Retraction Notice

The Editor-in-Chief and the publisher have retracted this article, which was submitted as part of a guest-edited special section. An investigation uncovered evidence of systematic manipulation of the publication process, including compromised peer review. The Editor and publisher no longer have confidence in the results and conclusions of the article.

KS, YY, and XZ either did not respond directly or could not be reached.

Real-time detection of panoramic multitargets based on machine vision and deep learning

Keyong Shen,[a,*] Yang Yang,[b,c] and Xiaoyu Zhang[b]
[a]Nanchang Institute of Technology, College of Computer Information and Engineering, Nanchang, Jiangxi, China
[b]Nanchang Institute of Technology, School of Electronics and Information, Nanchang, Jiangxi, China
[c]Bansomdejchaopraya Rajabhat University, Faculty of Education, Bangkok, Thailand

Retracted

Abstract. With the continuous development of computer vision technology, moving object detection technology has been paid enough attention and made great progress. Many new methods and new equipment have been developed. As an important part of computer vision, it has important applications in battlefield reconnaissance, video surveillance, image compression and retrieval, human–computer interaction and other research fields, moving object detection and tracking algorithm has always been a research hotspot. We mainly study the panoramic multitarget real-time detection based on machine vision and deep learning. By studying the principle of multitarget real-time detection based on machine vision and deep learning, the panoramic multitarget real-time detection model based on machine vision and deep learning is determined. The principle and correction effect of existing image distortion correction algorithm are analyzed, and the existing problems are summarized. Aiming at the problem that the existing image distortion correction effect is not good, a new method based on machine vision and deep learning is proposed. A real-time panoramic multitarget detection method based on degree learning is proposed. The experimental results show that when the target is moving at medium speed and slow speed, the success rate of tracking is 97% and 95%, respectively; the probability of successful target detection is 100% and 97%, respectively. Experimental results show that the improved method can solve the problem of particle degradation and improve the accuracy of real-time detection. © *2022 SPIE and IS&T* [DOI: 10.1117/1.JEI.31.5.051403]

Keywords: deep learning; image processing; machine vision; multitarget real-time detection.

Paper 210595SS received Sep. 22, 2021; accepted for publication Dec. 3, 2021; published online Jan. 11, 2022; retracted Jun. 24, 2023.

1 Introduction

Motion detection is often used in unattended surveillance video and automatic alarm. The images collected by the camera according to different frame rates will be calculated and compared by the CPU according to a certain algorithm. When the picture changes, if someone walks by and the lens is moved, the number obtained from the calculation and comparison results will exceed the threshold and indicate that the system can automatically make corresponding treatment. The research content of moving-target detection mainly includes four parts: image collection, image processing, image segmentation, and target tracking. The role of image processing is to process images to get better and more useful images for subsequent work. The function of image segmentation is to segment the image we are interested in from the image, that is, to segment the image into two parts, foreground and background, to prepare for tracking and subsequent target analysis and trajectory prediction. Therefore, accurate target detection is the basis for follow-up work and an important prerequisite for other important work. The moving-target tracking system studied in this paper can realize the automatic discovery and segmentation of moving targets in the panoramic field of view and control the pan-tilt to perform real-time tracking to obtain high-definition images of the target.[1] It can meet the requirements of completely no dead ends,

*Address all correspondence to Keyong Shen, jsj@nut.edu.cn

1017-9909/2022/$28.00 © 2022 SPIE and IS&T

all-weather time, clear shooting targets, and intelligent target tracking. This system can be applied to various important departments of monitoring and security, such as military departments, dangerous industrial departments, important commercial departments, etc.[2] By providing angle and spatial information of light, light field images are widely used in many applications. To capture the large field of view, light field images have been widely used in three-dimensional (3D) reconstruction, scene depth information extraction, fast multiview scene rendering, stereo vision matching, and so on. The existing methods either stitch a small field of view with a small aperture or use special equipment that is inaccessible to ordinary users. Zhao et al. proposed a method for extracting a full 360-deg field of view from a densely sampled handheld video. Their method can handle large exposure changes and motion blur issues, which are common in panoramic shooting and handheld video. To provide real-time rendering performance, their method applies light field sampling to the extracted panoramic light field. Based on unstructured representation, their method can sample the light field without explicit parameterization. They formulate the sampling problem as an ensemble multiple covering problem and use integer linear programming to solve it globally. The authors[3] method is not efficient and cannot meet the needs of practical applications. Chen et al. proposed a 360-deg situational awareness system. They gave the framework of the system and proposed key technologies, such as overall design and high-level design. They analyzed the throughput of the embedded image processing platform, real-time panorama stitching, target recognition, and threat estimation. Then, a solution was proposed. It has the ability to detect and threaten human and vehicle targets and can significantly improve the situational awareness of armored vehicles at close range. Their method cannot meet the detection purpose of multiple targets.[4]

Sun et al. proposed an image adaptive fast registration algorithm and an enhanced extended difference algorithm. Combining mathematical morphology, threshold segmentation, and global nearest-neighbor multitarget tracking algorithm, the functions of image background suppression, registration, suspicious target extraction, and multitarget tracking are realized. Their method is more complicated and cannot repair possible accidents in real time.[5] High dynamic range (HDR) images are usually obtained by capturing multiple images of a scene under different exposures. Popovic et al. designed a new multicamera platform that can construct and render HDR panoramic videos in real time with a resolution and frame rate of 25 fps. They used overlapping fields of view between cameras with different exposure levels to create HDR radiation patterns. They proposed a method for HDR frame reconstruction. Their method combined the previous HDR imaging technology with the panorama reconstruction algorithm. The field programmable gate array-based processing system developed can use the proposed method to reconstruct HDR frames and use a hardware-adaptive global operator to tone map the resulting image. Their method is more costly in practical applications.[6]

Based on the analysis of previous research results, this paper plans to complete the improvement of fisheye image distortion correction algorithm, multitarget detection and tracking algorithm, and video electronic image stabilization algorithm. This paper analyzes the principles and effects of common target detection and tracking algorithms and summarizes their advantages and disadvantages; for static and dynamic backgrounds, static background target detection and tracking algorithm based on selective matching of mean background and color histograms and optical flow aggregation are proposed. The dynamic background target detection and tracking algorithm with selective matching of categories and color histograms is tested through experiments.

2 Panoramic Multitarget Real-Time Detection Method

2.1 *Machine Vision*

Machine vision is a rapidly developing branch of artificial intelligence. In short, machine vision is to use machines instead of human eyes to measure and judge. The machine vision system converts the captured object into image signal through machine vision products (i.e., image pickup device, divided into CMOS and CCD) and transmits it to the special image processing system to obtain the shape information of the captured object, which is transformed into digital

Retracted

signal according to pixel distribution, brightness, color, and other information. The image system performs various operations on these signals to extract the characteristics of the target and then controls the on-site equipment action according to the discrimination results.

In a typical parallel axis model, the two cameras C_l and C_r only have a translation b in the x axis direction, that is, the optical centers of the left and right cameras are used as the origins of the left and right camera coordinate systems O_l and O_r. Take the connection direction of the optical centers as the direction of their x_c axis, and the distance between the optical centers is b, which is called the baseline. The optical axes of the two cameras are parallel, as their respective z_c axes. The y_c axis is perpendicular to the $x_c z_c$ plane, conforming to right-hand rule, and the focal length of both cameras is f.

Taking the left camera coordinate system as the world coordinate system, suppose the coordinates of the space point $P(x_w, y_w, z_w)$ in the C_l and C_r camera coordinate systems are $P_{cl}(x_{cl}, y_{cl}, z_{cl})$, $P_{cl}(x_{cr}, y_{cr}, z_{cr})$, and the projection points on the imaging plane are $P_l(x_l, y_l)$, $P_r(x_r, y_r)$, respectively, can be obtained from the triangular geometric relationship:

$$\frac{x_l}{x_{cl}} = \frac{y_l}{y_{cl}} = \frac{f}{z_{cl}}. \tag{1}$$

The left camera coordinate system coincides with the world coordinate system, and the x, y, and z coordinates are the same. The relationship between the points in the two coordinate systems is

$$x_w = x_{cl} = x_{cr} - b, \tag{2}$$

$$y_w = y_{cl} = y_{cr}, \tag{3}$$

$$z_w = z_{cl} = z_{cr}. \tag{4}$$

Then,

$$x_{cr} - x_{cl} = b. \tag{5}$$

The two transfer points are stored as flat and the y coordinates in the ^|^ standard system are the same. The x point coordinates are different, namely x_l and x_r. If $P_l(x_l, y_l)$ and $P_r(x_r, y_r)$ are known, then

$$x_{cr} = \frac{z_{cr} x_r}{f} = \frac{z_w x_r}{f}, \tag{6}$$

$$z_w = \frac{bf}{x_r - x_l}. \tag{7}$$

Let $x_r - x_l = d$, then

$$z_w = \frac{bf}{d}, \tag{8}$$

where d is the difference between the two image points relative to the coordinates of $P_l(x_l, y_l)$ and $P_r(x_r, y_r)$ in the image coordinate system, called disparity, that is, when the same 3D point is projected on two different camera images, the position of the corresponding point on the image difference; b is the distance between the optical centers of the left and right cameras, called the baseline; and f is the focal length of the camera. Therefore, to calculate the depth information of the obstacle, only the parallax d of the obstacle is obtained. To calculate the disparity, the premise is to find the same part of the left and right images, which involves the matching problem of image pairs.[7]

The most basic feature of the machine vision system is to improve the flexibility and automation of production. In some dangerous working environments that are not suitable for manual operation or where artificial vision has difficulty in meeting the requirements, machine vision is often used to replace artificial vision. At the same time, in the process of mass repetitive industrial production, the machine vision detection method can greatly improve the efficiency and automation of production.

2.2 *Error Analysis Method in the Algorithm Calculation Process*

There is a strict functional correspondence between the matching template and the corresponding pixels in the target area, which is the theoretical basis of the template matching algorithm based on gray-level correlation. Template matching is to find a subimage similar to the template image in another image according to the known template image. Template matching is one of the main research contents in graphics and image processing. When the change of the original image is not obvious, due to the neighborhood similarity of the processed image, the matching and tracking results of the gray-scale correlation algorithm generally will not produce errors.[8] However, in the target tracking process, the effects of reflections on the water surface, wide-angle distortion of the image, and noise interference during imaging will greatly affect the collected target information. Under the influence of noise, some pixels with their own background will be detected as moving areas in the detection results, or some areas in the moving target may be missed. The factors that affect the matching accuracy mainly include the following two aspects.[9]

The matching accuracy of the algorithm is affected on the one hand from the target template for matching, and the resulting error is called the template error. On the other hand, when inaccuracy comes from various external interferences, such as the noise generated when the camera collects the original image, an influence caused by the reflection of light on the surface of the pool, or imaging deviation caused by a problem in the vision subsystem hardware, etc., the resulting error is called random error.[10] To remove or weaken the noise in the image, the image can be smoothed, which is called image smoothing. The target matching template is very different from the target information in the image, which leads to template errors. This is a shortcoming of the algorithm itself, and the algorithm cannot automatically perform intelligent correction. If there is no corresponding error elimination measure, once the template error occurs, the template error will quickly become larger,[11] which greatly affects the subsequent target tracking. Therefore, measures must be taken to prevent or suppress template errors that may occur during the tracking process. The random error is different because there may be uncertain interference in a certain frame of image, which leads to unpredictable distortion of the target information of the frame, which causes errors in the matching process.[12] In general, the resulting deviation value is usually within a pixel range. In addition, the algorithm itself can weaken this influence by writing the algorithm properly. When the target is matched in the subsequent image frame sequence, the algorithm can still get the correct matching result. It can be seen that during tracking, whether accurate target tracking can be achieved depends on the degree of change of the image itself.

The reasons for the above two types of errors are different, and their respective effects are also quite different in the tracking process. However, when applied in reality, these two algorithms influence and restrict each other. Both of them may cause the deviation of the matching results in the tracking process and ultimately affect the performance of the system. Therefore, when suppressing errors, it is often required to comprehensively consider the effects of the above two types of errors and use targeted correction methods.[13]

2.3 *Convolutional Neural Network*

Convolutional neural network (CNN) model is a deep learning method from the aspect of reducing network weight parameters.[14] CNN has now become a research hotspot in the field of speech analysis and image recognition. Its weight-sharing network structure makes it more similar to a biological neural network, reducing the complexity of the network model. The several deep learning methods mentioned above, for the neurons in the hierarchy, the pixel matrix of the image is pulled into a one-dimensional vector, the form of processing,[15] which destroys the correlation of the adjacent pixels of the image, and the CNNs structure directly uses the image as input, especially in multidimensional images, which avoids the complex feature extraction in traditional recognition algorithms and data reconstruction process. In processing

Retracted

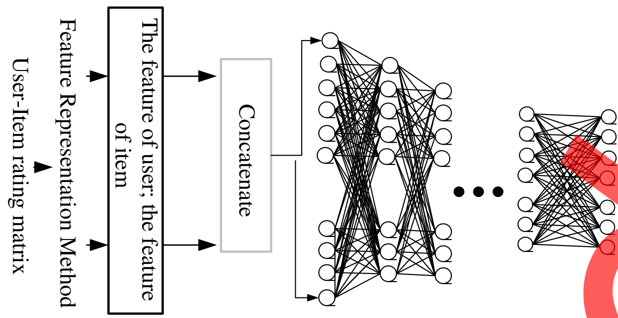

Fig. 1 CNN model.

two-dimensional (2D) images, this network structure has a high degree of invariance to translation, tilt, scaling, or other various deformations.[16]

As shown in Fig. 1, the CNN structure model mainly involves two layers: convolution layer and pooling layer. These two layers operate alternately in the network structure. CNNs contain one or more such layers. And the fully connected layer at the top. It is such a network structure that can utilize the 2D characteristics of image data, thereby exhibiting excellent performance. The input image is convolved for the first time with three trainable convolution kernels (corresponding to a short segment of the image) and trainable offset, and three feature maps are obtained in the C1 layer, and then the downsampling layer is performed, that is, the pooling operation. The size of pooling in this layer needs to be chosen by the designer. In most experiments, the size of 2×2 is used, so each group in the image obtained by C1 is a pixel summation, multiplied by a trainable parameter, plus a training bias is calculated by the sigmoid function to obtain the feature maps of the three S2 layers. As for the six feature maps of the C3 layer that appear in the above figure, they are derived from the corresponding combination of the feature maps of the S2 layer, and the operation is the same as the previous one. The resulting feature map is fully connected to a long vector and sent to the pattern classifier for classification.[17]

In fact, layer C is the feature extraction layer. Each layer may have multiple feature maps. Each feature map extracts a feature of the input through a convolution kernel. At the same time, each feature map has multiple neurons, each input of each neuron is connected with the local receptive field of the previous layer, and the local characteristics are extracted. Its characteristic is to strengthen the original signal characteristics. The S layer is a downsampling layer. Each feature map is a plane. This layer uses the principle of image local correlation to subabstract the image, which can reduce the amount of data processing while retaining useful information. The mapping from one plane to the next can be regarded as convolution, and the S layer can be regarded as a blur filter, which has the function of secondary feature extraction. The spatial resolution between the hidden layer and the hidden layer decreases, and the number of feature planes contained in each layer increases, which is conducive to detecting more feature information.[18]

Product neural network has long been one of the core algorithms in the field of image recognition. For character detection and character recognition/optical character reading, CNN is used to judge whether the input image contains characters and cut effective character fragments. Among them, the CNN directly classified by multiple normalized exponential functions is used for house number recognition of Google Street view image, and the CNN including conditional random field graph model can recognize words in the image.

2.4 *Mixed Gaussian Background Model*

The basic idea of the Gaussian mixture model is: for each pixel, define a state to represent the color it presents, and the K value is generally between 3 and 5 (depending on the computer memory and the speed requirements of the algorithm). Each of the K states is represented by a Gaussian distribution; some of these states represent the pixel values of the background, and the rest represent the pixel values of the moving foreground. Estimate the best matching state K through the observations at the current moment, and distinguish which states are foreground

Retracted

and which are background. The concept of background and foreground means that any meaningful moving object is the foreground under the assumption that the background is static. The basic idea of modeling is to extract the foreground from the current frame, and its purpose is to make the background closer to the background of the current video frame and use the result of this distinction to determine whether the best matching state love is foreground.[19] The parameter defining the Gaussian distribution corresponding to each state of the pixel is $\theta_k = \{u_k, \sum_k\}$, where u_k and $\sum_k$ are the mean and covariance matrix of the Gaussian distribution, respectively. Define the prior probability of each Gaussian distribution as $w_k = P(k)$, $k = l, 2, \ldots, K$, and $\sum_{k=1}^{k} w_k = 1$. Define $\Phi = \{w_1, \ldots w_k, \theta_1, \ldots \theta_k\}$ to represent the parameter set of K Gaussian distribution. If the color value of each pixel is represented by the variable x_t, its probability density function can be represented by the following K Gaussian distributions:

$$f_{x_t}(X_t|\Phi) = \sum_{k=1}^{k} P(k) f_{x_t|k}(x_t|k, \theta_k). \tag{9}$$

Among them, $f_{x_t|k}(X_t|\theta_k)$ represents the k'th Gaussian distribution function, which is defined as follows:

$$f_{x_t|k}(X_t|k, \theta_k) = \frac{1}{(2\pi)^{n/2}|\sum_k|^{1/2}} e^{-\frac{1}{2}(X-u_k)^T \sum_k^{-1}(X-u_k)}. \tag{10}$$

Given the current pixel value setting, estimate which of the K distributions is similar. The posterior probability $P(k|X_t, \Phi)$ of each state represents the observation likelihood generated by the pixel value X_t, which can be obtained by Bayesian equation, namely:

$$P(k|X_t, \Phi) = \frac{P(k) f_{x_t|k}(X_t|k, \theta_k)}{f_{x_t}(X_t|\Phi)}. \tag{11}$$

The current state can be estimated by the maximum posterior probability to get the current estimated state $\hat{k}$:

$$\hat{k} = \arg\ \max\ P(k|X_t, \Phi) = \arg\ \max\ P(k) f_{x_t|k}(X_t|k, \theta_k). \tag{12}$$

The Gaussian mixture model simulates an objective process, regardless of subjective foreground or background. That is to say, each Gaussian process in the mixture model may be either a foreground process or a background process.[20] Therefore, if the background is a dual model, there must be at least three Gaussian models in the mixed model, two of which simulate the background and the other simulate the foreground. If the mixture model is less than two, the Gaussian mixture model is equivalent to averaging over time to get the background and subtracting the background from the current frame to get the foreground. However, K is not as big as possible.

3 Design of Panoramic Multi-Target Real-Time Detection Experiment

3.1 *Video Capture, Transmission, and Display*

The system can realize video display in B/S mode. The steps to test it are as follows:

Read, write, and copy bootloader, kernel source code, file system, etc. to the ARM platform; transplant the Boa server program to the ARM platform; transplant the video capture program, image compression program, etc. to the ARM platform, and write the shell script program under the Linux system to migrate to the ARM platform allows these programs to run automatically after booting;[21] install the program that supports Java dynamic data exchange on the PC, this system uses the jre-1.5.0 version of the program; log in to the main interface of the video capture and monitoring, and then start the ARM platform. The video capture program, the real-time video of the monitoring site appears.

Click “turn on motion detection” under the page to turn on the moving object detection function, but because of the mutual calling of processes, an ARM collection point cannot run the moving object detection and video remote display functions at the same time. After the motion detection function is turned on, if a moving object appears in the monitoring area, the system will save the current frame as a picture, and the picture storage location is /root/motion/.

The test environment includes the main components of the system, such as IP cameras, gun-and-ball linkage camera equipment, client hosts, and personnel detection servers. All devices are interconnected through Ethernet switches.

3.2 *Function Test*

After adding and configuring the device in the device management interface, you can use the corresponding video device for video splicing. To test the video splicing function, the video stream is obtained in the laboratory scene to see the step-by-step effect of video splicing. In the video splicing experiment process, two video data sources are spliced to obtain the final spliced image.[22]

The main body of the back-shaped panoramic client interface is a video wall composed of multiple video panes, which can play and display multiple channels of video. The video splicing module in the multichannel splicing scheme includes two parts: preprocessing and splicing. In the preprocessing stage, the video data format conversion is performed first, and then the homography mapping matrix of the two pictures is calculated. The stitching stage is executed in the order of data format conversion, image mapping, and result image postprocessing (black border removal). The test content is mainly aimed at testing the efficiency of the video splicing module.

Test the time of the four parts of video data format conversion, image processing and result image postprocessing, and test the time required for the overall preprocessing module and splicing module. The test indicators are the average, maximum and minimum values of the single execution time of the test unit. Adopt cycle test, set the number of cycles and the average value of the time spent in the test. The maximum and minimum test execution time are executed a fixed number of times, the time of each execution of the test unit is recorded, and the maximum and minimum execution times of the test unit are counted.

When a moving object appears in the prison area, the gun and ball linkage device will recognize the moving object and send an alarm to the client. After the client receives the alarm, it captures the picture and sends a picture transmission request to the server. After the server receives the transmission request, it establishes a connection with the client for image transmission.[23] After the image transmission is completed, the server performs human target detection on the image and pushes the results to the client. The client prioritizes the detection results and displays them in the monitoring screen in real time. Among them, when an alarm event occurs on the client interface, the video splicing pane marks the location and credibility of the person in the corresponding panoramic screen.

The human target detection module is used to detect whether there are people in the video screen of the surveillance area. There can be single or multiple targets in the surveillance scene. The targets can present different categories, different postures, different orientations, different motion states, partial occlusions, etc. Aiming at different states of the target, multiple scenarios of different environments are simulated, mainly including different lighting conditions and different backgrounds, to increase the detection complexity and perform the test. The performance indicators of the human target detection module mainly include the accuracy of detection results and detection efficiency. Since the human target detection scheme in this paper uses video sequences as the detection samples, after the images are detected one by one, a certain strategy is used to determine whether the detection result is valid. Every 25 frames of images are used as a group, and the detection result of a single image is based on the sequence result. Make further judgments.

Retracted

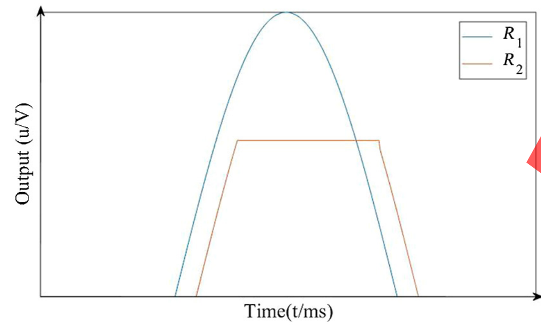

Fig. 2 The response of photodetectors to moving targets.

4 Real-Time Detection Method for Panoramic Multitarget

4.1 *Signal Acquisition and Processing Analysis of Moving Target Detection*

As shown in Fig. 2, in the detector signal detection process, when the target passes through the field of view of the photodiode, photodiodes work under the action of reverse voltage. When there is no light, the reverse current is extremely weak, which is called dark current; when there is light, the reverse current increases rapidly to tens of microamps, which is called photocurrent; the greater the intensity of light, the greater the reverse current. The change of light causes the change of photodiode current, which can convert the optical signal into electrical signal and become a photoelectric sensor. The obtained signal curve has a Gaussian distribution.

Among them, $R1$ is a signal curve with Gaussian distribution characteristics; $R2$ is a signal curve with overflow Gaussian distribution characteristics

According to the size of the target and the speed of the moving speed, the waveform of the moving-target motion signal obtained generally has two forms: one is a typical Gaussian-like distribution peak curve $R1$, the moving target passes through the sensor field of view at a constant speed, and there is no target to move the sensor. The phenomenon that the field of view is completely obscured. The other is the Gaussian distribution waveform with overflow, such as the signal curve $R2$, indicating that the target volume is large or the distance between the sensors is small, and the sensor field of view is completely obscured within a period of time.[24]

In this experiment, to analyze the characteristics of moving-target detection with Gaussian distribution signal curve, the signal acquisition of uniform moving target is realized through stepper motor control. Set the target movement speed to 2.5 cm/s through the program (when the speed setting parameter is 10, the actual speed is about 1.25 cm/s, the same below), change the distance D between the target and the sensor, and obtain the moving target at an equidistant speed through LabVIEW. The signal in the motion state is shown in Fig. 3. LabVIEW is a graphical programming language development environment. It is widely accepted by industry, academia, and research laboratories. It is regarded as a standard data acquisition and instrument

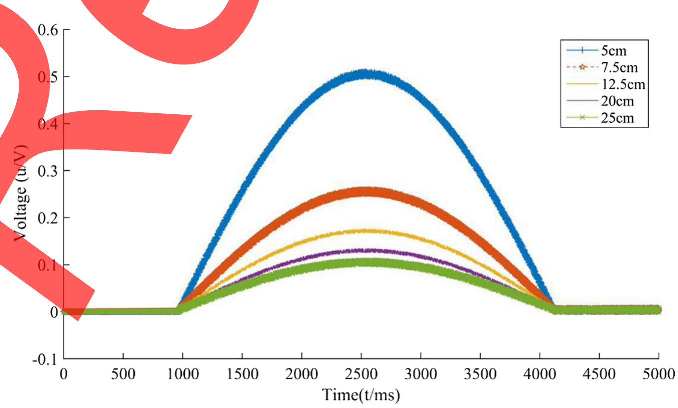

Fig. 3 Moving target signal under the same target at constant velocity and variable distance.

control software. The moving target signal detected by the photodiode is proportional to the target distance. The farther the distance is, the less obvious the characteristic signal of the moving target is collected; for the signal collection of the same target at the same speed, the detected signal peak value becomes smaller as the target is farther away. For the sensor array that detects the movement of the target, the change in the peak size of the Gaussian curve signal of the same moving target can be used as a judgment basis for detecting the movement of the moving target in the direction of the visual axis.

4.2 *Regularized Filter Recovery*

In the interface written based on MATLAB, open the picture to be processed and then click the regularized filter recovery option to call up the running interface. MATLAB is a mathematical software with strong computing and simulation ability, which can be used to deal with various data, statistical laws, and simulation of various physical models. Among them, long-range guidelines (LRG)-regularized restoration refers to the calculation result of adding the coefficient optimization range of the Lagrangian multiplier to the parameters. If different LRG parameters are selected, the restoration result will be affected. The peak signal-to-noise ratio is a standard for evaluating images, and its value is related to the mean square error of the original image and the result image. In general, the larger the value, the better the processing effect.[25] As shown in Fig. 4 and Table 1, the peak signal-to-noise ratios of the two result images and the original image are 31.322472 and 31.2947, respectively, and the mean square errors are 45.3321 and 48.2487, respectively. Although it can be proved from the experimental conclusion that the threshold can be selected as 10 or 100 for target discrimination, the simulation results cannot meet the requirements, so we no longer choose a fixed threshold but adopt a new method of dynamic

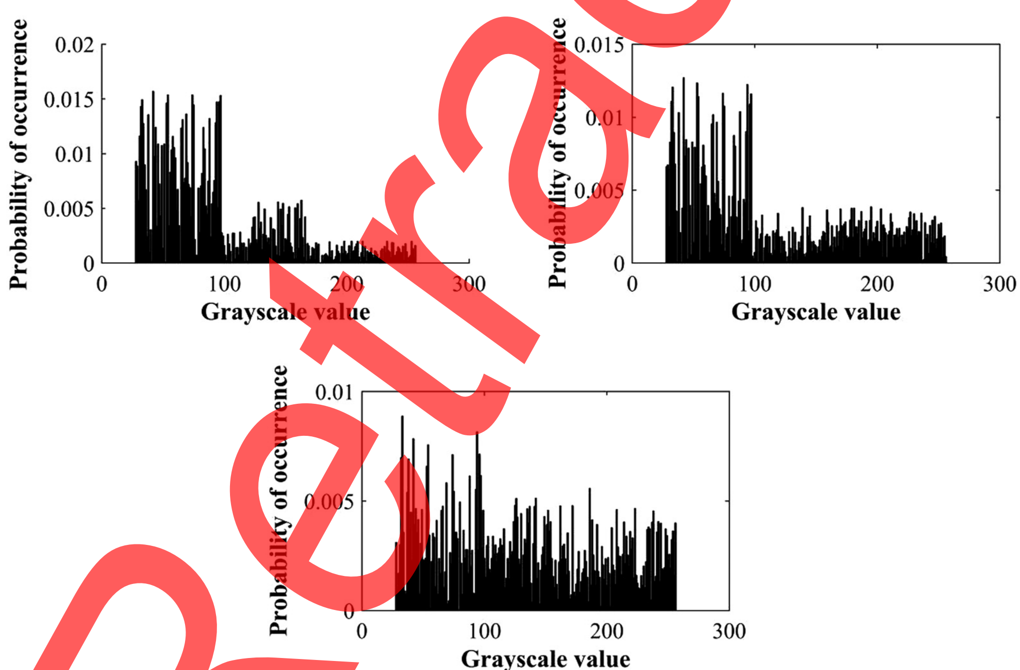

Fig. 4 LRG-regularized recovery histogram.

Table 1 Performance comparison of the two repair methods.

Performance	LRG-regularized recovery	Regularized filter recovery	LR iteration 5 times	LR iteration 15 times
Peak signal-to-noise ratio	31.322472	31.2947	31.4943	29.7599
Mean square error	45.3321	48.2487	46.0951	68.7205

threshold selection. Although the time spent in simulation has increased, the simulation effect is very good and basically meets the requirements of engineering practice. At the same time, this paper also uses the interframe difference method to detect moving targets. Through the analysis and comparison of the results, it is concluded that the background difference method can effectively eliminate the overlap of detection targets caused by the short sampling time. Background subtraction is a method to detect moving objects by comparing the current frame in the image sequence with the background reference model. Its performance depends on the background modeling technology used. Therefore, the simulation effect above is obviously due to the interframe difference method. Since the observation of moving targets is mostly carried out outdoors, sunlight exposure will cause shadows of the moving targets. If the shadow area is not effectively suppressed, the detection results will be biased and directly lead to the failure of tracking.

4.3 *Analysis of Motion Detection Success Rate*

When the wind and waves are very strong and have a great influence on the movement of the ship, although the surveillance system will not roll, it will perform random translation in space. At this time, a multitarget detection and tracking algorithm using optical flow clustering and selective matching of color histograms is used and an image stabilization algorithm is required.[26] To better verify the surveillance effect of the surveillance system, by simulating the movement of the ship, the camera is made to move in random directions in the x, y directions and xoy plane, respectively, where o is the imaging center, we can verify the surveillance effect of the surveillance system. As shown in Table 2 and Fig. 4, it is the success rate of the target detection effect when the test target moves at various speeds. It can be seen from this table that the probability of successful target detection is 100% and 97% when the target is moving slowly and at medium speed. The success rate of target detection is 93% when the target is moving fast, and there is a detection failure phenomenon. The analysis reason is that the target is moving fast, and the time in the scene panoramic field of view is too short. It has not had time to form stable sample information. The target has left the view. Therefore, the tracking fails. When the target is in the field for a long time, the target can be fully detected. In general, the target detection effect is very good, which can meet the requirements of follow-up tracking.

As shown in Table 3 and Fig. 5, when the target moves at various speeds, the test system PTZ tracking success rate data. In this experiment, the target moves in a circular arc and a straight line at various speeds. From the data in the table, it can be seen that when the target is moving at a

Table 2 Target motion detection success rate table at different speeds.

Target speed	Total number of experiments	Number of successes	Success rate (%)
Fast	15	14	93
Medium speed	30	29	97
Slow	20	20	100
Total	65	63	97

Table 3 Target tracking success rate table at different speeds.

Target speed	Total number of experiments	Number of successes	Success rate (%)
Fast	15	13	87
Medium speed	30	29	97
Slow	20	19	95
Total	65	61	94

Retracted

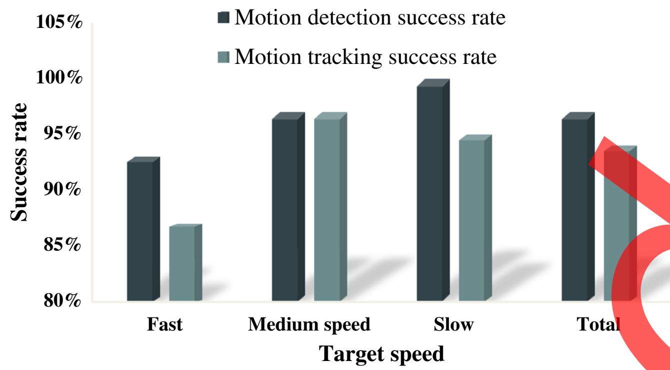

Fig. 5 Comparison of success rates.

Retracted

medium speed and a slow speed, the success rate of the gimbal tracking is 97% and 95%, respectively; when moving fast, the tracking rate is 87%. The analysis reason is that when the target is moving fast, the distance between the point and the point used by the trajectory prediction algorithm becomes larger, which leads to a larger deviation between the predicted point and the actual position of the target, and finally leads to the failure of the PTZ tracking. The overall performance is satisfactory, and the working performance is good. Only in individual cases such as the target suddenly turning, suddenly moving in the opposite direction, the target separation or the target speed is too fast, the success rate of the experiment is not ideal for the time being. It is easy to analyze the situation based on the theme of this article. The reasons for these undesirable phenomena, this article mainly studies the detection of moving targets. The tracking aspect mainly refers to controlling the tracking of the pan/tilt. The tracking target selected by the pan/tilt is based on the largest target detected. For the case of multiple targets, they can only be detected, but not identified. This is where we need to improve our future work.

5 Conclusions

In this paper, a moving-target localization algorithm based on ommatidium spatial region coding is established. Through computer simulation, the different trajectory movement of the simulated target in the field of view of the image detector is simulated; the simulation result shows that the spatial position code obtained based on the threshold value of the moving-target detection can realize the rapid estimation of the moving-target position within a small error range. The test results verify the feasibility of the method. Second, according to the motion signal characteristics of the moving target in the field of view of the photodiode and the hardware circuit design, the theory verifies that the hardware circuit positioning system based on the photodiode design has high real-time moving-target positioning capability.

An improved kernel density estimation target detection algorithm is proposed. Compared with other target detection algorithms, this algorithm can effectively detect moving targets while also meeting real-time requirements and has strong antibackground interference capabilities, which is more suitable for intelligent monitoring performance requirements. Improve the classic gray-scale interpolation algorithm, propose an edge-based interpolation algorithm, and integrate it into the image sphere restoration algorithm, successfully achieving image restoration, with high real-time performance, and initially achieving image quality improvement.

Image stitching, especially panoramic image stitching, has lower and lower requirements for images. It is no longer limited to the same angle or small angle stitching. It can realize the stitching of images from different multiangles and the corresponding feature point extraction. As well as image matching and image fusion, higher requirements are put forward.

Acknowledgments

This work was supported by China National College Student Innovation and Entrepreneurship Training Program-Upgrading and promotion of rail transit catenary smart wear detector based on the enterprise's internal entrepreneurial model (202012795001S).

Retracted

References

1. W. T. Lee et al., "High-resolution 360 video foveated stitching for real-time VR," *Comput. Graph. Forum* **36**(7), 115–123 (2017).
2. F. Pelissier et al., "Embedded multi-processor system-on-programmable chip for smart camera pose estimation using nonlinear optimization methods," *J. Real-Time Image Process.* **12**(4), 663–679 (2016).
3. Q. Zhao et al., "Panoramic light field from hand-held video and its sampling for real-time rendering," *IEEE Trans. Circuits Syst. Video Technol.* **30**(4), 1011–1021 (2020).
4. W. Chen et al., "Research on vehicle 360° situational awareness system," *J. Appl. Opt.* **39**(2), 165–168 (2018).
5. Q. Sun et al., "An adaptive real-time detection algorithm for dim and small photoelectric GSO debris," *Sensors* **19**(18), 4026 (2019).
6. V. Popovic et al., "Multi-camera platform for panoramic real-time HDR video construction and rendering," *J. Real Time Image Process.* **12**(4), 697–708 (2016).
7. J.-T. Ryu and D. Kwon, "An analysis of the surveillance image monitoring system using multi-image stitching," *Int. J. Imaging Rob.* **17**(3), 31–40 (2017).
8. R. Langseth et al., "An experimental evaluation of debayering algorithms on GPUs for recording panoramic video in real-time," *Int. J. Multimedia Data Eng. Manage.* **6**(3), 1–16 (2015).
9. L. Penaranda, L. Velho, and L. Sacht, "Real-time correction of panoramic images using hyperbolic Moebius transformations," *J. Real Time Image Process.* **15**(4), 725–738 (2018).
10. L. Yu, S. K. Ong, and A. Y. C. Nee, "A tracking solution for mobile augmented reality based on sensor-aided marker-less tracking and panoramic mapping," *Multimedia Tools Appl.* **75**(6), 3199–3220 (2016).
11. T. Rhee et al., "MR360: mixed reality rendering for 360° panoramic videos," *IEEE Trans. Vis. Comput. Graphics* **23**(4), 1379–1388 (2017).
12. T. H. Sun et al., "Automated thermal fuse inspection using machine vision and artificial neural networks," *J. Intell. Manuf.* **27**(3), 639–651 (2016).
13. H. Jiang, B. S. Duerstock, and J. P. Wachs, "A machine vision-based gestural interface for people with upper extremity physical impairments," *IEEE Trans. Syst. Man Cybern. Syst.* **44**(5), 630–641 (2014).
14. T. Andersson, M. J. Thurley, and J. E. Carlson, "A machine vision system for estimation of size distributions by weight of limestone particles," *Miner. Eng.* **25**(1), 38–46 (2012).
15. W. S. Qureshi et al., "Machine vision for counting fruit on mango tree canopies," *Precis. Agric.* **18**(3), 224–244 (2016).
16. C. Favret and J. M. Sieracki, "Machine vision automated species identification scaled towards production levels," *Syst. Entomol.* **41**(1), 133–143 (2016).
17. M. Ziatdinov, A. Maksov, and S. V. Kalinin, "Learning surface molecular structures via machine vision," *npj Comput. Mater.* **3**(1), 31 (2017).
18. F. Kurtulmuş, I. Alibas, and I. Kavdir, "Classification of pepper seeds using machine vision based on neural network," *Int. J. Agric. Biol. Eng.* **9**(1), 51–62 (2016).
19. Y. J. Chen, J. C. Tsai, and Y. C. Hsu, "A real-time surface inspection system for precision steel balls based on machine vision," *Meas. Technol.* **27**(7), 074010 (2016).
20. Z. Xiangyang et al., "Straight line detection algorithm in cigarette packet stamp inspection based on machine vision technology," *Tobacco Technol.* **49**(7), 105–109 (2008).
21. S. Cubero et al., "Automated systems based on machine vision for inspecting citrus fruits from the field to postharvest: a review," *Food Bioprocess Technol.* **9**(10), 1623–1639 (2016).
22. C. S. Nandi, B. Tudu, and C. Koley, "A machine vision technique for grading of harvested mangoes based on maturity and quality," *IEEE Sens. J.* **16**(16), 6387–6396 (2016).
23. L. N. Smith et al., "A 3D machine vision method for non-invasive assessment of respiratory function," *Int. J. Med. Rob. Comput. Assisted Surg.* **12**(2), 179–188 (2016).
24. W. Tang, L. Tian, and X. Zhao, "Research on displacement measurement of disk vibration based on machine vision technique," *Optik* **127**(8), 4173–4177 (2016).
25. H. Xu, B. Shi, and Q. Zhang, "A multitarget visual attention based algorithm on crack detection of industrial explosives," *Math. Probl. Eng.* **2018**, 8738316 (2018).

26. R. Wang et al., "Cascading classifier with discriminative multi-features for a specific 3D objects real-time detection," *Vis. Comput.* **35**(3), 399–414 (2019).

Keyong Shen is a professor at the College of Computer Information and Engineering, Nanchang Institute of Technology, Nanchang, Jiangxi, China. He received his master's degree from Huazhong University of Science and Technology, Wuhan, Hubei, China. His research areas include information processing and systems, computer applications, information security, and big data analysis.

Yang Yang is an associate professor at the School of Electronics and Information, Nanchang Institute of Technology, Nanchang Institute of Technology, Nanchang, Jiangxi, China. He received his master's degree from Nanchang University, Nanchang, Jiangxi, China. He is a PhD candidate at Bansomdejchaopraya Rajabhat University, Bangkok, Thailand. His research interests include image processing and recognition, control engineering, and intelligent systems.

Xiaoyu Zhang is currently a junior at the School of Electronics and Information, Nanchang Institute of Technology, Nanchang, Jiangxi, China. She participated in the related research works of image object recognition based on machine vision and deep learning under the guidance of professors.

Retracted