The algorithm of nighttime pedestrian detection in intelligent surveillance for renewable energy power stations

Bao Peng¹, Zhi-Bin Chen², Erkang Fu³ and Zi-Chuan Yi⁴

Abstract
Intelligent surveillance is an important management method for the construction and operation of power stations such as wind power and solar power. The identification and detection of equipment, facilities, personnel, and behaviors of personnel are the key technology for the ubiquitous electricity The Internet of Things. This paper proposes a video solution based on support vector machine and histogram of oriented gradient (HOG) methods for pedestrian safety problems that are common in night driving. First, a series of image preprocessing methods are used to optimize night images and detect lane lines. Second, an image is divided into intelligent regions to be adapted to different road environments. Finally, the HOG and support vector machine methods are used to optimize the pedestrian image on a Linux system, which reduces the number of false alarms in pedestrian detection and the workload of the pedestrian detection algorithm. The test results show that the system can successfully detect pedestrians at night. With image preprocessing optimization, the correct rate of nighttime pedestrian detection can be significantly improved, and the correct rate of detection can reach 92.4%. After the division area is optimized, the number of false alarms decreases significantly, and the average frame rate of the optimized video reaches 28 frames per second.

¹Shenzhen Institute of Information Technology, Shenzhen, China
²South China Academy of Advanced Optoelectronics, South China Normal University, Guangzhou, China
³Sichuan Agricultural University, Chengdu, China
⁴Zhongshan Institute, University of Electronic Science and Technology of China, Zhongshan, China

Corresponding author:
Erkang Fu, College of Landscape Architecture, Sichuan Agricultural University, Chengdu 611130, China.
Email: 368066539@qq.com

Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (https://creativecommons.org/licenses/by/4.0/) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
**Keywords**
Intelligent surveillance, nighttime pedestrian detection, support vector machine, Internet of Things

**Introduction**
The development of renewable energy to reduce reliance on traditional thermal, coal, or nuclear power generation methods has been the goal of governments in recent years (Johansson, 1975). China as the world’s most important country for wind power and solar power, had the world’s largest wind (169 GW) and solar (77.7 GW) capacities along with highly competitive wind and solar photovoltaic (PV) manufacturing. Wind power and solar power facilities are widely distributed in deserted or sparsely populated areas, where external communication conditions are poor, people, materials, and machines are dynamically concentrated, lead to a lot of safety risks at the construction and operation. It affects the normal operation of the power plant and causes serious losses to the enterprise (Aki et al., 2016; Schmarzo, 2015; Zeid and Davis, 2014). Many experiences show: regulatory difficulties are one of the key factors causing problems in the management of wind and solar companies. Intelligent monitoring system is the main technical means of the current power plant supervision. It automatically provides intelligent warning through the built-in computer vision algorithm, reduces personnel participation, and greatly improves the efficiency of monitoring system (Chen et al., 2017, 2018; Choudhury et al., 2017). This paper takes the typical application demand of nocturnal pedestrian detection as the research object, which is of high economic value and practical significance (Han et al., 2019; Pease, 2015).

Deep learning technology is an effective method to improve the performance of pedestrian detection algorithms (Liu et al., 2018). This kind of method is based on the statistical rules of a large number of pedestrian and non-pedestrian samples in the appearance mode and mainly includes two technical aspects: feature extraction and a classifier design. Currently, pedestrian feature descriptors applied to pedestrian detection include Haar-like feature applied, local binary pattern (LBP) feature applied, edgelet feature applied (Luo et al., 2019), scale invariant feature transform applied (Chen et al., 2018), histograms of oriented gradients (HOG) features applied, and multi-feature based on Hough forest applied (Peng et al., 2014). The classification algorithms commonly used in the field of pedestrian detection mainly include: artificial neural networks, various boosting methods, and support vector machine (SVM) (Su et al., 2012). The disadvantages of this method are as follows: deep learning requires a large number of data samples for training to obtain the pedestrian classifier model and maintain a sufficiently high accuracy. At the same time, because a large number of data need to be processed by a hardware system, the performance of the hardware system is very important. To improve the performance of the hardware system, most pedestrian detection systems run on a PC platform, but their volume and power consumption are relatively large. PCs are not suitable for specific application scenarios, such as a vehicle-mounted scene (Kim and Kim, 2018; Liem and Gavrila, 2014; Na et al., 2018).

To meet the demand for intelligent pedestrian detection at night, this paper proposes a vehicle-mounted nighttime pedestrian detection algorithm based on SVM and HOG. In the vehicle illumination recognition system, a camera is used to capture the image data of the road ahead, and the RGB image is converted into a hue saturation lightness (HSL) image. The image edge and erosion center are obtained by introducing an image processing
mathematical morphology method. All irregular reflectors are identified according to the symmetrical shape characteristics of the vehicle lighting. By dividing the image into upper and lower parts, we can identify the position of the headlights in the lower part of the road area in the image. In addition, the upper and lower boundary lines can be dynamically adjusted according to the farthest position of the lane line to adapt to actual road conditions, such as ramps and downhill roads. Then, the lighting direction of the matrix headlamp can be controlled according to the lighting environment. In the proposed system, liquid crystal display and wireless Bluetooth are configured, and Bluetooth communication technology and a mobile client interface are used to achieve image adjustments. The recognition calibration line added to the image display is used to solve the shooting angle offset problem and optimize the recognition effect. In short, the system can reduce the number of recognition errors and effectively improve the correct rate (Qazi et al., 2011).

System design principle

The problem of pedestrian detection at night includes the process of feature extraction in a broad sense. In a narrow sense, it includes matching methods, which are methods of determining whether the images match or not. Traditional matching methods are mostly based on simple distance operations or correlation operations, which have the advantages of good generality and high computational efficiency. However, in this particular style of image, it is difficult to achieve high precision. This paper mainly discusses the problem of selecting a measurement strategy and proposes a transformation of the problem of detecting nighttime pedestrians into a classification problem to overcome the shortcomings of traditional matching methods (Liu and Zhang, 2019; Su et al., 2012).

Image preprocessing

Before processing, feature extraction and classification of video images, chroma transformation, pixel edge acquisition, image segmentation and other image preprocessing should be carried out to optimize the acquired nighttime images.

Chromaticity conversion. The initial image acquired by the system may affect all three parameters of the image RGB values of the image due to different lighting conditions (Pan et al., 2015), but only the H parameter is affected in the image of HSL value of the image. To improve the image stability of the acquisition, in the image preprocessing steps, the RGB values of the image will be converted into HSL values. RGB values are converted into HSL values by the following

\[
h = \begin{cases} 
60^\circ \times \frac{g - b}{\max - \min} + 360^\circ, & \max = r, g < b \\
60^\circ \times \frac{g - b}{\max - \min} + 0^\circ, & \max = r, g \geq b \\
0^\circ, & \max = \min \\
60^\circ \times \frac{r - g}{\max - \min} + 240^\circ, & \max = b \\
60^\circ \times \frac{b - r}{\max - \min} + 120^\circ, & \max = g 
\end{cases}
\] (1)
In formulas (1) to (3), \( h \) represents hue, \( l \) represents lightness, and \( s \) represents saturation. Max and min represent the maximum and minimum RGB values, respectively. After converting the image, we recorded the pixel coordinates of the pixels with maximum brightness according to formula (3) and set it as the central target point \((X, Y)\) to facilitate the mask operation of the image.

**Pixel edge acquisition.** After the chroma conversion of the image was completed, the brightness of the pedestrians and the background were significantly different. The areas represented by pedestrians on the image are white with high brightness, while the background brightness is black. The purpose of this paper is to determine the position of the pedestrian area (the highlighted area) on the image. To determine the upper boundary, lower boundary, left boundary, and right boundary of the highlighted area of the image, the image is masked to smooth the image. We use the central target point \((X, Y)\) as the structural element to corrode the image in formula (2), create a \(3 \times 3\) convolution core as the mask, and use the HSL threshold to create a mask. The original image and mask perform bit-by-bit operations starting from \((X, Y)\). For the source image data \(F(x, y)\), the convolution operator is \(3 \times 3\), and the average mask in the neighborhood of 4 is calculated by the following formula (4)

\[
F(x, y) = \frac{Z(x-1,y) + Z(x+1,y) + Z(x,y-1) + Z(x,y+1)}{4}
\]  

(4)

In formula (4), \( x \) and \( y \) represent the horizontal and vertical coordinates of the image, respectively. By the above method, the center coordinates and boundary coordinates of the highlighted area on the image and the distance between the center and the boundary can be determined preliminarily. To further distinguish pedestrians from other highlighted objects, the pedestrian shape is analyzed in this paper. Ordinarily, pedestrians stand upright. A cuboid was used as the model to set up the basic conditions, and the length–width ratio of the cuboid was set in the range of 2:1–5:1

\[
5 \min(n_1 + n_2, n_3 + n_4) > \max(n_1 + n_2, n_3 + n_4) > 2 \min(n_1 + n_2, n_3 + n_4)
\]  

(5)

\( \min \) means the minimum value of \( n_1 + n_2 \) and \( n_3 + n_4 \) and \( \max \) means the maximum value of \( n_1 + n_2 \) and \( n_3 + n_4 \).

**Image segmentation.** The image area is roughly divided into upper and lower sections to avoid the interference from the sky, the street lights, and trees. However, in actual road conditions,

\[
s = \begin{cases} 
\frac{\max - \min}{\max + \min} \frac{\max - \min}{2l}, & 0 < l \leq \frac{1}{2} \\ 
0, & l = 0/\max = \min \\ 
\frac{\max - \min}{2 - (\max + \min)} \frac{\max - \min}{2 - 2l}, & l > \frac{1}{2} 
\end{cases}
\]  

(2)

\[
l = \frac{1}{2} (\max + \min)
\]  

(3)
such as uphill and downhill roads, the camera cannot make upward and downward adjustments in real time, so the image division area is not accurate, and the correct recognition area and nonrecognition area cannot be divided.

This system proposes a method of optimizing the effective region division. When it is necessary to divide the image into sky and road regions, the boundary line is dynamically adjusted according to the actual situation of the road. In addition, when there is no lane reference line on the road, Line1 is set to be a boundary line at 1/2 of the image width by default. By default, the distance from Line1 to Line2 is set to 1/10 of the image width (the height of the pedestrian in the image corresponding to the actual distance), and Line2 is aligned with the actual boundary line when the camera is actually tested and installed.

In Figure 1, the content of effective area division in the case of an uphill road is introduced. As shown in Figure 1(a), the original figure shows that there are tag lines on the road, and the furthest point from the lane is used as the reference point of the horizontal dividing line, Line2, in the system. The reference points for Line2 are shown in Figure 1(b), which is the brightened image after the chroma conversion. The relative distance between Line1 and Line2 is the approximate height of the pedestrian, as shown in Figure 1(c) and (d). The relative distances between Line1 and Line2 are fixed and the lines are dynamically adjusted upward and downward according to the actual boundary between the road and sky.

Figure 1. Regional division diagram of uphill conditions: (a) original image, (b) gray scale, (c) boundary line image, and (d) resulting image.
Then, the image is processed by the normalized segmentation method. The image is represented by an undirected weighted graph, and the pixels in the image are regarded as nodes in the graph. The segmentation criterion based on graph theory is used to perform image segmentation. The graph $G = (V, E)$ consists of some nodes $V$ and edge $E$, and each edge corresponds to the weight $W_{ij}$ of the two nodes it connects. If each pixel is regarded as a node, then $W_{ij}$ is the similarity between two pixels

$$W_{ij} = e^{-\frac{||x(i) - x(j)||_2^2}{2\sigma^2}} \cdot \begin{cases} e^{-\frac{||x(i) - x(j)||_2^2}{2\epsilon}} & , \|X(i) - X(j)\|_2 < r \\ 0, & \text{others} \end{cases}$$  

(6)

The degree of dissimilarity between regions $A$ and $B$ is defined as

$$Cut(A,B) = \sum_{u \in A, v \in B} w(u, v)$$  

(7)

Figure 2. Regional division diagram of downhill conditions: (a) original image and (b) resulting image.

Figure 3. Picture of the experimental setup.
The degree of correlation between parts $A$ and $V$ is defined as

$$Assoc(A, V) = \sum_{u \in A, t \in V} w(u, t)$$  \hspace{1cm} (8)

Then, image segmentation can be defined as

$$NCut(A, B) = \frac{Cut(A, B)}{Assoc(A, V)} + \frac{Cut(A, B)}{Assoc(B, V)}$$  \hspace{1cm} (9)

To find the optimal segmentation outcome, the minimum value of $NCut$ must be determined. It is very difficult to solve the above problems directly, and the amount of computation required to solve the problems is very large. Therefore, it is necessary to transform the problems into Rayleigh quotient minimization problems

$$\min_x NCut(x) = \min_y \frac{y^T (D - W) y}{y^T D y}$$  \hspace{1cm} (10)

$\min_y$ corresponds to the minimum value of $y$ and $NCut(x)_{\min}$ corresponds to the minimum value of $NCut$. The solution can be further converted to the generalized eigenvalue problem after the constraint conditions are relaxed

$$(D - W)y = \lambda Dy$$  \hspace{1cm} (11)

The solution can divide the graph according to the eigenvector corresponding to the second smallest eigenvalue, and the result is the optimal $NCut$ value, so the result image of the two-segmentation can be obtained. After obtaining the result of two-segmentation, the sub-region is divided into two parts by the recursive method, and finally, the result of multi-segmentation is obtained.

**HOG feature extraction**

To accurately distinguish pedestrians and other non-pedestrians in highlighted areas, this paper uses the HOG feature and machine learning method (Aleena and Anamik, 2016). The HOG descriptor is a feature descriptor used for object detection in computer vision and image processing. It calculates a statistical histogram of the local area of the image to form characteristics compared with other methods that describe the characteristics. HOG has many advantages. First, since HOG operates on the local grid unit of an image, it can maintain a favorable level of invariance to the geometric and optical deformations of the image. These two kinds of deformations can only occur in a larger space. Second, with coarse spatial sampling, fine direction sampling, and strong local optical normalization, as long as pedestrians can maintain a generally upright posture, pedestrians are allowed to have some subtle limb movements; the subtle movements can be ignored without affecting
the detection process. Therefore, HOG features are particularly suitable for human detection in this paper. The HOG feature extraction process is as follows:

**Gamma normalization.** The image is normalized to reduce the influence of local shadows and the background as shown below

$$I(x, y) = I(x, y)^{\text{gamma}}$$  \hspace{1cm} (12)

**Computing the image gradient.** The gradients of the horizontal and vertical coordinates of the image are calculated. The pedestrian contour and edge information can be captured in the derivation process according to the gradient direction values of each pixel position. The gradients of the pixels \((x, y)\) in the image are

$$G_x(x, y) = H(x + 1, y) - H(x - 1, y)$$  \hspace{1cm} (13)

$$G_y(x, y) = H(x, y + 1) - H(x, y - 1)$$  \hspace{1cm} (14)

\(G_x(x, y), G_y(x, y), \text{ and } H(x, y)\) represent the horizontal gradient, the vertical gradient, and the pixel value at the pixel point \((x, y)\) in the input image, respectively, and the gradient increment and the gradient direction at the pixel point \((x, y)\) are shown in equations (15) and (16), respectively

$$G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2}$$  \hspace{1cm} (15)

$$\alpha(x, y) = \tan^{-1}\left(\frac{G_y(x, y)}{G_x(x, y)}\right)$$  \hspace{1cm} (16)

In this paper, the \([-1,0,1]^T\) gradient operator is used to perform convolution operations on the original image to obtain the gradient components in the horizontal direction \(x\). Similarly, the \([1,0,-1]^T\) gradient operator is used to perform convolution operations on the original image to obtain the gradient components in the vertical direction \(y\). Then, the gradient size and direction of the pixel are calculated by combining the above formulas.

**Cell division.** The histogram of the oriented gradients of each cell element is constructed. The image is divided into several “cells”; in a 64 \(\times\) 128 image, each 8 \(\times\) 8 pixel constitutes a cell, and each 2 \(\times\) 2 cell constitutes a block. Using 8 pixels as the step size, then there are 7 scanning windows in the horizontal direction and 15 scanning windows in the vertical direction, with a total of \(36 \times 7 \times 15 = 3780\) features.

**Normalized gradient histogram.** The directional histogram corresponding to a single cell is transformed into a one-dimensional vector; the number of corresponding directional gradients is coded according to the prescribed group spacing (8, 10, 6, 12, 4, 5, 8, 6, 14), and nine features of a single cell are obtained. Each block (scanning window) contains 2 \(\times\) 2 cells, that is, \(2 \times 2 \times 9 = 36\) features. The final feature number of an image with a size of 64 \(\times\) 128 is \(36 \times 7 \times 15 = 3780\). In this way, an intuitive gradient graph is decomposed and extracted into a feature vector easily decoded by a computer.
Collecting HOG features. The last step involves collecting the HOG features from all overlapping blocks in the detection window and combining them into the final feature vector for classification.

Image classification based on SVM

The above HOG features can be classified by SVM to distinguish pedestrians from non-pedestrians. The equation of an hyperplane is

$$w^T x + b = 0$$  \hspace{1cm} (17)

where $x$ is the eigenvector. The optimal classification surface needs to meet the following condition

$$\min \frac{1}{2} \|w\|^2$$  \hspace{1cm} (18)

At the same time, the following conditions are met

$$y_i((w^T x_i + b) + 1) \geq 1, \quad i = 1, 2 \ldots n$$  \hspace{1cm} (19)

where $(x_i, y_i), \ i = 1, 2 \ldots n$ is the data sample set obtained by training, $y_i \in \{+1, -1\}$, $+1$ is a positive sample, $-1$ is a negative sample; the Lagrange multiplier $\alpha_i$ is used to convert to

$$\min \max \sum_{y_i} \left\{ \frac{1}{2} \|w\|^2 - \sum_{i=1}^{n} y_i [(w^T x_i + b) - 1] \right\}$$  \hspace{1cm} (20)

To solve $\alpha_i^*$, the following equations are used

$$w^* = \sum_{i=1}^{n} y_i \alpha_i^* x_i$$  \hspace{1cm} (21)

$$b^* = -\frac{1}{2} \left[ \min_{y_i = +1} (w, x_1) + \max_{y_i = -1} (w, x_1) \right]$$  \hspace{1cm} (22)

If the sample is not linearly separable, a slack variable $\xi$ needs to be added, which corresponds to the interval that allows the $i$th data point to deviate. If $\xi$ is arbitrarily large, any hyperplane meets the condition. Therefore, on the basis of the original objective, the total amount of $\xi$ should be as small as possible, and the new objective function should be changed into

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{n} \xi_i$$  \hspace{1cm} (23)
Equation (22) can be converted to

\[ y_i((w^T x_i + b) \geq 1 - \xi, \quad i = 1, 2 \ldots n) \]  

(24)

SVM can also transform features into high-dimensional spaces by a kernel function to solve nonlinear problems. The optimal classification function is

\[ g(x) = \text{sgn}\left( \sum z_i^* y_i K(x_i, x_j) + b^* \right) \]  

(25)

\( K(x_i, x_j) \) is the kernel function and the eigenvector \( x \) outputs the positive and negative categories of the target through the above formula.

**Experimental results and discussion**

**The problem system architecture**

In this paper, the Raspberry hardware platform and PC platform are used for analysis. As shown in Table 1, the frame rate of the video display can reach more than 30 frames per second. However, when comparing the volume, the hardware volume of the setup in this paper is much smaller than the volume of a desktop PC, which improves the portability of the system. At the same time, input/output (IO) ports exist on peripheral ports and improve the compatibility of peripheral devices. In addition, the Raspberry hardware platform is much cheaper than a PC (Liu Q, 2016). In conclusion, the hardware in this paper is more suitable for vehicle scenes.

**Lane line detection**

When the image area is divided, the adjustments of the dividing line are based on the farthest point in front of the lane line, when there is a lane line. To obtain the farthest point, the system carries out image edge detection and a series of image processing steps. The process of line detection is as follows:

- **Read the image** \( \text{img} = \text{cv2.imread(‘line.jpg’) \, and convert the image} \) \( \text{gray} = \text{cv2.cvtColor(\text{img}, \text{cv2. COLOR_BGR2GRAY})} \).
- **Edge detection of the processed image.** Canny operators in OpenCV can perform edge detection operations. The Canny function is \( \text{cv2.Canny(\text{Input, Minval, Maxval})} \). The first parameter of this function is the input image, and the second and third parameters are the minimum threshold and the maximum threshold, respectively. The setting of the threshold directly affects the effect of edge extraction result. To determine the true boundary, the two thresholds Minval and Maxval are set; as shown in Figure 4, the \( x \) axis corresponds to the pixel, and the \( y \) axis corresponds to the gradient of the pixel. When the gradient of pixel A is higher than the threshold Maxval, it can be considered as the true boundary point. Although pixel C is lower than Maxval, it is higher than Minval, and at the same time, it is on the same continuous line as pixel A, so it can also be considered as the true edge. The B-point of the pixel is between the two thresholds, and there is no connection with other lines beyond the threshold range, so the B-point of the
pixel is not the real boundary point. The threshold range of this system is 50–120, and the edges of edge images are cv2.Canny (gray, 50, 120). The effects of edge extraction are shown in Figure 4.

• Several straight lines are obtained by the probabilistic Hoff transform of the image after the edge detection step. The above edge image is calculated by using the statistical function cv2.HoughLinesP(DST, lines, rho, theta, threshold, minLineLength, maxLineGap) in OpenCV, where seven parameters are adjusted, DST is the binarized input image in Figure 5(b); the variable lines represent the parameters for storing the detected line; rho represents the resolution of the radius $\rho$ of the parameter polar coordinates in units of pixels; theta represents the resolution of the angle $\theta$ in the polar coordinates of the parameter; threshold represents the setting of the threshold, that is, the minimum curve intersection required for a straight line; minLineLength represents the minimum number of points that can form a straight line, as a line with an insufficient number of points is excluded; and maxLineGap represents the maximum distance of a point from a target line. The system was tested to select the ideal parameter value and detect the straight lines $=\text{cv2.HoughLinesP(edges, 1, np.pi/180, 100, 10, 10, 30)}$. The result is shown in Figure 6. The result is indicated by a green line.

• The straight lines are grouped, and the lane line is judged according to the condition of the slope of the straight line. In Figure 6, it can be seen that the detected straight lines are successfully marked with green lines; however, at the same time of correctly identifying the lane line, the pole beside the road is misidentified. This misidentification occurs because the pole and the lane line also show the straight line characteristics in the figure, and the previous step does not distinguish the specific straight line type. To select the required lane line, the difference between the pole and the lane line is analyzed. The largest difference is the difference in slope. Most of the poles are vertical objects on the image display, and the slope of the vertical line is infinite or non-existent. Most of the

| Hardware environment       | Video fluency           | Volume | IO | Price |
|----------------------------|-------------------------|--------|----|-------|
| The hardware platform in this paper | Up to 60 frames per second | Small | Yes | Low   |
| PC                         | Up to 60 frames per second | Big    | No  | High  |

Figure 4. Canny threshold schematic diagram.
tag lines have a slope with a diagonal line. The system sets a discrimination condition for the slope of the line. The conditional procedure as follows

for x1, y1, x2, y2 in lines1:
    if x1-x2 < y1-y2:
        cv2.line(img, (x1, y1), (x2, y2), (0, 255, 0), 2)
    elif (x1-x2) > (y1-y2) or (x1 == x2):
        cv2.line(img, (x1, y1), (x2, y2), (0, 0, 255), 2)

The result is shown in Figure 7(a), in which the green line indicates the correctly identified lane line, the red line indicates the misidentified pole, and Figure 7(b) shows the last display result of the recognition error.

According to Figure 7(b), the Y coordinate value of the farthest pixel of the successfully recognized lane line is obtained, and the pixel is used as the horizontal line; thus, the process of locating the image partition line, Line_2, is completed.
HOG feature extraction

The HOG descriptor is a very effective feature descriptor for target detection. Some local shape features of the image can be represented by using the gradient of the image and the direction density segment of the edge, which can well describe pedestrian shape features in static images or videos well.

In this system, $64 \times 128$ pixel images are taken as samples, and $8 \times 8$ pixel unit samples are divided into small cell units. Each cell unit is processed with gradient and edge histogram, and the histograms of the cell units are combined to form image features. To improve the detection performance of the feature, four adjacent cells are combined into one block. The formation of the block is generated by sliding each cell unit, and the block moves eight pixels at a time, that is, a cell width, so that the sample can obtain a total of 105 blocks of size $7 \times 15$. The pixel composition of each block is $16 \times 16$, and the relationship between the cells and blocks is shown in Figure 8.

For each cell unit of a block, it is required to project a gradient direction to obtain a gradient direction histogram for each individual cell unit. The direction of the gradient will be divided into nine intervals, the angle range is $0–180^\circ$, and every $20^\circ$ is an interval. The histogram of the gradient direction of four cell units in each block is connected in series to form the feature of the block, namely the feature of 36 dimensions. By traversing the entire sample, the feature vector of size $36 \times 105$ with a total of 3780 dimensions can be obtained.

Nighttime pedestrian classification based on SVM

In this paper, SVM is mainly adopted for classification decisions and is combined with HOG descriptor information. In the classification problem, SVM is faced with a binary classification problem. It was found in the experiment that the accuracy results of linear and nonlinear classification methods are nearly identical. In the end, the linear classification method is selected, as it is faster than the nonlinear method. Assuming that there is a joint probability $F(x, y)$ in the relationship between input $x$ and output $y$, the process of classifier learning involves obtaining an optimal function based on $n$ independent samples to be detected, so that the expected risk of prediction is minimized. The basic model of classifier

Figure 7. Straight line discrimination result chart: (a) discriminating result and (b) final result.
classification learning is shown in Figure 9, and the training process of the classifier is shown in Figure 10.

In the pedestrian detection test of the system, the results of the detection are obtained in an actual night environment. As shown in Figure 11(a), the light is sufficient under the street light, the pedestrian contour is obvious, the objects on the road are clearly visible, and the system can detect the pedestrian well. In Figure 11(b), with weak light, the pedestrian
contour is still visible, and the system can detect pedestrian well. However, due to the contrast between the light column generated by the street light and the surrounding dark light environment, the system detects the street light as a pedestrian. As shown in Figure 11 (c), in the dark environment, the infrared camera works, and the two pedestrians on the right are barely visible; however, the people on the left are blurred, so the system does not detect the pedestrian.

To verify the above results, the system performs data statistics in similar nighttime scenes. As shown in Table 2, under good light conditions, the accuracy of detection is the highest, reaching 91.8%. The second condition was weak illumination, with an accuracy of 90.1%. The lowest accuracy was in the case with no light, with an accuracy rate of 81.2%. Therefore, better the nighttime lighting conditions lead to the higher detection performance of the system. Regarding the false alarm rate, the highest rate was in the case with weak illumination, which indicates that street lights have a large impact on the detection results of the system under a weak illumination environment and that the system needs to be further optimized. In actual scenarios, most roads have street lights and car lights. The actual lighting conditions are mostly between weak and good lighting. Therefore, the system has high adaptability in a night environment.

To further improve the performance of the system, the system divides the image area into the sky and road in the image preprocessing step, eliminates the interference of the area above the image, and carries out frame separation processing to optimize the fluency of the video. As shown in Figure 12, pedestrians are correctly detected in (a) before optimization, but street light is detected incorrectly. After area division optimization in Figure 12(b), the pedestrian is correctly detected, and street light interference is reduced.

Table 2. Nighttime pedestrian detection data table under different lighting conditions.

| Night light conditions | Number of actual pedestrians, S | Number of pedestrians detected, r | Number of errors detected, f | False rate, F (%) | Missed detection rate, D (%) | Correct rate, R (%) |
|------------------------|---------------------------------|----------------------------------|----------------------------|------------------|-----------------------------|-------------------|
| Good                   | 1565                            | 1436                             | 81                         | 5.3              | 8.2                         | 91.8              |
| Weak                   | 1512                            | 1362                             | 143                        | 9.5              | 9.9                         | 90.1              |
| No light               | 1536                            | 1248                             | 54                         | 4.1              | 18.8                        | 81.2              |

Figure 11. Nighttime pedestrian detection results with different light intensities: (a) good lighting results, (b) weak lighting results, and (c) no lighting results.
The method before optimization is method 1 and the method after optimization is method 2. As shown in Table 3, the correct rate of method 2 is slightly higher than that of method 1, reaching 92.4%. The false positive rate is obviously lower than that of method 1, which is only 3.8%. The system maintains a high correct rate of pedestrian detection and effectively reduces the false positive rate. In addition, the average video frame rate of the video of method 2 reaches 28 frames/s, which is significantly better than that of method 1. In summary, method 2 is obviously superior to method 1. While maintaining a high accuracy rate, the false alarm rate and video fluency are also optimized.

**Conclusion**

The nighttime pedestrian detection system proposed in this paper can be applied to automobile driving, which is of great significance in night driving and provides important research value in the future development of unmanned driving system. At present, there are still many deficiencies in this system, and additional improvements are needed in the following aspects:

1. In the design of the pedestrian detection classifier, the correct rate of detection using the SVM is influenced by the number of training samples. This relation is not ideal at present. If we want to further improve the detection rate and reduce the number of false alarms, we need a sufficient number of training samples. At the same time, with an increase in the number of training samples, the processing time of the algorithm increases, which reduces...
the real-time performance of the system. In the following research, deep learning and SVM can be combined to improve the performance of the system.

2. The method of region division of the image proposed in this paper has not been fully adapted to actual road conditions. If a road without a lane line appears all the time and uphill and downhill situations occur at the same time; the effectiveness of this image region division method is halved. In a future area division method, it may be advantageous to add a gyroscope to determine whether the vehicle is in an uphill or downhill situation, and according to the offset angle with the horizontal direction, the image division area can be adjusted to improve the adaptability to the road.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Financial support for this work was provided by the Zhongshan Innovative Research Team Program (No. 180809162197886), the Zhongshan Institute high-level talent scientific research startup fund project in the University of Electronic Science and Technology of China (No. 416YKQ04), the National Key Research & Development Program of China (No. 2016YFB0401502), the Guangdong Province Higher Vocational Colleges & Schools Pearl River Scholar Funded Scheme (2016), and the Project of Shenzhen Science and Technology Innovation Committee (JCYJ20170817114522834).

ORCID iD

Bao Peng https://orcid.org/0000-0003-3964-3882

References

Aki M, Rojanaarpa T, Nakano K, et al. (2016) Road surface recognition using laser radar for automatic platooning. *IEEE Transactions on Intelligent Transportation Systems* 17(10): 2800–2810.
Aleena S and Anamik Y (2016) Data-mining-based fault during power swing identification in power transmission system. *IET Science Measurement & Technology* 10(2): 130–139.
Chen C, Jafari R and Kehtarnavaz N (2017) A survey of depth and inertial sensor fusion for human action recognition. *Multimedia Tools and Applications* 76: 4405–4425.
Chen K, Song X and Liu J (2018) Pedestrian tracking framework based on deep convolution network and SIFT. *Scientia Sinica Informationis* 48(7): 841–855.
Choudhury SK, Sa PK, Padhy RP, et al. (2017) Improved pedestrian detection using motion segmentation and silhouette orientation. *Multimedia Tools & Applications* 6: 1–40.
Han GJ, He Y, Jiang JF, et al. (2019) A synergetic trust model based on SVM in underwater acoustic sensor networks. *IEEE Transactions on Vehicular Technology* 68(11): 11239–11247.
Johansson G (1975) Visual motion perception. *Scientific American* 232: 76–88.
Kim T and Kim S (2018) Pedestrian detection at night time in FIR domain: Comprehensive study about temperature and brightness and new benchmark. *Pattern Recognition* 79: 44–54.
Liem MC and Gavrila DM (2014) Coupled Person orientation estimation and appearance modeling using spherical harmonics. *Image and Vision Computing* 32(10): 728–738.
Liu Q (2016) Pedestrian detection by combining guided probability map and saliency features. *High Technology Letters* 26(5): 464–474.
Liu X and Zhang XY (2019) Rate and energy efficiency improvements for 5G-based IoT with simultaneous transfer. IEEE Internet of Things Journal 6(4): 5971–5980.

Liu X, Zhang XY and Jia M (2018) 5G-based green broadband communication system design with simultaneous wireless information and power transfer. Physical Communication 25: 539–545.

Luo A, An F, Zhang X, et al. (2019) A hardware-efficient recognition accelerator using Haar-like feature and SVM classifier. IEEE Access 99: 1–1.

Na ZY, Wang YY, Li XT, et al. (2018) Subcarrier allocation based simultaneous wireless information and power transfer algorithm in 5G cooperative OFDM communication systems. Physical Communication 29: 164–170.

Peng HY, Chen WG, Wang HY, et al. (2014) Hybrid approach using map-based estimation and class-specific Hough forest for pedestrian counting and detection. IET Image Processing 8(12): 771–781.

Pan SY, Thomas M and Uttam A (2015) Developing a hybrid intrusion detection system using data mining for power systems. IEEE Transactions on Smart Grid 6(6): 3104–3113.

Pease G (2015) From Data to Action. Optimize Your Greatest Asset—Your People: How to Apply Analytics to Big Data to Improve Your Human Capital Investments. Hoboken, NJ: John Wiley & Sons, 2015, pp.41–58.

Qazi IUH, Alata O, Burie JC, et al. (2011) Choice of a pertinent color space for color texture characterization using parametric spectral analysis. Pattern Recognition 44(1): 16–31.

Schmarzo B (2015) Big Data MBA: Driving Business Strategies with Data Science. Hoboken, NJ: Wiley Publishing, 2015, pp.275–290.

Su SZ, Li SZ, Chen SY, et al. (2012) A survey on pedestrian detection. Chinese Journal of Electronics 40(4): 814–820.

Zeid A and Davis J (2014) Business Transformation: A Roadmap for Maximizing Organizational Insights. Hoboken, NJ: Wiley Publishing, 2014, pp.356–370.