WILEY | Hindawi

Research Article

Research on the Detection and Tracking Algorithm of Moving Object in Image Based on Computer Vision Technology

Chunsheng Chen [ID] **and Din Li**

Anqing Normal University, Anqing 246000, China

Correspondence should be addressed to Chunsheng Chen; chenchunsheng@aqnu.edu.cn

In order to improve the video image processing technology, this paper presents a moving object detection and tracking algorithm based on computer vision technology. Firstly, the detection performance of the interframe difference method and the background difference model method is compared comprehensively from both theoretical and experimental aspects, and then the Robert edge detection operator is selected to carry out edge detection of the vehicle. The research results show that the algorithm proposed in this paper has the longest running time per frame when tracking a moving target, which is about 2.3 times that of the single frame running time of the CamShift algorithm. The algorithm has high running efficiency and can meet the requirements of real-time tracking of a foreground target. The algorithm has the highest tracking accuracy, the time consumption is reduced, and the error of the tracking frame deviating from the real position of the target is the least.

1. Introduction

Before computers and the appearance of modern science and technology, people tend to rely mainly on vision when accessing outside information. However, in our country, under the background of computer technology development promotion, the emergence of computer vision technology has made developments to further extend the human eye vision, especially under the help of all kinds of sensor technology. In this way, people can track moving targets in real time so as to accurately grasp the specific morphological attributes of the target [1]. Under the background of the continuous enhancement of China's economic strength, the process of socialist modernization is also gradually advancing, and computer information technology has made great progress, which makes a lot of reform and innovation in the related image engineering and artificial intelligence technology. The detection of moving targets in video images has always been an important problem in the field of computers. Its application to urban management is of vital significance in combating illegal crimes, maintaining people's safety, and promoting social stability and harmony [2]. To solve

many practical problems in real life, people need to be able to accurately capture the moving information. Therefore, it is very urgent to study the way and algorithm of moving target detection at present. Detecting moving targets is the most important content in video sequence images; to observe the whole scene image and find out the moving object involves many contents, such as computer image analysis, image processing, and artificial intelligence. To detect a moving target is too late for target tracking, and to understand behavior and the important basis of classification belongs to the bottom of the video monitoring system. To a large extent, it affects the function of the video surveillance system [3].

Object tracking is an important branch in the field of computer vision, and in order for computers to better understand the real world and carry out efficient human-computer interaction, research on the related technology of object tracking is the top priority. The main task of target tracking is to extract the target of interest from the video image and obtain its characteristic information such as position, size, moving speed, and angle in space continuously. At present, few algorithms can accurately track any scene in

real life, especially when the target is occluded and deformed, and the color features change; the tracking frame will more or less shift the real target position, or even completely lose track of the target. At present, visual target tracking-related technology has been integrated more closely with modern people's life, so that computers can assist or completely replace humans to efficiently complete many types tedious and dangerous work. For example, monitoring equipment in a complex traffic area, video monitoring in coal mine safety operations, and visual navigation technology in the military neighborhood are all typical applications of target tracking technology in real life. Sergio et al. [4]. investigated how visual context affects motion's glasses, whether consistent or independent of natural gravity. To do this, 28 subjects followed a computer-simulated trajectory that was disturbed in the descent section, its gravitational effect ($0\,g/2\,g$) or fixed natural motion ($1\,g$). Shortly after the disturbance ($550\,ms$), the target disappears for 450 or $650\,ms$ and becomes visible again until landing. Target movements occur in quasirealistic graphical cues such as balance sequences or unified backgrounds. We analyzed the saccade and pursuit motions after 0G and 2G target-motion perturbations, and we also analyzed the corresponding intervals for undisturbed 1G trajectories after the corresponding occlusion. Furthermore, we consider the target to reappear at eye distance to the target. Tracking parameters differ significantly between scenes: through a neutral background, eye movements do not correspond to target movements, while under a patterned background, they show a significant dependency, indicating better tracking of accelerated targets. These results indicate that the motor control is adjusted to the realistic attributes of the visual scene [5]. Sudha et al. [6] proposed an advanced deep learning method called "enhanced," a V3 and improved visual background extractor algorithm for detecting multiple types and multiple vehicles in the input video. More precisely, tracking is done using a combination of the Kalman filter algorithm and particle filter technology to find the trajectory of the incoming vehicle. In order to improve the tracking results, the technology named the multicarrier tracking algorithm is further proposed and tested with different weather conditions, such as sunshine, rain, night, and fog, with input video at 30 frames per second. Major research issues found in its department's recent literature are closely related to real-time traffic environment issues such as occlusion, camera oscillations, background changes, sensors, clutter, camouflage, and different lighting variations during the day—sunny and night vision. The experimental results were tested with ten different input videos and two benchmark datasets, KITTI and DETRAC. Up to eight advanced features have been considered for automatic feature extraction and annotation. The attributes are length, width, height, number of mirrors, and wheels as well as windshield shielding glass to detect the target area (vehicle) on the road. In addition, further experiments were carried out using a high-definition multi-input video with a monocular camera, with an average accuracy of 98.6%. The time complexity of the algorithm was 0, and the tracking result reached 96.6% [7].

Based on this, this paper proposes an algorithm for detecting and tracking moving objects in images based on computer vision technology. Firstly, the detection performance of the interframe difference method and that of the background difference model method are compared comprehensively from both theoretical and experimental aspects, and then the Robert edge detection operator is selected to carry out the edge detection of the vehicle. In the problem of vehicle target segmentation, the algorithm can clearly distinguish the interference between each target and accurately describe the moving face target, so as to track the moving target better. Combined with the edge detection of the gray vehicle image, the maximum interclass variance can be obtained, and the accuracy and real-time performance of segmentation can be improved. It is proven that the modified algorithm has a better tracking effect [8].

2. Materials and Methods

2.1. Theoretical Framework of Visual Computing. The theoretical framework of computer vision affects the development of computer vision and is a guiding ideology in the field of computer vision. Marr's vision theory is a view of information processing that combines physics, neurophysiology, and image processing. Under this theoretical framework, the generation of visual images can be divided into three stages:

(1) Two-dimensional schematic diagram: the initial schematic diagram uses edge segments, lines, spots, and endpoints to describe the brightness changes in the image, and then it uses virtual lines to completely and explicitly represent the geometric relationship. Finally, the initial schematic diagram whose description level can cover a certain scale range is obtained

(2) 2.5-dimensional diagram: by performing a series of manipulations on the initial diagram, a representation of the geometrical features of the visible surface is derived, including surface orientation, observer distance, discontinuity between orientation and distance, surface reflection, and some rough description of the predominant illumination

(3) 3D model: realize the representation of the three-dimensional structure of the observed object in the object-centered coordinate system and some description of the surface properties of the object, so as to obtain the spatial structure of the image

There are two main approaches to the visual tracking problem. One is a bottom-up approach, and the other is a top-down approach. The method in this paper can be divided into three stages by using the bottom-up method and combining with the visual computing theory of the Marr visual process: layer vision, middle-layer vision, and high-level vision. From the low-level vision to the middle-level vision is the image feature description, from the middle-level vision to the high-level vision is the 2.5D description, and from the high-level vision is the 3D

description. The bottom-up tracking process is obviously able to get the position, speed, and acceleration of the scene. Therefore, we first need to detect the moving target and then determine whether the target is a tracking target. Finally, the target position, trajectory, and other information are obtained. We are now going to track a moving vehicle, and with the help of Marr's theory, the visual tracking framework is easy to implement. There are mainly the following three stages: vehicle detection is in an early stage; target extraction and recognition to determine whether there is tracking is in the middle stage; obtaining the target position and trajectory information is in the later stage [9].

2.2. Video Image Moving Target Detection Method

2.2.1. Interframe Difference Method.
As one of the most commonly used methods for moving object detection, the interframe difference method can effectively detect dynamically changing images, and more than between adjacent frames or between three frames. Therefore, it is called the frame difference method, which is mainly based on the principle of strong correlation between adjacent frames of a sequence of images and can realize the change detection of moving targets. Under the effect of filtering, the range and area of the moving target are determined, and the calculation formula can be described as follows: $d_k(x, y) = |f_k(x, y) - f_{k-1}(x, y)|$. Among them, $f_k(x, y)$ and $f_{k-1}(x, y)$ represent two successive frames of moving images and $d_k(x, y)$ represents the absolute difference image; this calculation formula only includes subtraction calculation of pixel intensity, and the whole calculation process is relatively simple, feasible, and easy to implement. This detection method also has some defects. It is easily affected by noise interference. When the occluded background appears under the influence of motion, the new background will be misdetected as a moving object; in order to effectively avoid the influence of a "ghost" on moving targets, a new improvement was made to the frame search method. By verifying the difference intersection of multiple frames of images, moving targets were detected, and symmetric difference detection was commonly used, as shown in Figure 1 [10].

2.2.2. Background Difference Method.
The so-called background difference method mainly refers to subtracting the current frame image presented in the video sequence from the prestored background image to detect and extract the target. The specific process is shown in Figure 2. The background difference method can be described as follows: $d_k(x, y) = |f_k(x, y) - B(x, y)|$. Among them, the difference image is represented by $d_k(x, y)$, the background is represented by $B(x, y)$, and the current frame image is $f_k(x, y)$. When detecting the target, the K-frame image is subtracted from the background image, and then the difference image is extracted, and then the threshold is selected, The obtained difference image is transformed into a binary difference image, in which the pixel of 0 is determined as the background region, and the pixel of 1 is determined as the moving target region. As a very important part of the background difference method, preprocessing refers to the simple filtering of video images to avoid camera noise and instantaneous noise of external links. If the camera shakes, then, before background modeling, the acquired continuous video frames must be integrated into image processing [11].

2.3. Algorithm Analysis of Moving Objects in Video Images

2.3.1. Multiframe Image Average Method.
The so-called multiframe image average method mainly implies that when detecting and analyzing the moving target, it is taken as the noise source, and the accumulated average method is adopted to gradually eliminate the noise, and the sequence image of the running target is taken as the key analysis object. The background image can be expressed as follows:

$$B(x, y) = \frac{1}{N} \sum_{i=1}^{N} f_k(x, y). \tag{1}$$

Using this algorithm, the corresponding background image can be obtained, and the image is affected by the average frame number. The larger the average frame number is, the better the noise elimination effect will be.

2.3.2. Continuous Frame Difference Method.
The so-called continuous frame difference method mainly refers to the calculation of the difference between the current frame image and the previous frame image in the process of extracting and calculating the moving object of the video image, and determines the moving region based on it; it can keep the background in the moving region unchanged, while for the background in the nonmoving region, the current frame can be used to realize dynamic update. Under the action of iteration, the background can be extracted. First, the original image B_0 can be represented by the first frame image I_0; then, we set the iteration parameter as $I = 1$ and obtain the difference of the binary image between the current frame and the previous frame, namely, BW_1. The specific calculation method is as follows:

$$BW_i \begin{cases} 1 \, \text{ads}(I_k - I_{i-1}) \geq T, \\ 0 \, \text{ads}(I_k - I_{i-1}) < T, \end{cases} \tag{2}$$

where the images of the current frame and the previous frame are represented by I_i and I_{i-1}, respectively, the difference between frames is $\text{ads}(I_i - I_{i-1})$, the grayscale histogram of interframe difference images is represented by T, that is, the gray value corresponding to 1/10 of the right side of the maximum peak value. Then, the background is updated through a binary image, and the specific calculation method is as follows:

$$B_i(x, y) \begin{cases} B_{i-1}(x, y) BW_i(x, y) = 1 \\ aI_i + (I - a)B_{i-1}(x, y) BW_i(x, y) = 0 \end{cases}, \tag{3}$$

where $B_i(x, y)$ and $BW_i(x, y)$ represent the positions of the two images in the coordinates, and the updating speed coefficient is set as 0.1. Set the condition $I = I + 1$, and then substitute it with BW_i for iterative calculation; when a certain number is
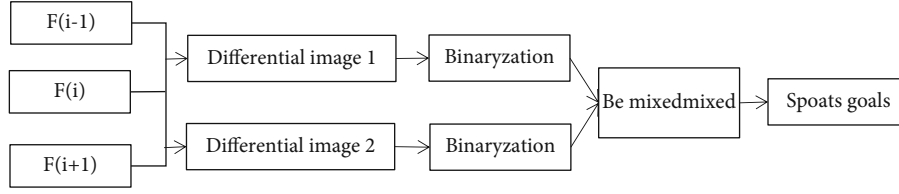
FIGURE 1: Symmetric difference method. By verifying the difference intersection of multiple frames of images, moving targets were detected, and symmetric difference detection was commonly used.
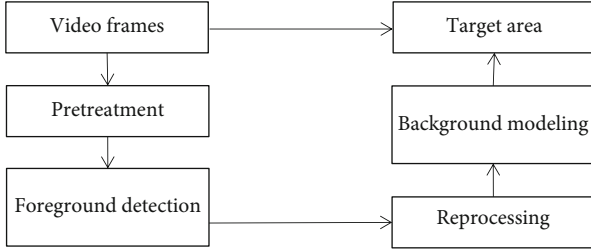

FIGURE 2: Detection flow. The so-called background difference method mainly refers to subtracting the current frame image presented in the video sequence from the prestored background image to detect and extract the target.

reached, the iteration ends and AA is set to extract the background.

2.3.3. Maximum Threshold Segmentation.
Usually, high-quality moving images are often needed in practical work. In order to meet the needs of high-quality moving objects in various fields, the threshold can be determined to segment the image, so that the binary image can accurately reflect the moving objects of the image. In general, the histogram of the image is in a bimodal state, and the optimal threshold is the valley of the bimodal histogram. But for multipeak histograms, the threshold true rule is faced with difficulties. After research, the probability distribution method can be used to represent the information. Generally, when the threshold is different, the amount of information in the target area and the background area is obviously different, so the total amount of information threshold T is the optimal segmentation value, that is, the optimal threshold.

2.4. Performance Analysis of Moving Target Segmentation Algorithm.
Image segmentation is an important method in computer moving object tracking. To solve the vehicle-tracking problem, the Robert edge detection operator is used to detect the edge of the target vehicle, which greatly improves the segmentation accuracy. In addition, we also propose a vehicle target segmentation strategy with maximum interclass variance, and the experimental results show that the algorithm has good segmentation effect. The OTSU method is a threshold segmentation method, which is based on the principle of discriminant analysis and least square method. In this method, the pixel threshold is divided into two parts, target C_0 and background C_1, and then the interclass variance is obtained:

Assume that the image gray value in a certain range is $\{0, 1, \cdots, l - 1\}$, the gray level I of pixel is expressed as n_i,

and the whole pixel is expressed as $N = \sum_{i=0}^{1-1} n_i$. The occurrence probability of gray level I is shown in formula (4).

Select the threshold T to distinguish between the target and the background area, $C_0 = \{0, 1, \cdots, t\}$; $C_1 = \{t + 1, t + 2, \cdots 1 - 1\}$. The probability of the possible occurrence of target C_0 and background C_1 is shown in formula (5).

The mean is shown in formula (6).

The average gray level of the image is shown in formula (7).

According to equation (5), equation (6), and equation (7), the interclass variance is calculated and shown in formula (8).

Next, we define the in-class variance as shown in formula (9).

The total variance is defined as shown in formula (10).

For second order statistics $0_W^2(t)$ and 0_r^2 independent of t, we made a simple judgment as shown in formula (11).

Under such a criterion, both types of t values belong to the optimal threshold, so $\eta(t)$ is the maximum criterion, as shown in formula (12).

2.5. Performance Analysis of Target Tracking Algorithm.
In many visual tracking algorithms, there are basically two categories: one is based on motion, and the other is based on a specific model. It is mainly based on the model tracking method, through mutual matching to achieve tracking. Template matching can be divided into two types based on target and target region. The target-based method matches by angle, color, etc. In a complex environment, its matching effect is better than the boundary matching method. In fact, due to the movement of the target itself, for the fixed target model that cannot be stable for a long time, it is necessary to update the target features in real time to adapt to the change of the target. If the target model of the current frame cannot accurately describe the current target, it will lead to wrong model update. To solve this problem, we proposed an image matching and tracking algorithm based on multiassociated templates. The algorithm flow is shown in Figure 3:

In the target tracking problem, the target tracking information is determined by the matching of the image relative to the original image. In fact, there is a certain degree of difference between the template involved in the image matching and the potential matching factors. Therefore, detecting a matching object on an unknown image is a complex task. The relationship between template T and the potential match object P is shown in formula (13).
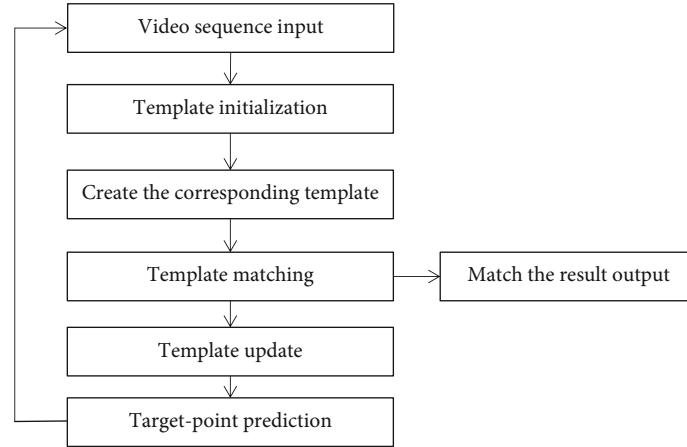
FIGURE 3: Tracking algorithm flow chart. To solve this problem, we proposed an image matching and tracking algorithm based on multiassociated templates.
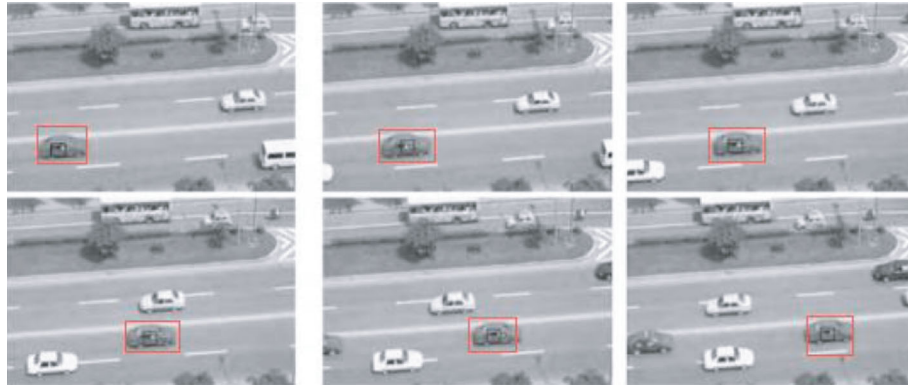

FIGURE 4: Tracking experiment results: tracking effect of black cars. In order to verify the accuracy of the algorithm, the experiment selected a PC configured with Intel core i5 2 GB memory.

Among them, $(x, y) \in T$, $(x', y') \in p$, and $\beta_{ij} \alpha_i$ are constants. By combining the similarity criterion, mean absolute difference method, and mean square error method, the similarity measure of the mean absolute error is obtained as shown in formula (14).

Here, the size of the reference image $f_1(x, y)$ is $m \times n$, the size of the real-time image $f_2(x, y)$ is also $m \times n$, and the mean square error similarity measure can be expressed as shown in formula (15).

In equation (14) and equation (15), the offset (x_0, y_0) that satisfies $D(x_0, y_0)$ is called the matching point, but when the target is heavily affected by illumination, the tracking effect will not be very ideal. The linear change of the image can be tracked by the normalization algorithm. The similarity measure of the normalized algorithm can be expressed as shown in formula (16).

Here, $\bar{f}_1$ and $\bar{f}_2$, respectively, correspond to the grayscale value of the template and the sample image [12].

3. Results and Discussion

In order to verify the accuracy of the algorithm, the experiment selected a PC configured with Intel core i5 and 2 GB

memory, and carried out experiments under two videos, respectively. The tracking results are shown in Figures 4 and 5.

Figure 6 is the tracking performance analysis of the above algorithm during the whole tracking period. As can be seen from the above table, the algorithm proposed in this paper has the longest running time of a single frame when tracking moving targets, which is about 2.3 times that of the CamShift algorithm. The algorithm has high running efficiency and can meet the requirements of real-time tracking of foreground targets. In terms of the tracking accuracy of the target, the algorithm in this paper has the highest tracking accuracy, reduced time consumption, and minimized error of the tracking box deviating from the real position of the target.

3.1. Advice on Equations.

$$p_i = \frac{n_i}{N},$$
$$p_i \geq 0,$$
$$\sum_{i=0}^{1-1} p_i = 1,$$

(4)

FIGURE 5: Tracking experiment results: tracking effect of white cars. In order to verify the accuracy of the algorithm, the experiment selected a PC configured with Intel core i5 2 GB memory.
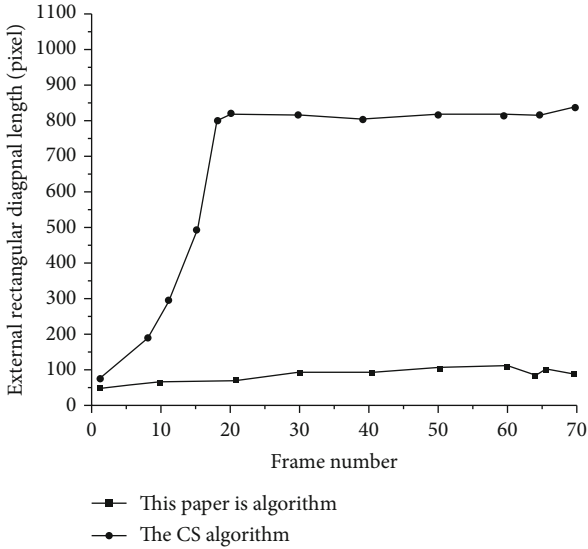

FIGURE 6: Dynamic changes of tracking box size under the complex background of two algorithms. The tracking performance analysis of the above algorithm during the whole tracking period is shown.

$$P_0(t) = \sum_{i=1}^{t} p_i,$$
$$P_1(t) = \sum_{i=t+1}^{1-1} P_i, \tag{5}$$

$$\mu_0(t) = \sum_{i=1}^{t} \frac{ip_i}{P_0(t)},$$
$$\mu_1(t) = \sum_{i=t+1}^{i=1} \frac{ip_i}{P_1(t)}, \tag{6}$$

$$\mu_T = \sum_{1=0}^{1-1} ip_i, \tag{7}$$

$$0_B^2(t) = P_0(t)[\mu_0(t) - \mu_T]^2 + P_1(t)[\mu_1(t) - \mu_T]^2, \tag{8}$$

$$0_w^2(t) = P_0(t) \sum_{i=0}^{t} \left[(i - U_0)^2 * \frac{P_i}{P_0(t)} \right]$$
$$+ P_1(t) \sum_{i=t+1}^{1-1} \left[(i - U_1)^2 * \frac{P_i}{P_1(t)} \right], \tag{9}$$

$$0_r^2 = 0_B^2 + 0_w^2, \tag{10}$$

$$\eta(t) = \frac{\sigma_B^2}{\sigma_T^2}, \tag{11}$$

$$\max_{0 < t < 1-1} \eta(t) = \max \eta(t^*), \tag{12}$$

$$\begin{bmatrix} x' \\ y' \end{bmatrix} \approx \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}, \tag{13}$$

$$D(x_0, y_0) = \frac{1}{mn} \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} |f_1(x, y) - f_2(x + x_0, y + y_0)|, \tag{14}$$

$$ID(x_0, y_0) = \frac{1}{mn} \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} |f_1(x, y) - f_2(x + x_0, y + y_0)|^2, \tag{15}$$

$$R(x_0, y_0) = \frac{\sum_{x=0}^{m-1} \sum_{y=0}^{n-1} [f_1(x, y) - \bar{f}_1][f_2(x + x_0, y + y_0) - \bar{f}_2]}{\sqrt{\sum_{x=0}^{m-1} \sum_{y=0}^{n-1} [f_1(x, y) - \bar{f}_1]^2} \sqrt{\sum_{x=0}^{m-1} \sum_{y=0}^{n-1} [f_2(x + x_0, y + y_0) - \bar{f}_2]^2}}. \tag{16}$$

3.2. Advice on Figures. Figures 1, 2, 3, 4, 5, and 6 are provided.

4. Conclusions

In recent years, the way to detect moving objects and the specific detection algorithm are two of the hot issues in the field of computers. It is an important practice of video image processing technology, and its application prospect is very broad. The detection and tracking algorithm of moving objects in images based on computer vision technology is studied. This algorithm improves the image processing

technology and establishes a new calculation method for vehicle tracking. In the problem of vehicle target segmentation, the interference between each target can be clearly distinguished, and the moving face target can be accurately described, so as to track the moving target better. Combined with the edge detection of a gray vehicle image, the maximum interclass variance can be obtained, the accuracy and real-time performance of segmentation can be improved, and the tracking effect can be made better.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

References

[1] Q. Ji, C. Dai, C. Hou, and X. Li, "Real-time embedded object detection and tracking system in Zynq SoC," *EURASIP Journal on Image and Video Processing*, vol. 2021, no. 1, 2021.

[2] C. Zhang, B. Liang, S. Ma, C. Xiang, Z. Bai, and J. Wang, "Research on the attribute extraction algorithm of people and vehicles based on video structure," *Journal of Physics: Conference Series*, vol 1802, no. 3, 2021.

[3] W. Zhang and W. Sun, "Research on small moving target detection algorithm based on complex scene," *Journal of Physics: Conference Series*, vol. 1738, 2021.

[4] W. Zhang, W. Sun, D. M. Sergio, L. Francesco, and B. Gianfranco, "Ocular tracking of occluded ballistic trajectories: Effects of visual context and of target law of motion," *Journal of Vision*, vol. 2019, no. 4, pp. 1–21, 2019.

[5] S. Javed, A. Mahmood, S. al-Maadeed, T. Bouwmans, and S. K. Jung, "Moving object detection in complex scene using spatio-temporal structured-sparse RPCA," *IEEE Transactions on Image Processing*, vol. 28, no. 2, pp. 1007–1022, 2019.

[6] D. Sudha and J. Priyadarshini, "An intelligent multiple vehicle detection and tracking using modified vibe algorithm and deep learning algorithm," *Soft Computing*, vol. 2020, no. 24, pp. 1–13, 2020.

[7] B. Zhang, L. Zhou, and J. Zhang, "A methodology for obtaining spatiotemporal information of the vehicles on bridges based on computer vision," *Computer-Aided Civil and Infrastructure Engineering*, vol. 34, no. 6, pp. 471–487, 2019.

[8] L. Ying, Z. Long, and N. Xin, "The technology of moving target detection and tracking based on sequential SAR image," *Journal of Physics: Conference Series*, vol. 1627, 2020.

[9] L. Liu, "Moving object detection technology of line dancing based on machine vision," *Mobile Information Systems*, vol. 2021, no. 2, 2021.

[10] X. Wang, B. Luo, and Z. Zhang, "Application of UAV target tracking based on computer vision," *Journal of Physics: Conference Series*, vol. 1881, no. 4, 2021.

[11] X. Wang, B. Luo, and Z. Zhang, "Application of UAV target tracking based on computer vision," *Journal of Physics: Conference Series*, vol. 1881, no. 4, p. 042053, 2021.

[12] Z. Yang, L. Fan, Y. Yang, Z. Yang, and G. Gui, "Generalized singular value thresholding operator based nonconvex low-rank and sparse decomposition for moving object detection," *Journal of the Franklin Institute*, vol. 356, no. 16, pp. 10138–10154, 2019.