriptor by a combination of type detector/descriptor while using the same classifier (SVM). The principle of our technique is detailed in the section 3.

This paper is organized as follows: Section 2 is devoted to the presentation of previous work on the detection of moving people. In Section 3, we present the principle and limitations of the HOG descriptor. The method put forward in this work to detect people are given in Section 4. The results obtained by our method will be detailed in Section 5. Finally, in conclusion, a discussion about the performances of the proposed method will be given in the last section...

2 Previous Research

Many methods of detection of persons have been proposed in the literature, the principle of which is the same as in the case of an image or a video sequence. These methods can be classified into two categories: those that use an explicit model (2D or 3D) of the shape of the human body [9, 10, 11,12] and those that are based on supervised learning techniques where a discriminator model is constructed based on characteristics of the shape of the human body extracted from an image base. Among the detection methods of this last category, we can cite the method of Papageorgiou et al. Proposed in [6] which is based on the combination of Haar wavelets and support vector machines (SVM); The one proposed by Viola et al. [7,8] which is based on the combination of the Haar filters and the Boosting algorithm, and finally the method proposed by Dalal et al. [3] which is based on the descriptor combination of the histograms oriented gradient (HOG) and SVM classifier.

3 Principle and Limitations of the HOG Descriptors

3.1 Principle HOG descriptors

The interest of using HOG descriptor by Dalal et al. [3] lies in its ability to isolate the most discriminating information of interest from the image. They have therefore proposed this descriptor whose principle is based on the fact that the appearance and the local form of an object in an image can be described by the distribution of the intensity of the gradient or the direction of the contours. The HOG descriptor computation can be done by dividing the image into adjacent small-sized regions, called cells, and calculating for each cell the histogram of the gradient directions or the edge orientations for the pixels within this cell. The combination of the histograms then forms the HOG descriptor. For the best results, a process of normalization of the cells with multiple covers can give better resistance to variations in illumination. Dalal et al. [3] extended the approach of the HOG descriptor to take into account the information provided by movement.

In the following, we will be interested in the descriptor HOG seeking to improve it by reducing its disadvantages. Indeed, this descriptor possesses a certain number of limits which are the following:

3.2 Limits of detection of people based on the HOG descriptors

The methods of detection of persons available present a good overall performance, but they are still perfectible. This is the case, for example, with those based on the HOG descriptor whose limits result from several causes and its errors are of two types: false positives and false negatives. The first errors, also called false detections, are manifested in boxes that do not contain people while the second correspond to the presence of persons who are not detected.

Copyright © Society for Science and Education United Kingdom

Figure 1 The images shows the false positive and negative.

Indeed, a pedestrian detection (person) is generally based on the contours of the elements of the scene to determine whether or not it is a person. Sometimes the detector confuses certain vertical structures such as window sills, poles and trees with people. Sometimes a detector may recognize both the person and the environment. As a result, the detection boxes obtained are larger than the person himself. One can also have the case where phantom detections appear near other detections. This is apparent when several nearby areas have a high score such as the case of several people together or a person approaching a parasitic structure (eg a pole).

Figure 2 The images shows the phantom detection.

To solve these problems of detection errors we propose to integrate the points of interest with the HOG descriptor, these points are belongs in all objects of interest. This integration should lead to a better consideration of the appearance of the pedestrians present in the scene. Thus, the system should be able to detect them even if the focus presents difficulties related to the perspective. Consequently, false positives will generally be much smaller after integration of points of interest with HOG since the system will have learned to recognize them correctly.

4 Proposed approach for the detection of people on the move

Extracting moving persons from a video sequence is a significant and fundamental step in many computer vision applications. A person detection method usually works one frame at a time, so it is not aware of the dynamics of the scene and the movement of objects, which makes it difficult to detect people. In this work, we used background subtraction method to extract the moving objects. The most common prototype for performing background subtraction is to find an explicit model of background. Foreground objects (objects in motion) are then detected by calculating the difference between the current frame and this background model [5].

After detecting the objects in motion, we will limit the search areas of our moving object and we bounded them with the minimal rectangles, then a Harris detector [4] is applied for positioning the points of interest [5] on these objects of interest. Once these points of interest are identified in the sequence, we propose to compute the HOG descriptor on these points instead of applying it on the whole image, that is to say, these descriptors are calculated around each region of interest identified by the Harris detector.

In order to improve the performance of methods based on the HOG descriptor in terms of object detection in a sequence of images, we propose the use of a combination of detector / descriptor type for the information isolation phase. More discriminating objects is presents in the scene.

4.1 The Principle of proposed combination of detector / descriptor

Many tests of the method of detection of people using the HOG descriptor, applied to several videos extracted from different databases show that the application of this method on scenes generates errors of false detections. That is why, we proposed a method based on a combination of detector/ descriptor type to reduce the rate of false detections of people present in the scene. The proposed method has enabled us to obtain performances detection exceeding the one based on the HOG descriptor.

For our method, the decision is obtained on the more reduced regions which were localized through the use of point of interest detector (the window detected by Harris). Then, in order to obtain a finer description of each window detected, this later is divided into 8×8 cells, the characteristics of the objects contained in the window are extracted by using the HOG descriptor. Thus, the feature vector relating to a detected window f is defined by:

$$\mathbf{V}_f = HOG_i \tag{1}$$

where i are the points of interest belonging to the windows detected by Harris. The use of this detector/ descriptor combination allowed us to reduce the false detections produced by the HOG descriptor (HOG/SVM).

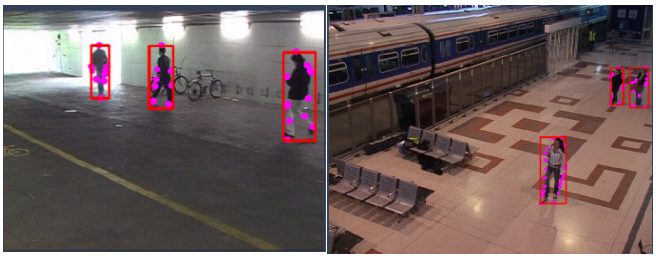

Figure 3 The results of ours combination detector/descriptor.

A schematic of the processing process is given in Fig. 4

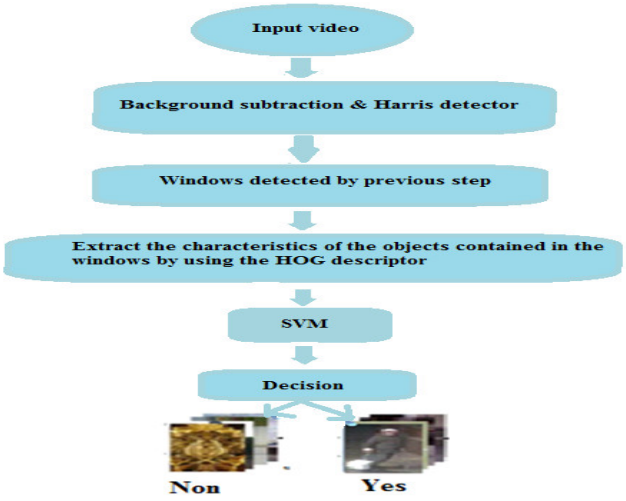

Figure 4 Processing process.

Copyright © Society for Science and Education United Kingdom

4.2 The SVM (Support vector machines) classifier

For our method of detection of persons, the classifier used is a Support vector machines (SVM) which constitutes the technique of discrimination most used for problems of detection of moving objects such as pedestrians. This classifier is also used successfully in the tasks of detection and recognition of faces [1]. In our case the role of the SVM is to provide a decision about the object detected on the basis of the vector of discriminating characteristics of the shape obtained with the detector/descriptor based on the combination of Harris detector and HOG descriptor.

4.3 Performance Measurement

In order to evaluate our technique for detecting people in motion, we will use two indices to assess the quality of the results obtained. This is the detection rate and bad detection rate:

- **Detection rate:** This index allows calculating the percentage of good detections of persons in a sequence of images. The more his value is raised the more the proposed approach is effective. Its expression is as follows:

$$\text{Detection rate} = \frac{r_d}{n} \tag{2}$$

Where r_d is the number of persons detected and n is the total number of persons present in the image.

- **Bad detection rate:** This index allows calculating the percentage of the bad detections, i.e the part located in the image is not a person, but indeed it is. The lower his value is, the more effective the proposed approach is. Its expression is given by:

$$\text{Bad detection rate} = \frac{r_m}{n} \tag{3}$$

Where r_m is the number of bad detections.

So, the evaluation of the performance of our technique of detection of the persons in motion is performed on a set of video sequences extracted from the PETS database 2006/2007 Benchmark Data.

5 Implementation and Results

In order to evaluate our technique of detecting people in a video sequence while comparing it to that using the HOG descriptor, we have implemented algorithms to implement the two detection methods. These algorithms have been applied to a set of video sequences and images taken simultaneously.

The obtained results are presented in figures. 5 and 6 where we have marked the detection errors produced by the technique based on the HOG descriptor by arrows of different colors. The sky-blue arrows indicate the false positives (boxes not containing people), the orange ones indicate the false negatives (corresponding to the people who are not detected in the scene) and finally those of orange color corresponds to the ghost detections (which appear near other detections).

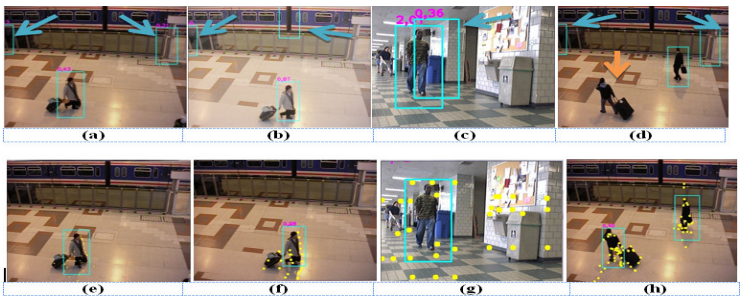

Figure 5 Simple scenes where the images (a), (b), (c) and (d) represent the results obtained by the HOG descriptor. Images (e), (f), (g) and (h) represent the results obtained with our detection technique.

The results obtained in FIG. 5 show that the use of the combination detector/descriptor has allowed the improvement of the quality of detection of persons in motion by reducing the rate of false detections produced by the use of HOG descriptor only while (a) (e), (b) (f), (c) (g) and (d) (h), for example, in the scene corresponding to (d) and (h), the images captured at the same moment show that our technique detects the two persons present in the scene, contrary to the one based on the HOG descriptor only, which detects just one person in addition to two boxes not containing people. So, the obtained results for the various scenes show that the application of the HOG descriptor alone can cause false detections. In fact, this descriptor can detect, in addition to people, forms that resemble a human.

In order to better evaluate the performance of our method in terms of detection of people, we also tested it on complex scenes. The results are presented in FIG. 6. These results show that the detection rate of our method is more important than that of the HOG descriptor which does not use the points of interest of Harris.

Consequently, the integration of a detector influences well the results obtained by the HOG descriptor since the combination of our detector / descriptor improves well the quality of detection. This improvement in the detection quality allows reducing the calculation time of the SMV classifier and consequently, the simplification of the task of tracking the persons in motion in a video scene.

Figure 6 Complex scenes where the images (a) and (b) represent the results obtained by the HOG descriptor. Images (c) and (d) represent the results obtained with our detection method.

5.1 Evaluation of the quality of the results

We have seen in section 4 (C) that the evaluation of the performance of our method of detection of persons can also be done using the indexes detection rate and bad detection rate. To do this, we computed these two indices by considering several sequences of test images. The test videos we have chosen present different scenes where the number of people varies from a scene to another. For example, for the Test_video_1 scene, the number of people present scene varies between 3 and 4, for the Test_video _2 sequence, this number varies between 4 and 7, and finally the Test_video_3 contains more than 9 people. The results are presented in the table below:

Table 1. Measurement of the detection rate of persons

	Detection of people by our method		Detection of persons by the HOG descriptor	
	Detection rate	Bad detection rate	Detection rate	Bad detection rate
Test Video 1	95% until 100%	2% until 0%	50% until 75%	33.33 %
Test Video 2	95% until 100%	5% until 2%	66,66% until 75%	33.33 %
Test Video 3	89% until 95%	6% until 2%	75% until 87.5%	14.29%

Copyright © Society for Science and Education United Kingdom

Thus, the results obtained for all the tests show that the performances of our technique of detection of people exploiting the integration of detector/descriptor are the best. Indeed, for the sequences of images containing a reduced number of people, the rate of detection can reach a 100 % with a low rate of bad detection. For the images containing a significant number of people, there is a decrease in the rate of detection and that of bad detection increase but always remains low.

We note that for the results obtained by the method of HOG show that the rate of bad detection increases with the complexity of the image (number of persons present, lighting, quality of the image ...).

6 Conclusion

In this paper, we proposed a new technique for detecting moving persons in a video scene. Our technique can be seen as an improvement of that based on the HOG descriptor. This improvement is obtained by integrating a detector / descriptor type process where instead of applying the HOG descriptor to the entire image, only the areas of interest detected by the Harris detector are used. The calculation of points of interest is a negligible step in computation time, which makes the proposed technique more suitable for real-time applications. The experimental results presented in this paper show the correct functioning of our detection algorithm. We can therefore think of extending the possibilities of our detection technique to facilitate the process of tracking people.

REFERENCES

[1]. R. Brown, J. *Ohmer and F. Maire, Implementation of Kernel Methods on the GPU.* Proceedings of the Digital Image Computing on Techniques and Applications*, 2005.*

[2]. N. Dalal and B. Triggs, *Histograms of oriented gradients for human detection.* In International Conference on Computer Vision & Pattern Recognition, vol. 2, pp. 886–893, 2005.

[3]. N. Dalal, B. Triggs and C. Schmid*, Human detection using oriented histograms of flow and apparence*. In European Conference on Computer Vision, pp. 7-13, 2006.

[4]. C. Harris and M. Stephens, *A combined corner and edge detector, in Alvey Vision Conference*, pp. 147–151, 1988.

[5]. B. Hdioud, A. Ezzahout, M. Y. Hadi and R. Oulad Haj Thami, *A Real-Time People Tracking System Based on Trajectory Estimation* at International Conference on Computer Applications Technology ICCAT'2013, pp. 20-22, Sousse, Tunisia, 2013.

[6]. 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.

[7]. P. Viola and M. Jones, *Rapid object detection using a boosted cascade of simple feature*, in IEEE Proc. of the conference on Computer Vision and Pattern Recognition, pp. 511–518, 2001.

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

[9]. L. Zhao, Dressed *Human Modeling, Detection, and Parts Localization*, PhD thesis, The Robotics Institute, Carnegie Mellon University, Pittsburgh, 2001.

[10]. Q. Zhao, J. Kang, H. Tao and W. Hua, *Part Based Human Tracking In A Multiple Cues Fusion Framework*, in Proc. of the International Conference on Pattern Recognition, pp. 450–455, 2006.

[11]. D. Gavrila and V. Philomin. *Real-time object detection for smart vehicles*. International Conference in Computer Vision, 1 :87–93, 1999.

[12]. I. Haritaoglu, D. Harwood, and L. S. Davis. W4: Who? when? where? what? *A real time system for detecting and tracking people.* In Third International Conference on Automatic Face and Gesture Recognition, pages 222–227, 1998.

[13]. PETS database; http://www.cvg.reading.ac.uk/PETS2006/data.html

[14]. PETS database; http://www.cvg.reading.ac.uk/PETS2007/data.html

Copyright © Society for Science and Education United Kingdom