Research on Abnormal Target Recognition of Full Information Mobile Monitoring Based on Machine Vision

Yudong Wei(✉) and Yuhong Xia

Chengdu College of University of Electronic Science and Technology of China, Chengdu 611731, China
weiyudong@cduestc.edu.cn

Abstract. When the color of moving object is close to the background, the accuracy of moving object recognition is affected. So the method of moving object recognition based on machine vision is designed. In order to reduce the distortion of image edge position, the moving object is calibrated and corrected by vision. In order to reduce the influence of noise to a controllable range, the full information mobile monitoring image is enhanced to preserve the image details. The edge features obtained from view and template are calculated by moment, and the similarity is obtained. Then the contour feature of moving monitoring target is extracted based on machine vision. Segmentation of the background region, according to the moving object trajectory center point information such as speed, direction and so on to determine whether the trajectory is abnormal events. The proposed method is tested on INRIA dataset and Vehicle Reld dataset, and the results show that the proposed method can improve the accuracy and recall rate and has good detection performance.

Keywords: Machine vision · Full information · Mobile monitoring · Abnormal target · Target identification · Monitoring objectives

1 Introduction

Full-information mobile monitoring can be networked transmission of information, and has the ability to remote real-time video viewing. Full-information mobile surveillance anomaly target recognition can analyze surveillance video streams, detect anomalous behavior and alert in a timely manner at less cost [1]. At present, most of the video surveillance is still a simple function that only provides the collection, storage and playback of video. If there is an abnormality, we need to rely on the monitors to view and analyze the video scene through human operation and human eyes detection. For the massive video data, data analysis is difficult, the workload is very large, and a lot of human resources are needed. How to get the information which we are interested in from the massive surveillance video data, this is the question which the video surveillance technology develops to the intelligent direction faces. Traditional video surveillance

W. Fu and L. Yun (Eds.): ADHIP 2022, LNICST 469, pp. 720–733, 2023.
https://doi.org/10.1007/978-3-031-28867-8_52

can not meet the needs of practical application, so the research of intelligent video surveillance system based on computer vision has been paid more and more attention.

Chen et al. [2] proposed the research of full information alarm and monitoring system based on mobile network. Through the location positioning class, the A-GPS module is used to obtain location information, store the mobile multimedia information center and automatically forward it to the preset number monitoring terminal, so as to realize the recognition and alarm of abnormalities in full information monitoring. Ma [3] proposed the design of video image tracking and monitoring system based on target detection, which uses DSP + FPGA to form the core part of hardware, converts video signals into recognizable TTL level signals, and applies drive servo technology to achieve stable target tracking. Hu et al. [4] proposed an end-to-end SSD real-time video monitoring abnormal target detection and localization algorithm. By setting a target preselection box in the convolutional neural network, the abnormal classification and abnormal target boundary box are obtained to complete the abnormal detection. However, the above three methods do not consider the problems of image distortion and noise, resulting in low detection accuracy.

Therefore, an abnormal target recognition method based on machine vision for full information mobile monitoring is proposed. Through the visual calibration and correction of the moving monitoring target, the distortion of the image edge position is reduced. The method of moments is used to calculate the edge features of the view and the template, and the similarity is obtained. According to machine vision, the contour features of moving objects are extracted to detect abnormal objects. This method has certain theoretical and practical significance for improving intelligence, real-time and efficiency.

2 Abnormal Target Recognition Method of Full Information Mobile Monitoring Based on Machine Vision

2.1 Visual Calibration and Correction of Moving Monitoring Targets

The object in the real environment is projected onto a plane by the imaging system to form an image. The color of each pixel in the picture shows the “reaction” of the real environment after receiving the light at the same position. The confirmation of coordinate system is the key to locate objects. Vision systems often contain worlds, pixels, images, and camera coordinates. For the determination of the placement of the camera and the object, there must be a reference to describe their corresponding positions. The position of the pixel in the image is geometrically related to the corresponding position in the real environment. The projection model of an imaging system concretizes and digitizes this geometric connection. The most useful optical imaging model is the central projection model, i. e. the pinhole imaging model. A coordinate system is established in space according to the datum, so that any object in the same space can describe its position in the coordinate system. Before using machine vision for target recognition, the first choice is to calibrate and correct the moving monitoring target visually. The camera coordinate system is generally based on the camera, and its origin is the center of plane projection. The focal length of the camera is the distance between the origin of its coordinate system and the origin of projection plane coordinate system. The pixel coordinate system is unique in the view and can not be set at will, so it is difficult to calculate the conversion relationship between the pixel coordinate system and the camera position. The focal length varies with the position of the camera. The basic unit is the millimeter. The camera coordinate system coincides with the optical axis, and the X and Y axes are parallel to each other, which are similar to the projection plane coordinate system. The camera is calibrated to obtain its internal and external parameters, which represent the change of the azimuth of the view and the camera as well as the position of the camera in the world coordinate system. The center projection can be used between the camera coordinate system and the image coordinate system, and the conversion relationship is as follows:

$$r\begin{bmatrix}\alpha\\ \beta\\ 1\end{bmatrix} = \begin{bmatrix}z\ 0\ 0\ 0\\ 0\ z\ 0\ 0\\ 0\ 0\ 1\ 0\end{bmatrix}\begin{bmatrix}p\\ q\\ r\\ 1\end{bmatrix} \tag{1}$$

In formula (1), z represents the camera focal length, (p, q, r) represents the camera coordinate system, and (α, β) represents the image coordinate system. Camera parameters are determined, not to be changed, is the camera's basic characteristics, scale factor, the coordinates of the pixel coordinates are included. The internal parameters will deviate from the values in the specification, so it is necessary to re-calibrate them in the application of machine vision. Camera calibration is the link between 2D image information and 3D information in reality. Zhang Zhengyou calibration method can keep high accuracy and robustness while considering lens distortion, and only need to print a piece of standard checkerboard as calibration object in practice. Zhang Zhengyou's calibration method has been widely used in the field of computer vision since its inception because of its high accuracy and flexibility. In this paper, Zhang Zhengyou calibration method, calibration template selected standard checkerboard. The homogeneous coordinates of corresponding points are obtained by projecting the points on the calibration template plane onto the image plane, and the mapping matrix can be expressed as follows:

$$W = \vartheta s\left[\varphi_1\ \varphi_2\ \lambda\right] \tag{2}$$

In formula (2), W represents the mapping matrix, ϑ is the camera internal parameter matrix, s represents any standard vector, φ_1, φ_2 and λ are the rotation matrix and translation vector of the camera coordinate system relative to the world coordinate system. In the process of calibration, the image will be distorted or distorted, and the position of the real object in the picture will be different from the expected position. This is called distortion. The distortion parameters are calculated by maximum likelihood estimation. Because the location of distortion is usually the edge of the image, the selection region should be as close to the edge as possible, and the average value should be obtained after calculating multiple selection points. Based on this, the visual calibration and correction of moving monitoring target are completed.

2.2 Full Information Mobile Monitoring Image Enhancement Processing

In the process of image acquisition, the image quality is often damaged by the interference of external factors. In addition, the result of image post-processing is not ideal. Eliminating these disturbances is a necessary condition for the next operation. The image processing process is shown in Fig. 1.

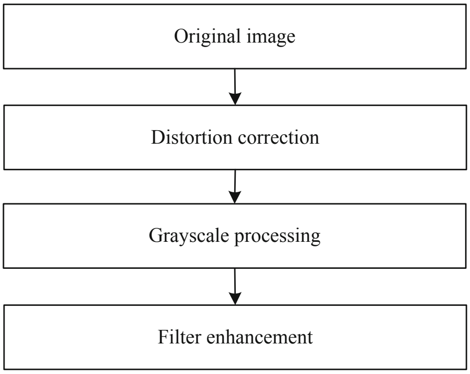

Fig. 1. Image processing process

Because the image captured by the camera is described and stored based on RGB color model, the color image is gray-processed. The key of image preprocessing is to reduce the influence of noise to a controllable range while preserving image detail features. Aiming at the characteristics of noise gray value, the filter is used to suppress the influence of noise and keep the useful and reliable information is the common denoising method. Because the gray image contains only a single channel array of pixels, the amount of data is small, which is conducive to improve the processing speed of image video, and can meet the basic requirements of video applications. Therefore, when processing and monitoring video images, color images in RGB format are usually converted to gray images [5]. The mean values of R, G and B components are taken as the gray values of pixels after gray-leveling, and the calculation formula is as follows:

$$G(\alpha, \beta) = \frac{X_1(\alpha, \beta) + X_2(\alpha, \beta) + X_3(\alpha, \beta)}{3} \tag{3}$$

In formula (3), $G(\alpha, \beta)$ is the gray value of grayed pixels, $X_1(\alpha, \beta), X_2(\alpha, \beta), X_3(\alpha, \beta)$ respectively correspond to the R, G and B components of the pixel at position (α, β). Because of the aberration of optical system, the diffraction of optical imaging, the nonlinear distortion of imaging system, the random noise of image capturing device and other factors, the image quality will be degraded in the process of capturing information moving monitoring image. The Gaussian filter takes into account the influence of the near distance, so its ability to reduce the influence of noise to a controllable range is better than that of the neighborhood mean filter. In this paper, Gaussian low-pass filtering is used to enhance the image. This process can be expressed as follows:

$$F(\alpha, \beta) = e^{\frac{-h^2(\alpha,\beta)}{2\varepsilon^2}} \tag{4}$$

In formula (4), $F(\alpha, \beta)$ represents the enhanced image, e represents the natural constant, h represents the distance between the pixels at position (α, β) and the central pixel, and ε represents the standard deviation. The adjustment of ε can change the smoothness of the preprocessed image. The larger the ε parameter, the smoother the image. Image enhancement can highlight or suppress the image information of a specific part according to the specific needs to achieve rapid extraction of feature information. But the quality of image preprocessing does not mean that the higher the smoothness is, the more blurred the image features will be, which will lead to errors in the process of object extraction, especially in texture detection. Gray linear transform is a common method to enlarge the range of image contrast difference. This method enlarges the range of image contrast difference in the way of keeping the pixel value and gray value unchanged. Therefore, it is necessary to adjust the smoothness degree of ε parameter reasonably according to the noise level and the detection of image target features. Image enhancement is usually a method of enhancing image quality to solve the problems such as blurring caused by the small difference of image gray value, small range of contrast difference and so on, which are caused by improper parameter setting [6]. The image is converted to HSV space, the lightness value of video image is calculated and counted, and then the lightness graph is formed. According to the global distribution of V-value, V-value can be divided into different ranges, and the number of ranges can be selected according to the needs of feature detection and calculation. The contrast difference range is enlarged to improve the edge contour feature of the processed image and increase the success rate and accuracy for the subsequent edge contour extraction.

2.3 Extraction of Contour Features of Full Information Mobile Monitoring Targets Based on Machine Vision

When the color of a moving object in an image is the same or similar to that of a moving object in the background, the correctness of the tracking may be affected by the number of iterations. Therefore, combining the hue and contour features of the target, the local texture of the background object with the same color as the target in the window can be further judged, thus improving the accuracy of target tracking under complex background. The convex polygon is the convex hull of an object, which consists of all the points in the set of points. A useful method of understanding the shape or contour of an object is to calculate the convex hull of the object, and judge the degree of defect by the size of the convex hull area and the contour area. Edge extraction is to set the region with obvious difference between the gray values of each pixel in the image as the boundary line. Because the foreground and background of the binary image can be distinguished by image segmentation, the difference between the gray value of the boundary region and the gray value of its adjacent region can be distinguished obviously, and the edge extraction becomes very easy. Object contour detection is such a process, specifically refers to a digital image containing the target and background, ignoring the impact of the target interior texture and background, using a specific algorithm to extract the target peripheral point set. Object recognition, object detection, object tracking and shape analysis are all inseparable from the basic contour detection. As a feature based on the surface of an object, which is not dependent on the change of color and illumination intensity, the contour features reflect the rough and smooth degree of the object. At the same time, in the image, the contour mainly reflects the local structural features of the image, specifically, some changes of pixel gray level or color within a certain region [7]. First, the gray difference of pixels is calculated according to the convolution template in the x and y directions, and then the gradient in the x and y directions is calculated according to the formula (5).

$$\begin{cases} W(\alpha, \beta) = \sqrt{G_1(\alpha, \beta)^2 + G_2(\alpha, \beta)^2} \\ \tau = -\arctan\left(\dfrac{G_1(\alpha, \beta)}{G_2(\alpha, \beta)}\right) \end{cases} \tag{5}$$

In formula (5), $W(\alpha, \beta)$ represents the gradient value, τ represents the gradient direction, $G_1(\alpha, \beta)$ and $G_2(\alpha, \beta)$ represent the gray difference in X and Y directions. Contour features are invariant in rotation, scale and noise, so the statistical or structural relationship between pixel values of local images is needed to be calculated for extracting contour features. The scale-invariant feature method decomposes the target image into squares. Finally, the feature points of the whole image are composed of the directions of these squares. Then the Hu moment feature matching is performed on the contour of the moving monitoring target. The essence of Hu moment is to construct several moment characteristics of normalized center moment by linear change, and it has good invariance of affine, translation and image transformation. The description effect of the same object in different directions is consistent [8]. If an image is described as a large rectangular box, it contains sixteen small rectangular boxes, or subregions. There are twenty-five characteristic points in each small rectangular box, through which the wavelet eigenvalues of small rectangular boxes can be obtained. Calculate the moments of the edge features obtained from the view and the template map, take their similarity, and then identify them. The specific process is as follows:

$$\begin{cases} \varpi = \sum_{7} \left| \dfrac{u_1 - u_2}{u_1} \right| \\ u_1 = sign(d_1)\lg|d_1| \\ u_2 = sign(d_2)\lg|d_2| \end{cases} \tag{6}$$

In formula (6), ϖ represents similarity, u_1 and u_2 are Hu moment parameters, d_1 represents template contour, and d_2 represents matching contour. For the center pixel, the positions of the other 8 peripheral pixels are uniformly distributed around the center pixel at the same distance, and their effects are uniform and equal. But when the center pixel is given the weight, if the starting point is at a different pixel position, then it will be encoded differently, and these different encodings have exactly the same meanings and effects. According to the above method, the target contour features of full information moving monitoring are identified, and the flow chart is shown in Fig. 2.

The target contour extracted from the template image is basically the same as the actual contour of the target object. This contour is obtained by manually simulating

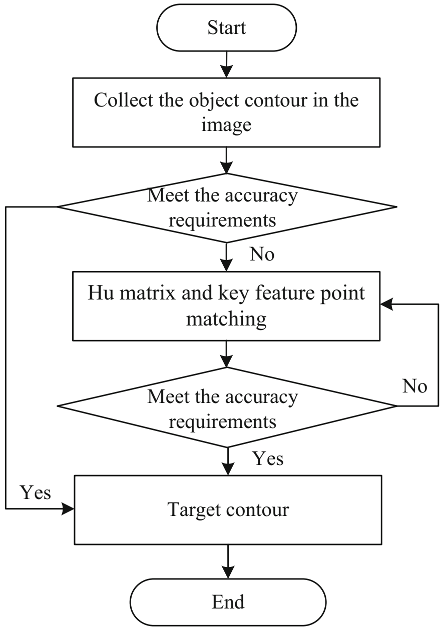

Fig. 2. Contour feature recognition process of full information mobile monitoring target

the similar environment. It can be compared with the object contour in the image to be recognized.

2.4 Establish the Abnormal Target Recognition Model of Mobile Monitoring

In video surveillance applications, surveillance cameras are usually placed in places where they can monitor important information, such as entrances, corridors, roads, etc., to monitor possible targets. However, the scene of video surveillance usually represents the moving range of the moving object in the surveillance area, and it is also the "visual" range of surveillance. Therefore, it is of great significance to analyze and study the monitoring scene area for the acquisition and analysis of moving objects. The normal video stream is divided into many segments in P frames, and then the convolution neural network self-encoder is trained with these segments. For the video stream waiting to be detected, the segment is divided into P frames, and the reconstruction error of P frames is used to measure the degree of anomaly. When the reconstruction error is large, the probability of the detected P frame containing abnormal behavior is considered to be high. In the full information mobile surveillance, it is difficult to analyze the scene and its content in the compressed domain, so in this paper, we first need to decode a small number of I frames, i. e. obtain the monitored scene model. Computing the background information of video requires decoding a small amount of I-frame weighting to obtain a relatively clean background information, and the background is processed accordingly. This process needs to continuously update the background as the video frame plays. Weighted gray histogram or color histogram is used to model the target, so that the pixels near the center of the target have greater weight, so that the edge noise and partial occlusion can be overcome. The specific background acquisition process is as follows:

$$V_t = \frac{Z_{t+c} + Z_{t+2c} + \cdots + Z_{t+mc}}{m} \tag{7}$$

In formula (7), V_t represents the background image after I frame superposition, Z represents the pixel value, t represents the time, c represents the interval between I frames in the video sequence, m represents the number of I frames, and the number of I frames required to obtain a background model does not exceed three. The behavioral trajectory of moving object is analyzed by segmenting the background region. Firstly, the histogram of luminance component Y is selected as the feature input for clustering segmentation. If the H value of the tracking target background is similar to the target, it is easy to drift.

Considering that the brightness is easily affected by illumination, while the saturation is not sensitive to illumination and is easy to distinguish between background and target, saturation S and hue H are introduced here to jointly count the color histogram of the target area. The number of clusters is set, and then the final clustering segmentation result is obtained by iterative operation according to the clustering rules of color histogram, and the classification result is used as the basis to distinguish active region and inactive region. The color distribution histograms of H component and s component of HSV color space are counted respectively, and then the back projection calculation is carried out to obtain the probability projection map. In the compressed domain, particle filter is used to track the target. For the tracked target, four tracking points are set, that is, the coordinates of the four corners of the tracking block diagram. The coordinates of the center point of the target can be calculated from the coordinates of the four points of the tracking frame. Images containing specific targets can produce very sparse and stable features through the deep convolution network with excellent classification performance. In video surveillance, since most pedestrians or other targets walk in a normal way or pass through the monitoring area, whether the track is an abnormal event is determined according to the information carried by the track center point, such as speed and direction. When judging the speed, use the following formula:

$$\eta = \sqrt{(B_N - B_{N-1})^2 + (A_N - A_{N-1})^2} \tag{8}$$

In formula (8), η represents the speed of tracking the target, (A_N, B_N) and (A_{N-1}, B_{N-1}) respectively represent the tracking point coordinates of the tracking target in frame i, N represents the number of frames the target spans. In this feature space, the changes of a small number of pixels and other ordinary changes in the input image will be ignored. Due to the pre training in the target classification training set, these convolution networks only respond to the structures that can represent the type of target and the structures of human interest.In order to detect the abnormal trajectory, this paper uses the velocity and direction information of the moving target as the input of the generalized regression neural network. For each moving target trajectory input, predict whether the current input trajectory is an abnormal event according to the historical information. If the trajectory is abnormal, it can be considered that its point velocity and direction change greatly, and then extract the abnormal target. So far, the design of abnormal target recognition method for full information mobile monitoring based on machine vision has been completed.

3 Experimental Study

3.1 Experimental Preparation

The moving targets in full information mobile surveillance video are mainly vehicles and pedestrians. In order to facilitate the experiment, this paper takes vehicles and pedestrians as the main objects. The algorithm framework in the experiment is implemented by tensorflow. The hardware environment is i7-5960x CPU, 64GB memory and NVIDIA Titan x graphics card. The running speed when detecting the test video is 20fpsa. This paper uses INRIA public pedestrian data set and vehiclereld vehicle data set as the training sample set. INRIA Pedestrian dataset (https://zhuanlan.zhihu.com/p/106216763)It is a video data containing pedestrians. The training set contains 614 positive samples (including 2416 pedestrians) and 1218 negative samples; The test set has 288 positive samples (including 1126 pedestrians) and 453 negative samples, which can be used for machine vision tasks such as pedestrian detection and recognition. Vehiclereld vehicle data set (https://zhuanlan.zhihu.com/p/106216763)It contains more than 50000 images of 776 vehicles, which were captured by 20 cameras and covered an area of 1.0 square kilometers in 24 h, which makes the dataset scalable enough for vehicle re ID and other related studies. The images are captured in the real-world unconstrained monitoring scene, and are marked with different attributes. The sample size is adjusted according to the scale characteristics of vehicles and pedestrians. Adjust the pedestrian sample size to 128x256 and the vehicle sample size to 256x256. The specific settings of the experimental data set in this paper are shown in Table 1.

Table 1. Experimental data set

Parameter	INRIA data set	VehicleReld data set
Number of positive samples	2358	2652
Number of negative samples	1164	1026
Sample size	128×256	256×256
Feature dimension	3685	1758

Adjust the size of the training negative samples of the two data sets for random cutting, and each negative sample generates five sample images with the same size as the pixels of the positive sample. The training sample set is used to train the model. The prediction network used is full convolution. The spatial resolution of the training data can be different from the test data. Therefore, the trained network can be applied to any resolution video.

This experiment measures the detection effect by calculating the accuracy and recall. The larger the ratio of accuracy and recall, the better the recognition effect of the proposed method.

Accuracy indicates the proportion of the detected abnormal targets in the total number, where *TP* indicates the actual abnormal data set, *TN* indicates the predicted abnormal data set, *FP* indicates the actual normal data set, and *FN* indicates the predicted normal data set:

$$Accutacy = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

Recall rate refers to the proportion of the number and total number of abnormal targets predicted and actually:

$$\mathrm{Re}call = \frac{TP}{TP + FN} \tag{10}$$

3.2 Results and Analysis

After the training of target recognition network, the performance of the designed full information mobile monitoring abnormal target recognition method based on machine vision is tested on the test data set. In order to measure the performance of the abnormal target recognition method designed in this paper, target recognition is carried out on the test samples, and the detection effect is measured by accuracy and recall. The test results of this method are compared with the abnormal target recognition methods based on inter frame difference and motion vector. In the test of INRIA data set and vehiclereld data set, the accuracy comparison results of different recognition methods are shown in Table 2–Table 3.

In the test of INRIA data set, the accuracy of the full information mobile monitoring abnormal target recognition method based on machine vision is 96.26%, which is 9.09% and 10.44% higher than the other two methods respectively.

In the test of vehiclereld data set, the accuracy of the full information mobile monitoring abnormal target recognition method based on machine vision is 92.83%, which is 9.26% and 10.04% higher than the abnormal target recognition methods based on inter frame difference and motion vector, respectively. From the perspective of detection accuracy, the recognition effect of abnormal targets in this paper is better than the two comparison schemes. In the tests of INRIA dataset and vehiclereld dataset, the comparison results of recall rates of different recognition methods are shown in Table 4–Table 5.

In the test of INRIA data set, the recall rate of the full information mobile monitoring abnormal target recognition method based on machine vision is 96.00%, which is 9.84% and 12.72% higher than the abnormal target recognition methods based on inter frame difference and motion vector, respectively.

In the test of vehiclereld data set, the recall rate of the full information mobile monitoring abnormal target recognition method based on machine vision is 93.29%, which is 10.92% and 10.19% higher than the abnormal target recognition methods based on inter frame difference and motion vctor, respectively. Based on the test results of accuracy and recall, the proposed method has good performance and can meet the recognition requirements of abnormal targets in full information mobile monitoring.

Table 2. Comparison of accuracy of INRIA data set (%)

Number of tests	Abnormal target recognition method of full information mobile monitoring based on machine vision	Abnormal target recognition method of full information mobile monitoring based on inter frame difference	Abnormal target recognition method of full information mobile monitoring based on motion vector
1	96.17	88.42	84.46
2	95.04	87.08	85.87
3	97.51	88.27	86.64
4	96.22	85.34	88.58
5	95.66	86.61	84.21
6	96.81	87.55	87.45
7	95.52	85.26	85.69
8	96.54	86.93	86.33
9	97.36	87.65	84.22
10	95.76	88.32	86.07
11	96.88	87.09	85.52
12	95.64	87.55	84.84

Table 3. Comparison of accuracy of vehiclereld data set (%)

Number of tests	Abnormal target recognition method of full information mobile monitoring based on machine vision	Abnormal target recognition method of full information mobile monitoring based on inter frame difference	Abnormal target recognition method of full information mobile monitoring based on motion vector
1	92.48	82.63	81.79
2	93.57	83.17	82.81
3	92.81	84.74	82.47
4	91.65	85.39	82.54
5	92.36	82.62	83.19
6	93.58	82.84	82.68
7	92.22	83.57	82.43
8	94.13	83.22	84.25

(*continued*)

Table 3. (*continued*)

Number of tests	Abnormal target recognition method of full information mobile monitoring based on machine vision	Abnormal target recognition method of full information mobile monitoring based on inter frame difference	Abnormal target recognition method of full information mobile monitoring based on motion vector
9	92.52	84.58	82.71
10	93.20	82.47	83.08
11	92.64	83.20	82.62
12	92.81	84.46	82.96

Table 4. Comparison of recall rate of INRIA dataset (%)

Number of tests	Abnormal target recognition method of full information mobile monitoring based on machine vision	Abnormal target recognition method of full information mobile monitoring based on inter frame difference	Abnormal target recognition method of full information mobile monitoring based on motion vector
1	94.49	86.17	82.51
2	95.54	85.59	82.06
3	96.62	84.85	82.59
4	97.38	86.66	83.38
5	96.16	87.32	84.51
6	94.17	88.24	82.14
7	95.02	85.57	82.27
8	96.58	86.83	83.63
9	97.54	85.64	84.05
10	97.96	86.37	84.48
11	95.37	85.26	84.62
12	95.22	85.45	83.11

Table 5. Comparison of recall rate of vehiclereld dataset (%)

Number of tests	Abnormal target recognition method of full information mobile monitoring based on machine vision	Abnormal target recognition method of full information mobile monitoring based on inter frame difference	Abnormal target recognition method of full information mobile monitoring based on motion vector
1	92.49	83.47	83.97
2	96.18	80.74	82.56
3	90.56	81.58	84.41
4	96.22	82.16	82.18
5	90.60	83.03	82.43
6	95.85	80.25	83.74
7	93.24	82.31	83.86
8	92.37	82.64	84.58
9	92.61	83.88	84.05
10	92.52	81.26	81.32
11	94.73	84.50	81.61
12	92.16	82.65	82.48

4 Conclusion

Video surveillance technology is widely used to maintain the security of public places and homes. The video surveillance system that relies on manpower has been increasingly unable to adapt to this era of big data. Therefore, integrated artificial intelligence and intelligent computer vision technology become the inevitable development direction of video surveillance in the future. Based on machine vision, this paper designs an abnormal target recognition method for full information mobile monitoring, which can improve the accuracy and recall rate, and has good detection performance. In this study, only THE HOG feature is used to describe the target, and then the target representation with multi-feature fusion can be considered to describe the target more fully to improve the target recognition effect. In this study, the research on target recognition is still insufficient, and the research on the recognition of human target pose is insufficient. In view of such problems, in-depth research is needed in future research.

References

1. Zhao, X., Ding, B., Xi, Y.: Weak supervised abnormal behavior detection using improved YOLOv3 under video surveillance. Opt. Technique **47**(1), 120–128 (2021)
2. Chen, Z., Yan, B., Zhang, S., et al.: Research of whole information alaeming and monitoring system based on mobile network. Laser J. **41**(10), 144–148 (2020)
3. Ma, Y.: Video moving target design of video image tracking and monitoring system based on target detection. Modern Electron. Technique **43**(11), 47–50+54 (2020)

4. Hu, Z., Zhang, Y., Li, S., et al.: End-to-end SSD real-time video surveillance anomaly detection and location. J. Yanshan Univ. **44**(05), 493–501 (2020)
5. Mo, H., Gan, J.: Network digital video image anomaly target detection area recommendation simulation. Comput. Simul. **37**(6), 396–400 (2020)
6. Hu, Z., Zhang, Y., Li, S., et al.: End-to-end SSD real-time video surveillance anomaly detection and location. J. Yanshan Univ. **44**(5), 493–501 (2020)
7. Dong, Y., Hu, G.: Detection and identification of the abnormal behavior in video surveillance systems. Mach. Des. Manuf. Eng. **49**(3), 66–70 (2020)
8. Ji, G., Xu, Z., Li, X., et al.: Progress on abnormal event detection technology in video surveillance. J. Nanjing Univ. Aeronaut. Astronaut. **52**(5), 685–694 (2020)