ISSN: 2472-3703 DOI: 10.6911/WSRJ.202503_11(3).0013

Target Recognition and Analysis Based on Infrared Thermal Imaging Technology

Yiwang Dong[1, a]

[1]College of Physics and Electronic Engineering, Northwest Normal University, Lanzhou, China

[a]yiwdong@126.com

Abstract

Target recognition is a core task in computer vision and a key challenge in intelligent perception. Current research primarily focuses on target recognition in visible light images, but traditional methods often face significant challenges in low-visibility environments such as nighttime or foggy conditions. To address this issue, this paper is based on infrared thermal imaging technology. It extracts features from infrared images by analyzing the gradient direction and gradient magnitude, and classifies the extracted pedestrian features using a Support Vector Machine (SVM). The method aims to achieve pedestrian detection under low-visibility conditions. Experimental results demonstrate that the proposed algorithm significantly enhances pedestrian detection performance, exhibiting strong practicality and flexibility.

Keywords

Infrared thermography; Image segmentation; Target recognition.

1. INTRODUCTION

With the rapid development of intelligent transportation, public safety, and autonomous driving, target recognition technology has become increasingly widespread in the field of computer vision, particularly pedestrian detection, which has emerged as one of the core tasks in intelligent perception systems [1]. Traditional pedestrian detection methods typically rely on visible light images for target recognition. However, in low-visibility environments, such as at night, in heavy fog, or under harsh weather conditions, visibility is significantly reduced, and insufficient lighting makes it difficult to capture visual information from visible light images. This not only greatly diminishes the accuracy of pedestrian detection but also severely impacts its robustness, especially in complex or unpredictable environments. To address these challenges, infrared thermal imaging technology has garnered widespread attention and research in recent years as a method that remains effective even in low-light or completely dark environments [2]. Unlike visible light images, infrared thermal imaging captures the thermal radiation emitted by objects, allowing clear detection of targets even in total darkness, which significantly enhances the feasibility of pedestrian detection under low-visibility conditions. Infrared images are particularly suitable for nighttime surveillance, security patrols, and autonomous driving scenarios [3]. Fig 1 shows pedestrian images captured under visible light and infrared conditions. However, despite the unique advantages of infrared thermal imaging in low-light environments, it still faces certain challenges, mainly due to issues such as low contrast and high noise levels in infrared images [4]. These factors make target recognition in infrared images more difficult, especially in crowded or complex backgrounds, where the accuracy and robustness of pedestrian detection still leave significant room for improvement. Therefore, enhancing the quality of infrared images and effectively improving the accuracy of pedestrian detection has become an important research topic in the current field.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202503_11(3).0013

To address the aforementioned issues, this paper proposes an innovative pedestrian detection method based on infrared thermal imaging technology, aiming to improve the accuracy and robustness of pedestrian detection in low-visibility environments. Specifically, the method first performs statistical analysis of the gradient direction and gradient intensity of the infrared images to extract effective pedestrian features. This process effectively enhances the edge information in the images and emphasizes the thermal radiation characteristics of pedestrians. Subsequently, a Support Vector Machine (SVM) classifier is employed to classify and identify these extracted features, thereby achieving accurate pedestrian detection. Compared with traditional pedestrian detection methods based on visible light images, this method demonstrates stronger adaptability and robustness in complex environments such as nighttime, heavy fog, rain, and snow, effectively overcoming the difficulties caused by insufficient lighting or poor visibility.



Figure 1. Comparison of visible light imaging and infrared thermal imaging performance at night. From left to right are visible light image and infrared image.

Experimental results show that the pedestrian detection method based on infrared thermal imaging significantly outperforms traditional methods in target detection performance under low-visibility conditions, particularly in complex scenes and low-visibility environments, where it maintains a high detection accuracy. With this method, pedestrian detection is not only more reliable but also capable of efficient recognition without relying on strong light conditions, offering considerable practical utility and flexibility. Moreover, infrared thermal imaging technology enables the detection system to operate continuously and stably in all weather conditions, regardless of day or night. This study provides a novel technological approach for pedestrian detection under low-visibility conditions and offers valuable insights and references for applications in intelligent perception systems, particularly in fields such as autonomous driving and security surveillance.

2. EXPERIMENTAL PROCEDURE

2.1. Image Preprocessing

Preprocessing of infrared images is the first step in the target recognition process and a crucial link to ensure the effectiveness of subsequent detection and recognition tasks. In infrared images, distinguishing between the target and the background is often challenging due to various factors such as the unevenness of thermal radiation, sensor noise, and environmental influences [5]. Therefore, preprocessing not only helps to effectively remove noise and improve image quality, but also enhances the contrast between the target and the background, thereby making subsequent target detection and recognition more accurate and reliable. The quality of

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202503_11(3).0013

preprocessing directly impacts the overall performance of the target recognition system, making it an extremely important step in the image processing pipeline [6].

Common infrared image filtering methods include linear filtering, nonlinear filtering, and frequency domain filtering, each with its own unique advantages in noise removal and contrast enhancement. Median filtering, a common method in nonlinear filtering, operates by replacing the original value of a pixel with the median value of all the pixels in its neighborhood. This method effectively removes salt-and-pepper noise and other types of interference while avoiding the blurring effects that may result from linear filtering. Consequently, it preserves the edge information of the image during the denoising process [7]. Median filtering not only protects the structural details of the image during noise removal but also adapts better to various types of noise distributions, especially in cases where the noise is significant or the image quality is poor, demonstrating strong denoising capabilities.

Despite the excellent performance of median filtering in noise removal, it may still cause blurring of image details in certain specific scenarios, particularly when noise is sparse or uniformly distributed. To overcome this issue and further enhance the effectiveness of image preprocessing, this paper proposes an improved hierarchical median filtering method. This method applies different window sizes for filtering at different scales, processing the image details and noise at various levels. This multi-scale filtering strategy allows for flexible adjustment of the filtering strength according to the different features of the image, effectively preventing detail blurring while enhancing noise suppression. By using hierarchical median filtering, the key features of the image can be better preserved, reducing interference with important information, and thereby improving the accuracy of subsequent target detection and recognition. The mathematical expression of this method is as follows:

$$g(x,y) = medf\{f(x-k, y-l)\}, (k,l \in \omega) \quad (1)$$

where $f(x,y)$ represent the original image, $g(x,y)$ represent the filtered image, and ω is a two-dimensional template, with its size manually specified. This paper compares several common filtering methods, as illustrated in Fig 2. Hierarchical median filtering operates at different scales, allowing for more effective handling of complex noise with varying types and scales. As a result, it is more flexible and efficient than single-scale median filtering.



Figure 2. Processing results of different filtering methods. From left to right: original image, layered median filtering, median filtering, gaussian filtering, and mean filtering.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202503_11(3).0013

2.2. Feature Extraction

In the process of target recognition in infrared images, feature extraction is a critical step that directly impacts the final recognition accuracy and performance [8]. Effective features in the image can provide important information about the target's shape, texture, and structure. Therefore, the quality of feature extraction is crucial for improving target recognition effectiveness. In practical applications, selecting appropriate feature extraction methods is one of the key steps to enhance recognition performance. For infrared images, the significant features of the target include heat source distribution, edge structure, and texture information [9]. Hence, during the feature extraction phase, it is essential to optimize and select features that accurately describe the target while minimizing interference from environmental factors such as lighting and noise.

Histogram of Oriented Gradients (HOG) is a widely used feature extraction algorithm in object detection and image description, particularly effective for extracting edge and shape information from images. The HOG features capture edge and contour information by statistically analyzing the gradient directions within local regions of the image. This is especially important in infrared image target recognition, as infrared images contain rich information about the target's shape and edges. The advantage of HOG features lies in their ability to effectively capture the shape information of the target, and since they rely solely on gradient information, they are not affected by the image's absolute brightness, making them robust to lighting variations [10]. Moreover, HOG features are well-suited for detecting targets at different scales and in various background environments. In the context of infrared images, HOG features can accurately capture the target's thermal contours and variations in thermal intensity, aiding the system in recognizing different target objects. Therefore, in this study, we have chosen the HOG algorithm to perform feature extraction for target recognition in infrared images, with the aim of further improving recognition accuracy and stability.

The HOG feature extraction process typically involves several steps: first, the image is divided into small cells, and the gradient magnitude and direction of each pixel within the cell are computed. Next, based on these gradient values, a histogram of gradient directions is constructed, representing the frequency of occurrence of different gradient directions within each cell. To enhance the robustness of the features, these histograms are often aggregated within neighboring cell regions to form larger blocks, followed by normalization to reduce the impact of external factors such as lighting variations. Finally, by combining the gradient direction histograms over the entire image, a high-dimensional feature vector is obtained, representing the global edge information of the image. The main process is illustrated in Figure 3.

The detailed process of HOG feature extraction is as follows:

(1) Gamma Correction

In infrared image processing, due to the nonlinear response of infrared imaging devices to thermal radiation, variations in external lighting conditions, and noise from the device, the brightness distribution in the image may become uneven, which can adversely affect the accuracy of subsequent feature extraction and target recognition. Gamma correction adjusts the pixel values of the image nonlinearly to enhance the details of both the dark and bright areas, thereby improving the visual quality of the image. The basic idea behind gamma correction is to adjust the image's brightness according to a given gamma value (γ). Specifically, each pixel value $g(x, y)$ in the image is transformed according to the following equation:

$$I(x, y) = g(x, y)^{gamma} \tag{2}$$

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202503_11(3).0013

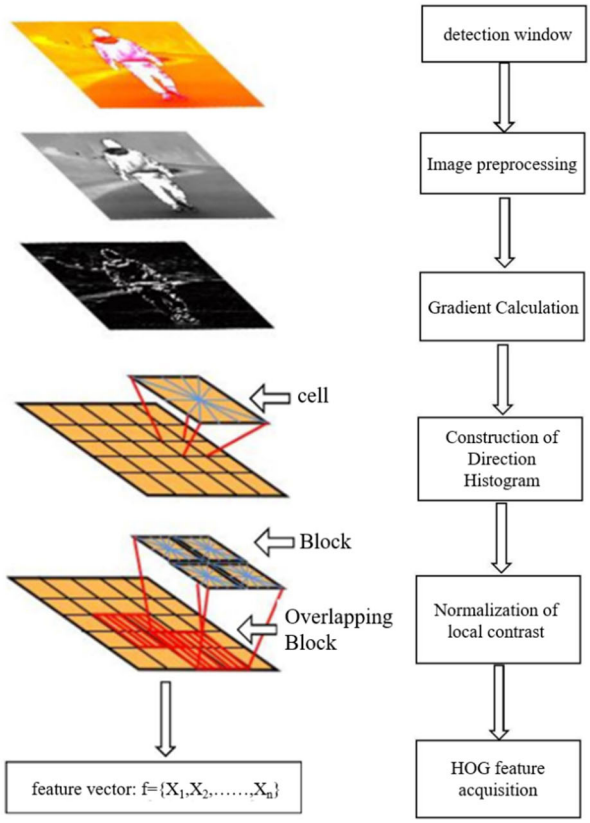


Figure 3. The flowchart of HOG Feature Extraction

Where $g(x,y)$ is the pixel value of the input image, $I(x,y)$ is the pixel value after gamma correction, and gamma controls the degree of brightness adjustment.

(2) Gradient Calculation

The gradient values along the x-axis and y-axis are calculated for each pixel in the image. The gradient magnitude and direction for each pixel are then computed to reduce the impact of lighting variations and to capture edge features of the image. In simple terms, a larger gradient value indicates a more significant change in brightness, which typically corresponds to the edges of the image. Since gradient calculation focuses on the local variation in pixel values rather than the overall brightness difference, it effectively minimizes the influence of lighting changes in the image.

The formulas for calculating the gradients along the x-axis and y-axis are as follows:

$$G_x(x,y) = H(x+1,y) - H(x-1,y) \tag{3}$$

$$G_y(x,y) = H(x,y+1) - H(x,y-1) \tag{4}$$

Where $H(x,y)$ represents the grayscale value at pixel point (x,y). Based on the obtained values $G_x(x,y)$ and $G_y(x,y)$, the gradient magnitude and direction for the target can be computed as follows:

$$G(x,y) = \sqrt{G_x(x,y)^2 + G_y(x,y)^2} \tag{5}$$

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202503_11(3).0013

$$\alpha(x,y) = tan^{-1}\left(\frac{G_x(x,y)}{G_y(x,y)}\right) \quad (6)$$

(3) Constructing the Gradient Orientation Histogram

The HOG (Histogram of Oriented Gradients) feature for each cell is computed by accumulating the gradient information of the pixels within the cell and categorizing the gradients into specific orientation bins. The gradient histogram consists of 9 bins, each of which effectively represents the gradient intensity within a particular angular range from 0° to 360°. The distribution of these directional bins is shown in Figure 4.

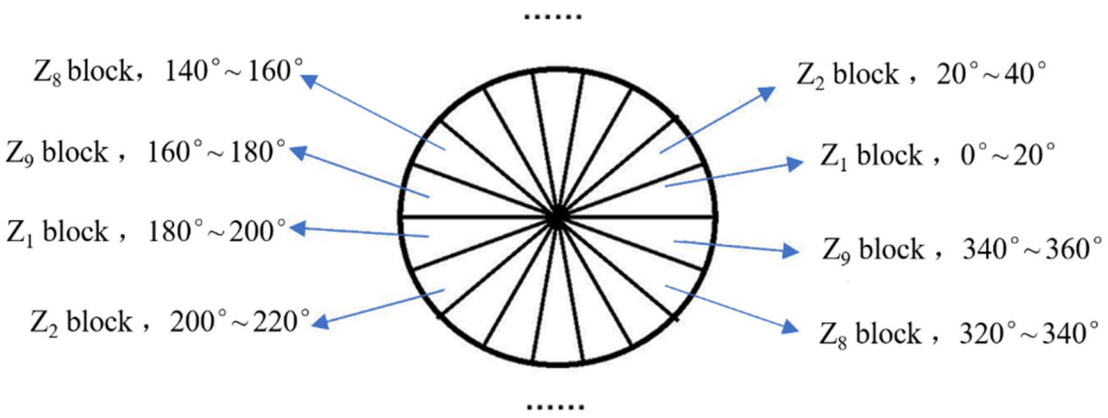


Figure 4. The distribution interval of striation

(4) Statistical HOG Feature Vector of the Object Window

The HOG feature extraction method computes local gradient information within an image to construct feature descriptors. Specifically, detection blocks are moved step-by-step across the image along both horizontal and vertical directions, covering every part of the image to ensure that the entire region is scanned and processed. Within each detection block, the gradient value of each pixel is first computed, yielding the local gradient directions and magnitudes within that region. Then, based on the gradient directions, these values are quantized and divided into multiple bins corresponding to specified angular ranges, forming histograms for each direction. Finally, the gradient values from these local directions are accumulated to generate the feature vector for the detection block. Fig. 5 illustrates the HOG feature maps for both positive and negative samples, along with their corresponding gradient orientation histograms. These example images clearly show the distribution of HOG features across different sample types and the variation in gradient directions.

Furthermore, the HOG feature extraction method demonstrates strong robustness to external lighting variations, maintaining consistent feature representation under varying illumination conditions. This property allows HOG to effectively capture the shape and contour information of the object, even in the presence of lighting changes or shadows, thereby mitigating interference caused by such variations. Additionally, the HOG method excels in capturing local information within an image, particularly in recognizing object contours and edge details, demonstrating exceptional recognition capability. These attributes make HOG features highly effective and widely applied in tasks such as pedestrian detection and object recognition.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202503_11(3).0013

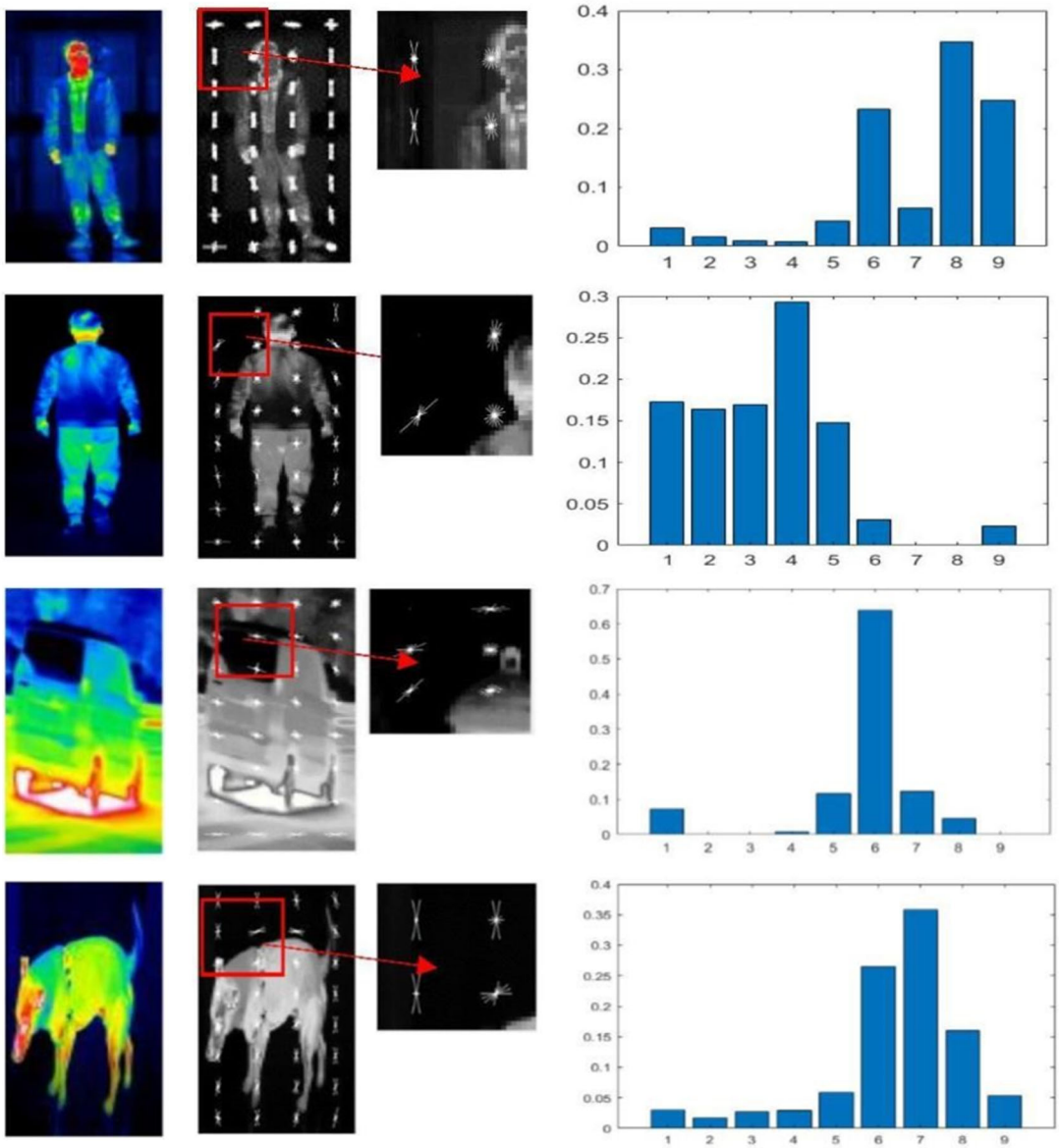


Figure 5. The HOG features and Gradient Distribution Histograms of samples. From left to right: Sample, HOG features, and Gradient Direction Distribution Histogram; the first and second rows represent positive samples, while the third and fourth rows represent negative samples.

2.3. Classifier Training

Support Vector Machine (SVM) is a binary classification algorithm that aims to find a hyperplane in the feature space that maximizes the margin between the two classes, thereby effectively separating the samples. Specifically, SVM performs classification by constructing an optimal hyperplane, which maximizes the margin $\frac{2}{||\omega||}$ between the hyperplane and the closest samples from both classes, known as the support vectors. To enhance the accuracy of the classifier, the objective of SVM is to select the hyperplane that maximizes the margin, ensuring that the positive and negative samples are perfectly separated, thus minimizing the risk of classification errors. Support vectors are the data points located closest to the decision boundary and are crucial in determining the position of the hyperplane. These points are key to constructing the classification model. By maximizing the margin between the support vectors and the hyperplane, SVM improves both the classification ability and the generalization performance of the model.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202503_11(3).0013

Based on geometric principles, to maximize the classification margin $\frac{2}{||\omega||}$, it is equivalent to minimizing $||\omega||$. The value of $||\omega||$ can be obtained by solving the following mathematical expression:

$$\begin{cases} \frac{min}{w,b} \frac{||w||^2}{2} \\ y_i(w \cdot x_i + b) \geq 1, i = 1,2, \dots \dots, n \end{cases} \tag{7}$$

The expression contains inequality constraints, and optimizing the solution is a key step. Typically, the Lagrange multiplier method is employed to formulate the following expression:

$$\mathrm{L}(\mathrm{w}, \mathrm{b}, \alpha) = \frac{||w||}{2} - \sum_{i=1}^{n} \alpha_i (y_i(w \cdot x_i + b) - 1) \tag{8}$$

According to the solution method of the Lagrange multiplier, by taking partial derivatives of the above expression with respect to w,b, and α, and setting the derivatives equal to zero, the following formula is obtained:

$$\frac{\partial \mathrm{L}}{\partial \mathrm{w}} = 0 \rightarrow \begin{cases} w = \sum_{i=1}^{n} \alpha_i y_i x_i \\ \sum_{i=1}^{n} \alpha_i y_i \end{cases} \tag{9}$$

Substituting the above expression into the Lagrange multiplier method, and considering the corresponding constraint conditions, the final mathematical expression is obtained:

$$\begin{cases} max_{\alpha} = \sum_{i=1}^{n} \alpha_i - \frac{\sum_{i,j}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j}{2} \\ \alpha_i \geq 0, i = 1,2, \dots \dots, n \\ \sum_{i=1}^{n} \alpha_i y_i = 0 \end{cases} \tag{10}$$

After solving using the Lagrange multiplier method, the values of w,b and the support vector machine (SVM) for the given sample can be determined based on the positive and negative samples, as well as the optimized algorithm. The test sample vector can then be input into the SVM classifier for classification and recognition.

3. EXPERIMENTAL RESULTS

3.1. Dataset Collection and Preparation

In this experiment, infrared image data were collected using the VarioCAM®HD980 infrared thermal camera, which included 300 positive samples and 500 negative samples, along with three sets of test samples. Considering the large-scale positive and negative sample requirements for Support Vector Machine (SVM) training, the sample set was expanded through data augmentation techniques. Specifically, methods such as noise addition and image rotation were employed, resulting in a dataset containing 1500 positive samples (with dimensions 64×128) and 4500 negative samples (with dimensions 64×128) to meet the needs of SVM training. Fig. 6 shows some typical positive and negative samples from the training set, as well as the infrared images after augmentation.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202503_11(3).0013

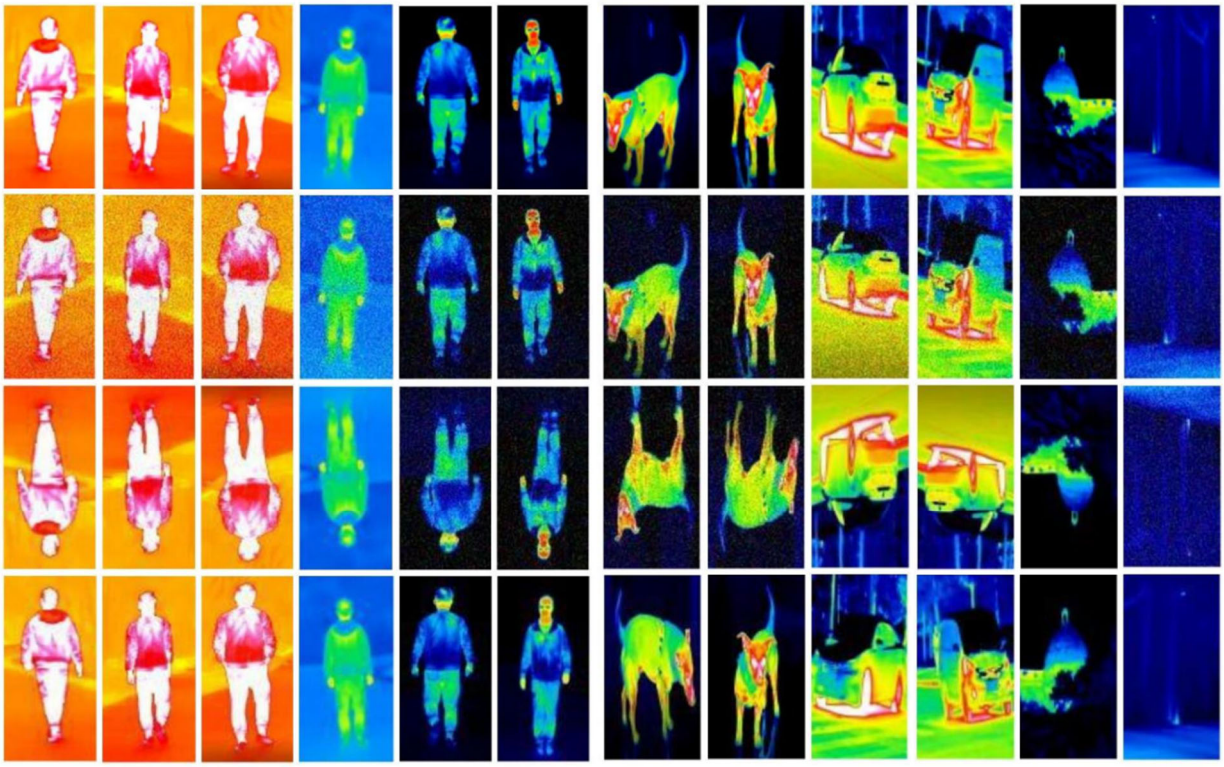


Figure 6. Some samples from the training set. From left to right: positive sample set and negative sample set. From top to bottom: original samples, noisy samples, flipped samples, and mirrored samples.

3.2. Recognition Performance

In the experiment, the augmented positive and negative samples underwent HOG (Histogram of Oriented Gradients) feature extraction to obtain high-dimensional feature vectors for each sample. These feature vectors were then fed into the Support Vector Machine (SVM) model for training. Through this process, SVM effectively learned the distinguishing features between different classes and automatically constructed the optimal classification decision boundary.

The working principle of SVM is to find an optimal hyperplane by maximizing the margin between the positive and negative classes, thereby accurately distinguishing between positive and negative samples. During the training process, SVM continuously adjusts model parameters based on the features of the samples, optimizing the decision function, and ultimately enabling the model to minimize classification errors on the training set, while improving its generalization ability to unknown samples. To ensure the model's robustness, the training and test sets were strictly divided according to a certain ratio, ensuring that the model could demonstrate good recognition performance across different datasets.

After training, to validate the model's actual recognition effectiveness, we evaluated the model using images from the test set. Fig. 7 shows the recognition results of crowd counts in different test images. As can be seen, the model successfully recognized the crowd count in most test images, and as the crowd size increased, the model's recognition accuracy continuously improved. This indicates that the SVM model, after sufficient training, is capable of accurately recognizing crowds in real-world applications, laying a foundation for further application development.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202503_11(3).0013

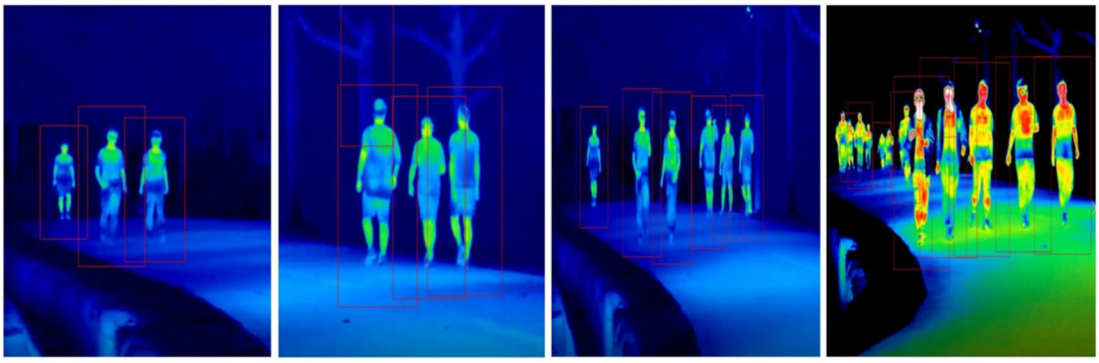


Figure 7. Recognition performance with varying numbers of individuals.

To quantitatively evaluate the algorithm, several objective evaluation metrics are introduced in this paper, with accuracy being the most commonly used metric. A higher accuracy indicates better classifier performance; however, relying solely on this metric is insufficient for evaluating the algorithm, as accuracy can be misleading in cases of imbalanced positive and negative samples. In this paper, accuracy, detection rate, and detection precision are used as evaluation metrics to objectively assess the detection results. Their mathematical expressions are as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

$$FNR = \frac{FN}{TP + FN} \tag{12}$$

$$Precision = \frac{TP}{TP + FP} \tag{13}$$

where, TP (True Positive) refers to the number of positive samples correctly predicted as positive, TN (True Negative) refers to the number of negative samples correctly predicted as negative, FP (False Positive) refers to the number of negative samples incorrectly predicted as positive, and FN (False Negative) refers to the number of positive samples incorrectly predicted as negative. In target detection results, Accuracy, Detection Rate (FNR), and Precision are key metrics for evaluating the performance of classification models [12]. Accuracy measures the proportion of correctly predicted samples across all predictions. However, in cases of uneven data distribution or imbalanced sample sizes, accuracy may fail to comprehensively reflect the true performance of the model. The Detection Rate focuses on the proportion of actual positive samples that are not correctly identified as positive, making it suitable for scenarios where false negatives are costly. Precision, on the other hand, measures the proportion of samples predicted as positive that are actually positive, and is commonly used in tasks where false positives carry a higher cost. In result evaluation, these three metrics are typically used in combination to provide a more comprehensive view of the model's performance, thereby avoiding potential biases introduced by relying on a single metric [13].

The experiment analyzed the results of three sets of test samples (500 images per set) on both the proposed algorithm and OpenCV, as shown in Table 1.

Fig. 8 shows the detection results of the OpenCV pedestrian detector and the proposed algorithm. It can be observed that in images with low contrast, the OpenCV detector incorrectly

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202503_11(3).0013

identifies the background as pedestrians in three instances, leading to false positives and multiple detections of the same object.

Table 1. Statistical Analysis of Target Recognition Results

Method	Test-set	Number of samples	Number of targets/(P)	Number of correctly detected items/(TP)	Number of missed detections/(FN)	Number of errors detected/(FP)
OpenCV	1	500	1336	1132	196	115
	2	500	1027	887	124	73
	3	500	873	754	107	67
Ours	1	500	1336	1182	151	87
	2	500	1027	925	103	61
	3	500	873	781	91	53

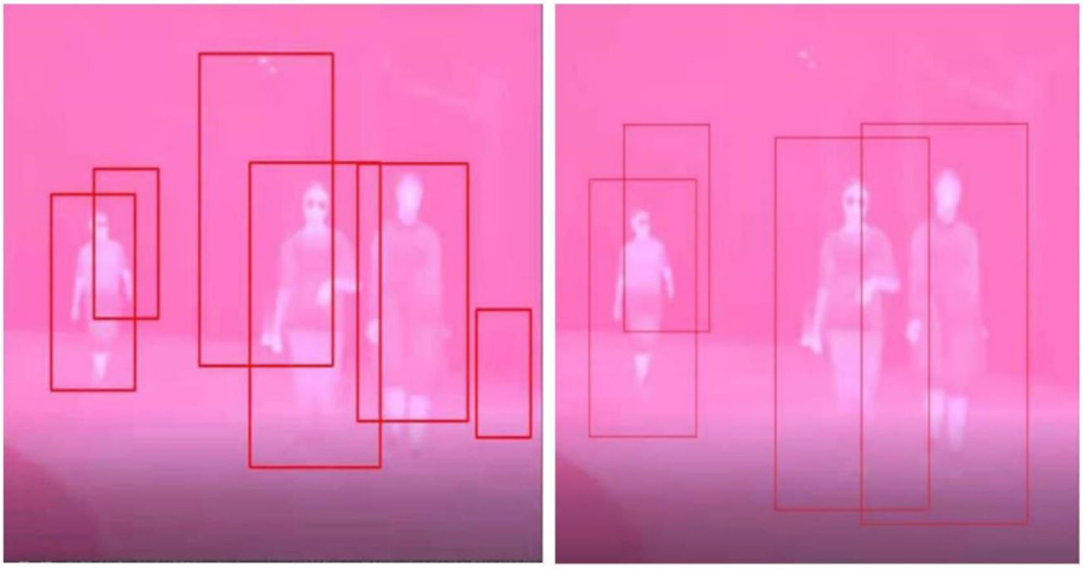


Figure 8. Pedestrian recognition results. From left to right: results of the OpenCV and our algorithm.

4. CONCLUSION

This paper presents a method for pedestrian detection and recognition in nighttime conditions based on infrared thermal imaging technology. The detailed process of Histogram of Oriented Gradients (HOG) feature extraction is analyzed, and a Support Vector Machine (SVM) classifier is trained to obtain the optimal hyperplane, ultimately achieving pedestrian detection at night. By extracting HOG features, local shape information of the target is captured, and SVM is employed for classification, effectively distinguishing pedestrians from the background. To validate the effectiveness of the proposed method, the experimental results are compared with the detection performance of the OpenCV pedestrian detector. The experiments demonstrate that the proposed method outperforms the OpenCV detector in terms of higher accuracy and lower false positive rates in low-light nighttime environments.

REFERENCES

[1] Q.F. Li, and Y.H. Zhang: Rainy pedestrian detection method based on improved YOLOv8, Journal of Guangxi University of Science and Technology (2025) p.1-11.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202503_11(3).0013

[2] S. Yang: *Research on Human Pose Recognition Method Based on Infrared Images* (MS., Shenyang Aerospace University, China 2019), p.21.

[3] Q.H. He, Z.X. Li and L.F. Reng: Infrared UAV Object Detection Based on Detailed and Global Multistage Fusion, Infrared Technology (2024), p. 1-8.

[4] Y.T. Dong: Research on Infrared Thermal Imaging for Human Object Detection and Recognition Technology (MS., Yanshan University, China 2018), p.14.

[5] R.Z. Ge and Y. Wu: Image fusion pedestrian detection based on stage feature fusion, Electronic Measurement Technology, Vol. 47 (2024) p. 103-109.

[6] Y. Man: *Research on Infrared Image Preprocessing and Small Target Detection Algorithm* (Ph.D., University of Chinese Academy of Sciences, China 2023), p.18

[7] X.Y. Li, J.Z. Yang and J.X. Yang: Infrared image enhancement algorithm based on improved histogram equalization and edge preserving smoothing filtering, Computer Applications and Software,Vol. 36 (2019) No.3, p.96-103.

[8] S.H. Zhou, X.Z. Liu and X.L. Li: Infrared ship target detection model integrating cross attention and dual feature interaction, Computer Engineering (2025) p.1-13.

[9] J.C. Gong, D.X. Sun and R.H. Peng: Intelligent detection technology for camouflage targets under low light conditions based on visible infrared feature level fusion, Command Control & Simulation (2025) p.1-10.

[10] Syaputra R, Syamsuar D, and Negara E S: *Multiple Smile Detection Using Histogram of Oriented Gradient and Support Vector Machine Methods*. IOP Conference Series: Materials Science and Engineering, 2021, 1071(1).

[11] X.M. Fang, H.L. Wang and H. Xu: Multi component trace gas recognition and detection based on SVM and near-infrared TDLAS technology, Spectroscopy and Spectral Analysis, Vol. 44 (2024) No.10, p.2909-2915.

[12] B. Zhang and Y.P.Zhao: Pedestrian detection algorithm based on progressive multispectral images, Computer Engineering and Design, Vol. 45 (2024) No.10, p.3042-3050.

[13] W.M. Huang, Z. B. Wu and M. Tao: Design and Implementation of Embedded Pedestrian Detection and Tracking System Based on HOG and SVM, İnternet of Things Technologies, Vol.13 (2023) No.8, p29-32.