Journal of Computers Vol. 31 No. 4, 2020, pp. 211-221
doi:10.3966/199115992020083104016

# Improved Pedestrian Detection Algorithm Based on HOG and SVM

Renwei Tu, Zhongjie Zhu*, Yongqiang Bai

Ningbo Key Lab of DSP, Zhejiang Wanli University, Ningbo, China
zhongjiezhu@yeah.net

Received 31 January 2019; Revised 13 June 2019; Accepted 14 July 2019

Abstract. Pedestrian detection becomes an acknowledged challenging problem to the development of intelligent video surveillance and vehicle active safety. At present, there are some shortcomings in pedestrian detection, such as missing detection, false detection, inaccurate detection frame location. An improved algorithm based on Histograms of Oriented Gradients (HOG) and Support Vector Machine (SVM) is proposed to tackle these problems. This algorithm employs face detection technology to further enhance the accuracy of pedestrian detection. Firstly, face detection and pedestrian detection are adopted to accurately locate the face contour and the human body contour, respectively. And then, detection frame is redraw by synthesizing the coordinates of the upper left corner of the face contour and the coordinates of the lower right corner of the human body contour. A new test database is established with numerous images in complex scenes. And then, relevant experimental results demonstrate the effectiveness of proposed algorithm and shows better target detection accuracy and real-time performance compared with existing methods.

Keywords: face, HOG feature, pedestrian detection, SVM classifier

## 1 Introduction

Pedestrian detection based on machine vision is one of the most popular research topics at present, which is widely used in the fields of intelligent transportation, intelligent monitoring, and pedestrian abnormal behavior detection. To achieve better detection rate and detection speed, the current hot deep learning algorithm has been introduced into the research field of pedestrian detection, but the research and application of deep learning requires a lot of hardware support, which results some technical difficulties in being ported to small mobile devices. Hence, it is still necessary to deeply analyze and improve the existing methods.

Most of the current traditional pedestrian detection methods use statistical learning based methods. Dalal and Triggs proposed a combination of HOG (Histograms of Oriented Gradients) and Support Vector Machine (SVM) on CVPR [1]. Although this method achieved a detection success rate of approximately 90% on the INRIA pedestrian database containing changes in perspective, illumination and background, it still had more missed detections and false detections in actual detection, accuracy and real-time detection. Therefore, the accuracy and real-time performance of the detection needs to be improved. Later, Wu et al. proposed the concept of “Edgelet” features [2], i.e., short straight lines and curve segments, and applied them to pedestrian detection of single images in complex scenes. The algorithm achieved a detection rate of approximately 92% on the CAVIAR database. However, it is time consuming and laborious as requiring manual calibration for each Edgelet feature. Meanwhile, it is difficult to obtain a “small edge” feature by manual calibration for some more complex curves. To address this problem, Sabzmeydani proposed a feature that can be automatically obtained using machine learning methods, namely Shapelet features [3]. The Shapelet feature is more in line with the human body curve and better describes the human body. Based on the Shapelet feature, 90% of the pedestrian

* Corresponding Author

detection success rate is achieved on the INRIA dataset, but the formation of its features lacks adaptability, and some discriminative details may be lost for some cases.

Recently with the improvement of computer computing power and deep learning theory, deep learning has developed rapidly in the field of computer vision, and its detection effect is significantly higher than traditional algorithms. In 2015, R-CNN [4] designed by Ross B. Girshick detected 66% of the results on the PASCAL VOC2007 test dataset. However, the training process of R-CNN is too complicated, and the information redundancy of the detection process is too large, so the detection speed is very slow. Generally, this method takes tens of seconds for ordinary image detection, which cannot meet the purpose of real-time detection. The proposed Faster R-CNN [5] has a significant improvement in detection speed and accuracy compared to the previous algorithm. In 2018, the YOLO v3 [6] proposal further solved the problem of detection accuracy and real-time performance, which made the single picture detection speed reach tens of milliseconds. However, deep learning algorithms require a large number of data sets, and it is possible to obtain good detection results by performing high-speed GPUs for tens of thousands of iterations for a long time.

To tackle these problems in pedestrian detection, such as missing detection, false detection, inaccurate detection frame location, this paper proposes a HOG+SVM pedestrian detection algorithm combined with face detection. The contour of the face is accurately positioned by Haar+Adaboost [7], and the contour of the human body is positioned according to HOG+SVM. The contour coordinates of the face and human body are integrated to accurately segment the position of the pedestrian in the figure, and the multi-feature fusion detection pedestrian is realized. In this paper, a new test database is established with numerous images in complex scenes. And then, relevant experimental results demonstrate the effectiveness of proposed algorithm, on the experimental platform of Visual Studio 2013 and open CV 3.0. The fusion of multiple features effectively eliminates the influence of pedestrian headwear and hats on HOG feature extraction, improves the accuracy of human contour positioning, and solves the problem of false detection of HOG pedestrian detection to some extent. Compared with existing method based on HOG+SVM, the proposed algorithm reduces the false detection rate, and detection rate by 5.1% and 11.1% respectively, and increases the detection accuracy by 8.1%, showing better target detection accuracy and real-time performance.

## 2 Feature Extraction and Classifier

The core issues in pedestrian detection are feature extraction and classification which largely determine the accuracy of pedestrian detection. In 2005, Dalal and Triggs proposed the concept of HOG and used it for pedestrian detection, achieving good detection results. The directional gradient histogram feature is characterized by calculating and statistic the gradient direction histogram of the local region of the image. Extracting the HOG feature requires calculating the gradient of each pixel in the image, including size and direction.

The gradient of the pixel point (x, y) in the image is [8]:

$$G_x(x,y) = H(x+1,y) - H(x-1,y), \tag{1}$$

$$G_y(x,y) = H(x,y+1) - H(x,y-1). \tag{2}$$

Where $G_x(x,y)$ is the gradient of the input image in the horizontal direction, $G_y(x,y)$ is the gradient of the input image in the vertical direction, and $H(x,y)$ represents the pixel value of the input image.

The following two formulas are to find the gradient magnitude and gradient direction of the pixel at (x, y):

$$G(x,y) = \sqrt{G_x(x,y)^2 + G_y(x,y)^2}, \tag{3}$$

$$\alpha(x,y) = \tan^{-1}\left(\frac{G_y(x,y)}{G_x(x,y)}\right). \tag{4}$$

As shown in Fig. 1, the feature extraction needs to first divide the cell, the block and the gradient direction, and divide the sample image into a plurality of cells, each block consisting of four adjacent cells, and the gradient direction is divided equally. 9 intervals, in each unit, the gradient direction of all pixels is subjected to histogram statistics in each direction interval to obtain the feature vector. After the 4*9-dimensional feature vectors in the block are connected in series, the block is scanned in one unit by the scanning step size. The sample image is scanned, and finally all the block feature vectors are concatenated to obtain a human body feature expressed in a vector form.

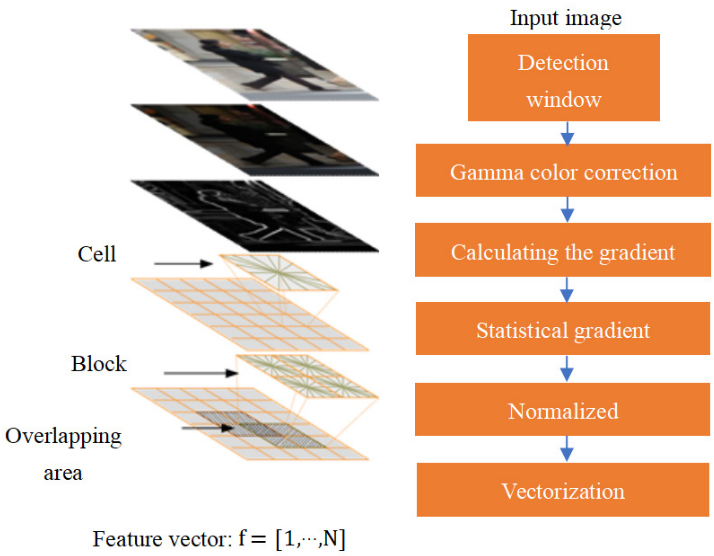



**Fig. 1.** HOG feature acquisition flow chart

The training of the SVM classifier is actually to find the optimal segmentation hyperplane. The mathematical meaning is based on the premise of dividing the positive and negative samples, and the maximum geometric distance from the support vector to the segmentation hyperplane is obtained. Let the segmentation hyperplane be: g(x)=ω·x+b, then the geometric spacing of the support vector to the segmentation hyperplane is:

$$\delta_i = \frac{y_i\left(\omega \cdot x_i + b\right)}{\|\omega\|}. \quad \textbf{(5)}$$

Where $y_i$ represents the category to which the sample belongs. In order to get the maximum δ, you need to minimize $\|\omega\|$. Then the training of the SVM classifier can be transformed into a mathematical problem: find the solution of ω in the inequality $y_i\left(\omega \cdot x_i + b\right) - 1 \geq 0$, and then find the smallest $1/2\|\omega\|$. Finally, the extracted HOG feature vector is brought into the SVM classifier to calculate the confidence, as shown in Fig. 2.

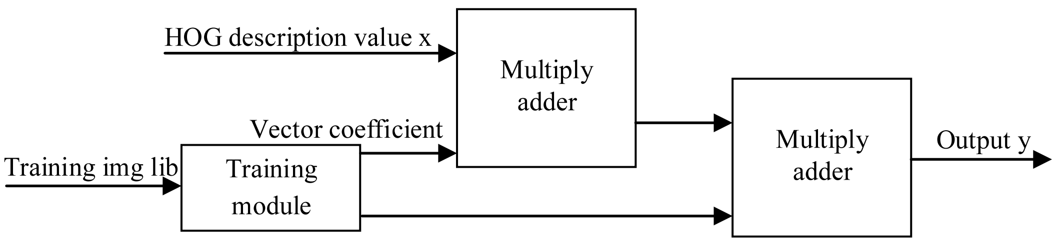



**Fig. 2.** SVM classifier calculation process

## 3 Pedestrian Detection Based on HOG and SVM Combined with Face Analysis

The overall design scheme of pedestrian detection based on HOG and SVM combined with face analysis is shown in Fig.3. Firstly, the face detection is realized by using the Haar feature + Adaboost classifier, and then the pedestrian contour pre-detection is performed on the image according to the HOG feature + SVM classifier. Finally, detection frame is redraw by synthesizing the coordinates of the upper left corner of the face contour and the coordinates of the lower right corner of the human body contour.

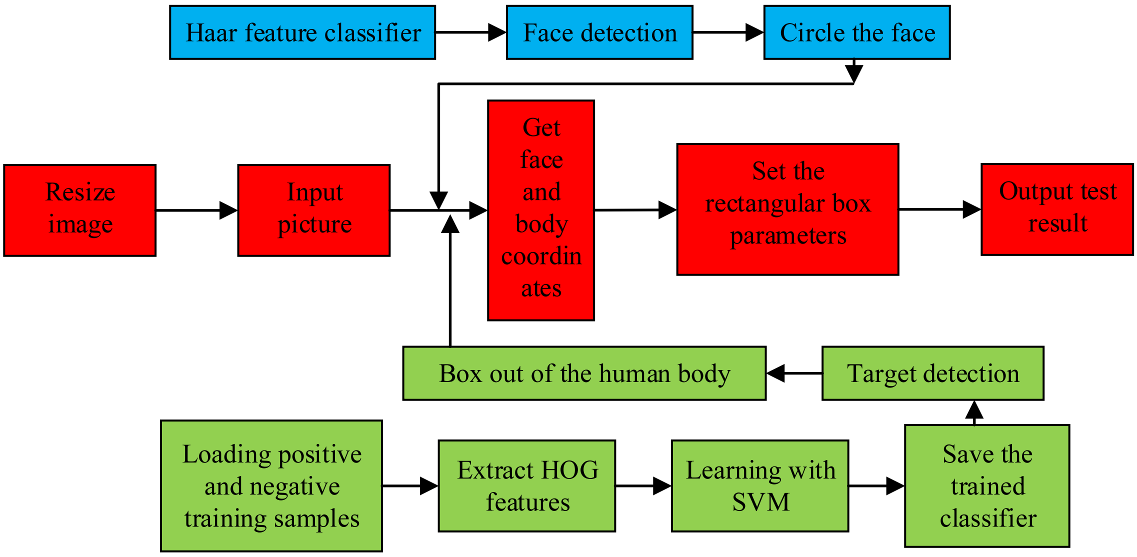



**Fig. 3.** Flow chart of the proposed pedestrian detection scheme

The inspection process is divided into three parts:

(1) Face detection determines the position of the face. Face detection is performed using the trained haar feature + Adaboost classifier in the OpenCV environment. The detected face area is a rectangular area, and the detection result is circled by a blue circle, which is the first step of detecting.

(2) Pedestrian pre-detection determines the initial contour of the human body. By extracting the HOG features of a large number of positive and negative samples, the SVM classifier is trained to obtain a discriminant function, which can directly identify the HOG features. In general, the larger the sample set of training, the more accurate the parameters of the discriminant function, and the higher the accuracy of the detection. The test results are framed by a green rectangle.

(3) Combine face detection with pedestrians to complete accurate pedestrian detection. According to the coordinates of the face and the human body, the position of the pedestrian in the figure is comprehensively determined. Since the width of the face region detected by the face detection is generally smaller than the width of the human body region detected by the HOG pedestrian detection, it is necessary to fine-tune the parameter of the upper left corner coordinate of the face region as the origin of the final detection of the human body rectangle. Normally, the origin moves to the upper left corner, and through continuous debugging procedures, the appropriate points are found based on the feedback of the test results. The detection results of the HOG pedestrian detection combined with the face are marked with a red rectangular frame.

### 3.1 Face Detection

Face detection utilizes the Haar feature and the Adaboost classifier [9] in the OpenCV environment. Haar features consist of three categories: linear features, edge features, center features, and diagonal features. These three categories constitute their feature templates. The Haar eigenvalue reflects the grayscale variation of the image. The characteristics of the face can be described by the rectangular features in the Haar feature template. The mouth is darker than the skin around it, the eyes are darker than the eyelids, and the nose is darker than the nose. The Haar feature template has two rectangles in black and white, a white rectangle pixel and a black rectangle pixel minus and is the feature value of the template. By comparing the feature value with the pre-calculated threshold, the classification can be completed to determine whether it is a human face. But this Haar feature is just a weak classifier, and its detection rate is low, only a little better than random guessing. Only by combining a large group of such features can a credible judgment be reached, and these Haar features are combined to form an Adaboost cascade classifier [10]. This is a filtered cascading classifier. Each node is a classifier composed of multiple trees. If the algorithm stops immediately in any level of calculation, it means that no face is detected at this level. Only through all the levels in the classifier will it be considered that a human face is detected.

Fig.4. is a schematic diagram of the Adaboost cascade classifier. When identifying, different areas of the original image will be scanned. 70%-80% of the non-face areas will be rejected in the first two nodes of the screening cascade, so the detection speed quickly. Obviously, this screening cascade makes the correct recognition of face detection high.

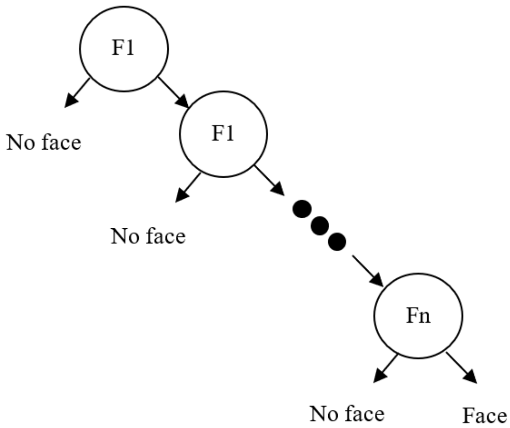



**Fig. 4.** Adaboost cascade classifier

Through programming and verifying with a large number of images, the algorithm does have a good face detection performance, whether it is one or two faces or complex multi-faces can be accurately detected. Fig. 5 shows the results of a multi-face image. Eight faces were successfully detected and no false positives occurred.

**Fig. 5.** Face detection image

### 3.2 Pedestrian Pre-detection

The HOG+SVM pedestrian detection algorithm is used as the pedestrian pre-detection of the proposed algorithm. The 64*128 pixel value picture is selected as the training sample. Therefore, the extraction feature can use the whole image as the feature extraction object. Set the detection window size to 64*128, the block size to 16*16, the sliding step size of the block to 8, the cell size to 8*8, and the cell gradient direction to 9.

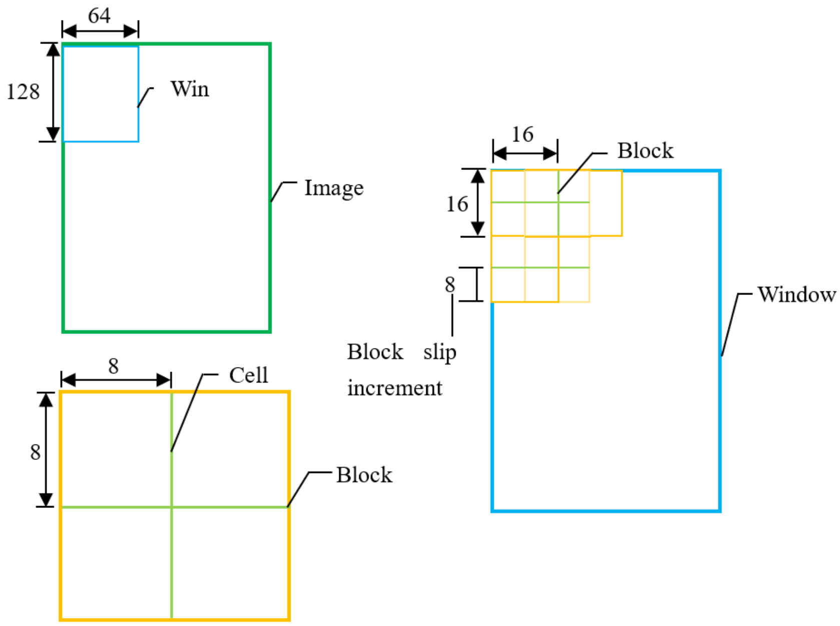



**Fig. 6.** Detection window, block, cell structure diagram

The training sample used the public MIT database 924 pedestrian image as a positive sample, using 943 online download images as a negative sample, after SVM training, it can be tested.

The most critical function used for detection is:

HOG.detectMultiScale (img, found, 0, Size (8, 8), Size (32, 32), 1.05, 2);

This is multi-scale detection. found is the list of target areas for detection,0 is the distance from the detected feature to the SVM classification hyperplane. Size (8, 8) is the distance that the sliding window moves each time, and Size (32, 32) is the image expansion size.1.05 is the ratio of the compressed image to be compressed each time (the detection window is fixed, the image to be detected is scaled down), and 2 is the correction coefficient.

When detecting, the area with internal and external inclusion relationships in the space is removed, and the large one is retained. Finally, the detected target rectangle should be processed in some way. For example, if the two target frames are nested, the outermost frame is selected. Because the rectangular frame at the HOG detection is slightly larger than the actual human frame, make some adjustments to the size of the rectangular frame. Fig. 7 is a result of detection of pedestrian pre-detection, and the green frame can substantially frame the contour of the human body.

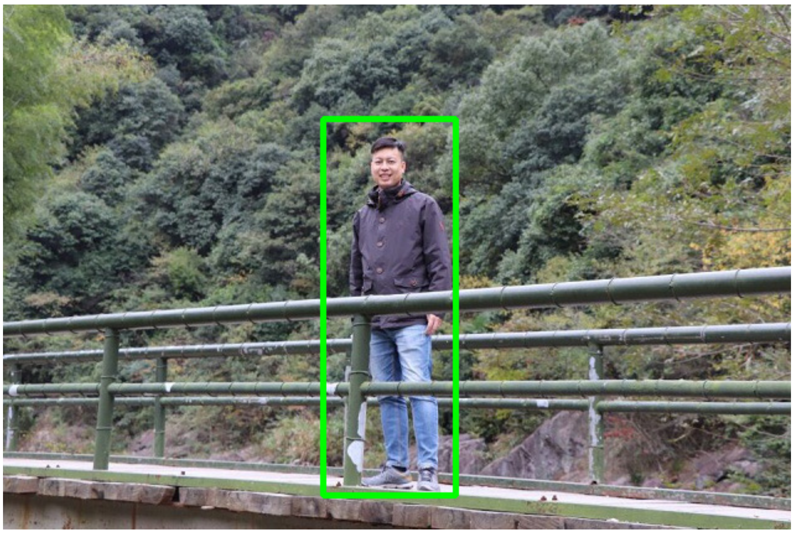


**Fig. 7.** Pedestrian pre-detection result

### 3.3 Pedestrian Detection Combined with Face Analysis

By detecting the face, the coordinates of the face in the image are determined, and the HOG+SVM pedestrian detection algorithm detects the contour of the person as a pedestrian pre-detection. In this paper, the coordinates of the upper left corner of the contour of the face and the coordinates of the lower right corner of the human contour are re-established. A rectangular box is drawn to determine the basis of the position of the pedestrian in the image.

HOG+SVM pedestrian detection algorithm sometimes has misdetection, especially when the part of the image that is not the human body is misdetected as the human body, or the human body part is identified, or the detection frame does not completely frame the whole person. Even some people have completely missed the detection. For these cases, a face detection algorithm with very high detection accuracy is used in this paper. Firstly, it is ensured that the face in the detected image can be detected, and then the coordinates of the head of all people in the image are known. Secondly, the coordinates of the human body are determined according to the contour of the human body detected by the HOG pedestrian pre-detection. Among the 8 coordinate points, the point at the upper left corner is selected as the origin (ie, the coordinates of the upper left corner of the overall contour), and the point at the lower right corner is used as the coordinates of the lower right corner of the overall contour. The coordinates of the upper left and lower right corners are determined to frame the outline.

There are three pictures in Fig. 8, each picture has a blue circle, two red dots, and a red rectangular frame, which are the face detection frame, the key coordinate points for determining the human body area, and the final detection frame. The green detection frame is the HOG pedestrian detection frame. The left picture represents that the HOG pedestrian detection frame does not completely frame the human body, and the head part is not in the frame. The middle picture represents that the HOG pedestrian detection only detects the area below the human head and does not detect the head at all. The two detection effects are obviously not ideal and need to be improved. The picture on the right represents that HOG pedestrian detection has not successfully detected the human body area and draws a rectangular frame. There are many reasons for this phenomenon. It is possible that the illumination distribution in the picture is very complicated, and there may be occlusion of the human body part of the detected picture, etc., which will affect the detection performance. In the fourth case, the HOG pedestrian detection is completely misdetected, and the area that is not the human body is incorrectly framed. And this happens more frequently. There is also a situation where HOG pedestrian detection needs to be improved. For images of multiple people, there will be more obvious missed detections. It is possible that HOG pedestrian detection can only detect two people though there are four or five people in the image.

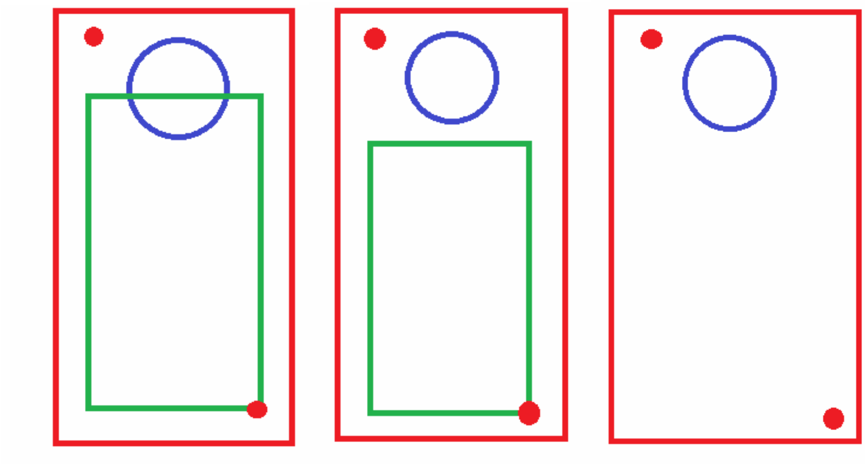


**Fig. 8.** Detection window, block, cell structure diagram

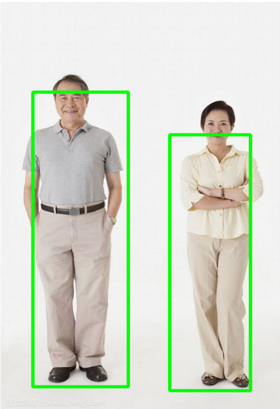
For some HOG pedestrian detections that do not have a frame drawn, it is not always because the human body area is not detected at all, so it is sometimes possible to provide the lower right corner coordinates. Once face detection and HOG pedestrian detection provide key coordinate points for the body region, pedestrian detection involving the entire region of the head and body can be achieved. Especially for the shortcomings of misdetection of HOG pedestrian detection, the pedestrian detection algorithm combined with the face effectively solves this problem. If there is no positive face detected in the image but a green frame appears, it must be a false check.

## 4 Experimental Results and Analysis

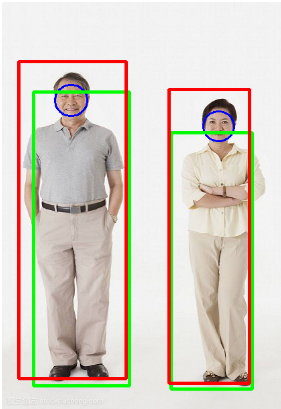
In this paper, 500 sets of positive face pedestrian data sets are collected from the network for verification of the algorithm. The data set contains various images of different illuminations, different backgrounds and different numbers of people to verify the reliability of the algorithm.

**Fig. 9.** Examples of the self-built detection data set

The experimental results are as follows: The HOG pedestrian detection combined with the face has indeed made significant progress in the place where the original HOG pedestrian detection is still defective. The problem that the original algorithm can’t picture frame accurately in some images is successfully solved. Also, it solved the problem of missed detection in some cases. The following six groups representative images are selected for comparing and analyzing the result of detection of HOG pedestrian detection algorithm and the HOG pedestrian detection algorithm combined with the face.



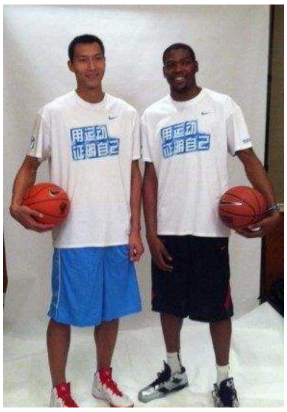
(a) HOG detection



(b) HOG + face detection

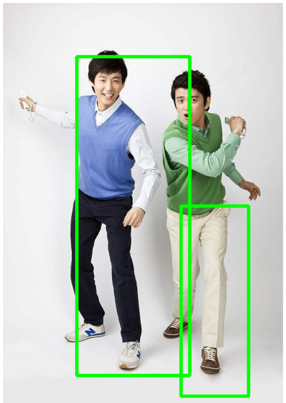
(c) HOG detection

(d) HOG + face detection

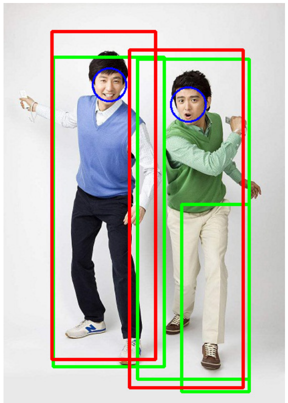


(e) HOG detection

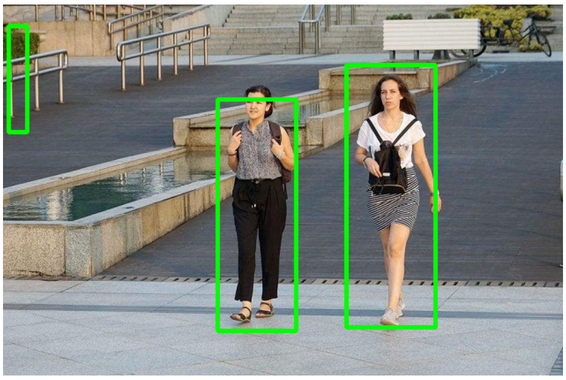
(f) HOG + face detection



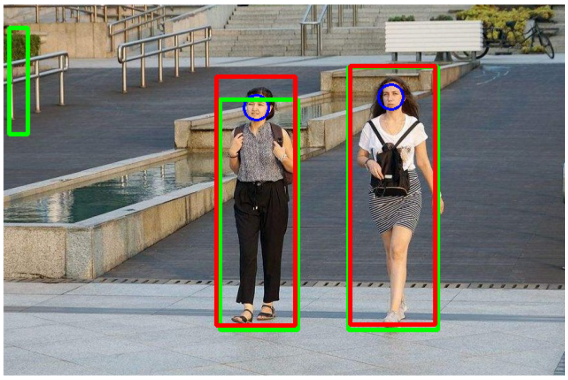
(g) HOG detection



(h) HOG + face detection

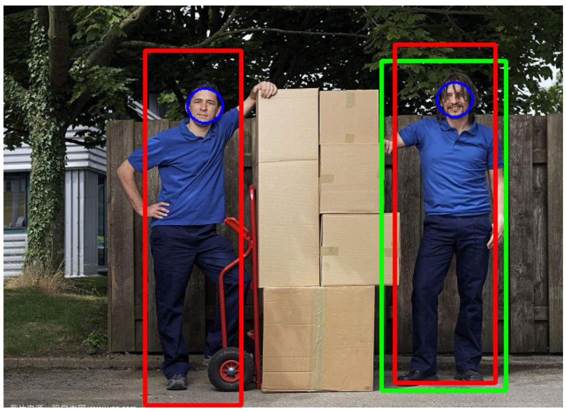


(i) HOG detection



(j) HOG + face detection

(k) HOG detection



(l) HOG + face detection

**Fig. 10.** Comparison results of different detection algorithms

The comparison of the above six groups of pictures shows that the HOG pedestrian detection combined with the face has a significant improvement in the detection accuracy compared to the HOG pedestrian detection for the above-mentioned several cases. The comparison between (a) and (b) shows that the pedestrian detection combined with the face is more accurate for the positioning of the contour of the human body, and the human body can be completely contained within the detection frame. Comparing (c) and (e), the HOG pedestrian test did not detect the person in the picture at all. (d) and (f) show that the HOG pedestrian detection combined with the face accurately detects the face and the human body area. A comparison between (g) and (h) shows that the HOG pedestrian detection does not completely identify the human body area in some pictures. It only detects the human torso part or only the upper body of the person. The HOG pedestrian detection combined with the human face will simultaneously frame two incomplete HOG pedestrian detection frames on the left side, and the detection accuracy is obviously improved. In (i), the HOG pedestrian detection successfully detected the person in the image, but there was a misdetection in the upper left corner of the figure, which mistakenly recognized the railing as a person. In (j) the figure shows that the HOG pedestrian detection combined with the face successfully solves the problem of false detection of the figure. If no face is detected, the algorithm does not recognize the frame. There are 2 people in (k), but the HOG pedestrian detection only identifies the person on the right, the person on the left is missed the detection.

**Table 1.** Comparison results of different detection algorithms

| Algorithm | Detected target | False detections | Missed detections | Semi-accurate number | Accurate number |
|---|---|---|---|---|---|
| HOG+SVM | 345 | 65 | 155 | 95 | 175 |
| HOG+SVM combined with face | 365 | 50 | 85 | 110 | 205 |

Table 1. shows the results of the HOG+SVM pedestrian detection algorithm and the improved algorithm for the detection of 500 data sets. From the number of detected targets, the improved algorithm is 20 more than the original algorithm, the number of false detections is less than 15, and the number of missed detections is 70, and the accuracy of recognition is higher. The semi-accuracy in the table covers the part of the body that detects the frame, or two pedestrians in a frame, which accurately refers to the complete frame of pedestrians.

**Table 2.** Comprehensive comparisons of different detection algorithms

| Algorithm | False detection rate | Missed detection rate | Semi-accuracy rate | Accuracy | Total accuracy |
|---|---|---|---|---|---|
| HOG+SVM | 18.8% | 31% | 27.5% | 50.7% | 78.2% |
| HOG+SVM combined with face | 13.7% | 18.9% | 30.1% | 56.2% | 86.3% |

Table 2. shows that the improved algorithm has a 5.1% lower false detection rate than the original algorithm, the missed detection rate has decreased by 12.1%, and the coarse accuracy rate has increased by 8.1%. The overall detection effect has been significantly improved compared with the original algorithm.

## 5 Conclusion

This paper addresses one problem for constructing an effective frontal pedestrian detection. Haar feature + Adaboost classifier and HOG feature + SVM classifier are combined to propose an improved pedestrian detection algorithm based on HOG and SVM. With the face detection and pedestrian detection, the face and the position of the human body in the images are comprehensively combined to detect the pedestrian. It makes the accuracy of the detection is significantly improved when detecting the front of the human body than the HOG pedestrian detection, especially in the case of the missing or false detection of the HOG, to some extent, the deficiencies of the detection are compensated. The algorithm only combines face detection in the detection, and cannot avoid false detection or missed detection of pedestrians. In future research, other features can be considered to improve the accuracy of pedestrian detection on the side, back and front. It is believed that with the continuous improvement of algorithms, pedestrian detection technology will be more and more perfect, and will be applied in more fields.

## Acknowledgements

This work was supported in part by the National Natural Science Foundations of China (No. 61671412); Zhejiang Provincial Natural Science Foundation of China (No. LY19F010002); the Natural Science Foundation of Ningbo, China (2018A610053); Ningbo Municipal Project of Science and Technology beneficial to People (No. 2017C50011); Innovation and consulting project from Ninghai Power Supply Company, State Grid Corporation of Zhejiang, China.

## References

[1] N. Dalal, B. Triggs. Histograms of oriented gradients for human detection, in: Proc. IEEE Computer Society Conference on Computer Vision & Pattern Recognition, 2005.

[2] A. Loesch, S. Bourgeois, V. Gay-Bellile, M. Dhome, Generic edgelet-based tracking of 3D objects in real-time, in: Proc. 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2015.

[3] W. Yao, Z. Deng, A robust pedestrian detection approach based on shapelet feature and Haar detector ensembles, Tsinghua Science and Technology 17(1)(2012)40-50.

[4] C. Chen, M.-Y. Liu, C. Tuzel, C.O. Oncel, J. Xiao, R-CNN for small object detection, in: Proc. 13th Asian Conference on Computer Vision, 2016.

[5] H. Jiang, E. Learned-Miller, Face detection with the faster R-CNN, in: Proc. 12th IEEE International Conference on Automatic Face & Gesture Recognition, 2017.

[6] J. Redmon, A. Farhadi, YOLOv3: An incremental improvement. <https://arxiv.org/abs/1804.02767>, 2018.

[7] Y. Zhou, D. Liu, T. Huang, Survey of face detection on low-quality images, in: Proc. 13th IEEE International Conference on Automatic Face & Gesture Recognition, 2018.

[8] V. Hoang, M. Le, K. Jo, Hybrid cascade boosting machine using variant scale blocks based HOG features for pedestrian detection, Neurocomputing 135(2014) 357-366.

[9] J. Zhu, Z. Chen, Real time face detection system using adaboost and Haar-like features, in: Proc. 2nd International Conference on Information Science and Control Engineering, 2015.

[10] W. Yu, J. Xiu, C. Liu, Z. Yang, A depth cascade face detection algorithm based on adaboost, in: Proc. IEEE International Conference on Network Infrastructure and Digital Content, 2016.