Real-Time Detection and Spatial Segmentation of Difference Image Motion Changes

KAI ZHOU¹, YINGPING HUANG², ENQING CHEN³, RUI YUAN²,4, AND ZHENDONG ZHANG²

¹Department of Physical Education, Henan Institute of Engineering, Zhengzhou 450001, China
²School of Physical Education (Main Campus), Zhengzhou University, Zhengzhou 450001, China
³School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China
⁴Department of Marine Sports, Pukyong National University, Busan 48513, South Korea

Corresponding author: Yingping Huang (zhzhdo@zzu.edu.cn)

ABSTRACT In this article, real-time detection of moving targets is performed, and an improved moving target detection algorithm combining background modeling and three-frame difference method is studied. An improved anisotropic differential filtering algorithm is proposed. The algorithm mainly compares the gradient difference between the target and the background in eight nearby directions. Choose the average of the three directions with the smallest spread function value to filter the space, which will effectively highlight the difference between the targets, so that the target signal is well preserved in the differential space domain; then based on obtaining the difference map. The average gray level of the target and the local signal-to-noise ratio of the spatial domain obtained by the energy enhancement algorithm combining the spatial and temporal motion characteristics have reached 278 and 14.84 dB, which further improves the target signal. Strict registration of heterogeneous spatial domains is completed by affine transformation and bilinear interpolation, and in many cases, the invariance and robustness of the method are verified. Focus on the proposed color space fusion method based on spatial frequency and fuzzy transformation. The superiority of the methods is evaluated and compared from both subjective and objective aspects. Finally, it lays a good foundation for moving target detection and better dealing with complex scenes. The improved inter-frame difference method combined with background subtraction validates and compares the method in different scenarios. It shows the advanced nature of the method in both qualitative and quantitative aspects. And effectively overcome the shortcomings of the particle filter tracking algorithm based on a single feature.

INDEX TERMS Differential space, motion change, real-time detection, space segmentation.

I. INTRODUCTION

The most important and intuitive way for humans to obtain information is through human vision. With the development of modern science and technology, computer vision has gradually been supplemented and improved and is in a leading position among functions that human vision cannot achieve [1]. Among them, video and airspace information are the main information carriers for computer vision [2]. Therefore, video-based moving target detection tracking technology has gradually become an important research direction in the field of computer vision [3]. As a hot research direction, moving target detection and tracking technology is an important branch of computer vision science [4]. According to the monitoring scene and the relative motion relationship of the camera, the detection and tracking of moving objects can be divided into the static background and dynamic background. In the above two cases, there are many theories and algorithms, and it is mature to detect and track moving targets in a static background and get ideal results in practical applications [5]. However, there are few existing algorithms for moving target detection and tracking in a dynamic background, and the adaptability of various algorithms is not strong. With the popularity of mobile camera equipment, moving target detection and tracking algorithms under dynamic background are urgently needed to be studied [6]. The application of the moving target detection and tracking algorithm in the military under the dynamic background includes military target photoelectric tracking and targeting system and battlefield situational awareness system; in the civil field, it includes...
airspace retrieval system, intelligent video surveillance [7], medical diagnosis and drone aerial photography surveillance [8]. In recent years, the wide application of moving target detection and tracking under dynamic background has made the moving target detection and tracking technology in the video shot under the motion of the camera become a hot research topic [9]. Because of this, it is imperative to study algorithms for moving target detection and tracking in a dynamic background. At the same time, both academic theory and social applications are of great significance [10].

The moving target detection algorithm is to separate the moving target in the video from the complex background area [11]. The moving target is also called the foreground. As an important basic algorithm of computer vision, the moving target detection algorithm has attracted the attention of many domestic and foreign scholars [12]. According to whether the camera is moving in the scene, the algorithm can be divided into a moving target detection algorithm under a fixed camera and a moving target detection algorithm under a moving camera [13]. The camera may miss important video frames. The research focus of this article is on the moving target detection algorithm under the fixed camera [14]. The classic moving target detection algorithm can be divided into inter-frame difference method, background difference method, and optical flow method [15]. The background difference method is to use the information of the previous frames to establish a background model for each pixel, and determine whether each pixel belongs to a moving target through the matching degree of the current frame pixel to be detected and its background model [16]. The algorithm has a good detection effect and real-time high sex has been widely used in practical engineering. Chandrasekar KS et al. proposed a sample consistency algorithm, which determines whether a pixel is a front sight or a background by performing sample consistency judgment on each pixel [17]. Most of the background difference methods proposed later. It draws on the idea of SACON algorithm sample consistency. The Visual Background Extractor (ViBe) algorithm proposed by Yin J et al. uses a random update and domain update strategy in the background model update [18]. According to the principle of domain similarity, each pixel updates its background model at the same time. The background model of domain pixels is selected for update, and the background model randomly selects background samples to update, ensuring that the life cycle of all background samples decays exponentially. Chen PH et al. improved the ViBe algorithm in 2011 [19]. According to the idea that the spatial distribution of pixels is like the time distribution, a single frame initialization background model is used [20]. To solve the ghost problem introduced by the ViBe algorithm single-frame initialization background model, Dong B et al. combined the three-frame difference method in the preprocessing stage to obtain the real background, thereby avoiding the ghost problem [21]. Shao J et al proposed a pixel-based adaptive segmentation (Pixel-Based Adaptive Segmenter, PBAS) algorithm in 2012, which automatically updates the decision threshold and background model update rate based on the detection results of the current frame to be detected [22]. It has certain robustness under the dynamic background. In 2013, the adaptive multi-resolution background extraction (AMBER) algorithm proposed by van Amerom JF et al. solved the problem of noise points [23]. By processing the full-resolution and zoom-resolution spatial domains, all the pixels with resolution airspace detection as the front sight and the pixels with zoom resolution airspace detection as the background point are judged to be noise points [24]. At the same time, the AMBER algorithm introduces the idea of background sample validity and selects the background sample with the lowest effectiveness during the background model update process compared with the random selection strategy, the validity of the background model is guaranteed by updating and replacing, but the calculation of the background sample validity is more complicated, which may cause the validity of multiple background samples to be equal to 0 at the same time [25].

As computers look smarter, people are increasingly expecting computers to have human-like perception and understanding capabilities [26]. This makes computer science and technology, especially artificial intelligence, one of the hottest and fastest-growing fields [27]. To allow computers to understand the world like humans to adapt to different tasks, researchers have designed corresponding algorithms based on humans’ perception of the outside world to give computers similar sensory and understanding capabilities [28]. Among them, the algorithm that allows the computer to obtain "vision" is very important for the computer to effectively perceive the outside world. Computer vision is an inseparable part of various application fields. Airspace is the foundation of vision. The first step of computer vision is to understand airspace [1]. If the computer cannot effectively understand the airspace, it will not be able to process the airspace. To understand airspace, airspace segmentation is at the most critical position. Airspace segmentation technology has a wide range of applications in the fields of scene understanding, manual analysis, automatic driving, medical diagnosis, and military engineering. Besides, airspace segmentation technology also plays an important role in other key technologies such as scene reconstruction and object recognition.

The moving target detection algorithm is to separate the moving target from the complex background. After research by many scholars in recent years, the moving target detection algorithm has made some progress, but there are still many interference factors in the complex actual scene that will affect the detection result of the moving target detection algorithm [29]. After continuous research by many scholars, we found that the detection accuracy and algorithm complexity of the moving target detection algorithm are improved. However, the current classic moving target detection algorithm is still difficult to solve the above problems at the same time or the detection results are not ideal. Therefore, while solving the above difficulties, it has important application
value for ensuring the detection accuracy of the algorithm. The main research goals of this article are moving target detection algorithm and shadow elimination algorithm under complex background, to provide accurate moving target information for subsequent target recognition and behavior analysis algorithms in intelligent video surveillance systems [30]. This article first studies several classic moving target detection algorithms analyzes their respective advantages and disadvantages through simulation experiments and theoretical basis, and selects the background detection method with high detection accuracy and good real-time performance as the main research algorithm in this article; several difficult problems encountered, such as dynamic background, ghosting problems, and intermittent target motion, etc., this article designs an adaptive model size background extractor (AMSBE) background difference method; to solve the actual scene, there are common shadow problems. In this article, the motion shadow detection algorithm that combines color and LFSP texture features is used to eliminate the shadow area. Finally, the AMSBE algorithm and the shadow elimination algorithm in this article are objectively evaluated through evaluation indicators. The difference method eliminates the effect of spatial domain segmentation on moving target detection, and the obtained target is relatively complete and clear. The multi-feature fusion method is used to model the tracking target, and the color feature and the texture feature are selected.

II. DIFFERENTIAL SPACE MOTION DETECTION IN REAL-TIME

A. AN IMPROVED DIFFERENTIAL SPACE DETECTION ALGORITHM

The core idea of the inter-frame difference method is to use the difference in pixel values between adjacent frames to distinguish the background from the moving target. In the adjacent frames of the video sequence, the moving target area has a more obvious pixel value change relative to the background area, so the moving target can be detected by the difference between the adjacent frames, as shown in Figure 1.

First, the differential space $D_s(x, y)$ is obtained by performing a differential operation on the gray space of the space $f_s(x, y)$ in the frame $t-1$ and the space $f_{s-1}(x, y)$ in the frame $s$, and then by determining the threshold $T$ The differential space $D_s(x, y)$ is binarized to obtain the moving target detection result $f_s(x, y)$ [31].

$$\begin{align*}
D_s(x, y) &= |f_s(x, y) - f_{s-1}(x, y)| \\
H_s(x, y) &= \begin{cases} 
0 & |D_s(x, y)| \leq T \\
1 & \text{otherwise}
\end{cases}
\end{align*}$$

In formula (2), 0 is a pixel point with a small change in the pixel value of the adjacent frame. The pixel point belongs to the background area, and 1 represents a pixel point with a large change in the pixel value of the adjacent frame, which belongs to the motion target area.

It can be seen from the detection results that the two-frame difference method selects a different decision threshold $T$ has a great influence on the detection result. When the decision threshold $T$ is small, the edge contour of the detected moving target is relatively complete, but the noise will be generated in the background area; when the decision threshold $T$ is larger, the generation of noise can be better suppressed, but the generated edge contour is incomplete [32]. By selecting a suitable decision threshold $T$, you can get a noise-free, complete detection of the edge contour, but there is still a certain gap compared with the ideal result of manual labeling because most of the moving target pixel values are distributed.
in blocks, such as cars. When the speed of the moving target is relatively slow, the position of the moving target in the adjacent frame will overlap, and the difference in the pixel value of the overlapping part is close to 0. It is detected as the background, which causes voids in the detection result; on the contrary, if the moving target speed is very large, there is no overlap of the moving target positions in adjacent frames. The two-frame difference method will detect two moving targets, one in the previous frame and the other in the current frame to be detected, resulting in a double shadow phenomenon.

The SACON algorithm initializes the background model through the first N frames of the video sequence, where the number of background samples of the background model is M. The background model of pixel m after initialization is shown in formula (3), where B is the pixel value of pixel m in the i-th frame, c represents the number of channels, and the number of grayscale spatial channels is 1 [33].

\[
\{ B_i^c(m) | i = 1, \ldots, M \} \tag{3}
\]

By performing pixel-level consistency judgment on the current frame to be detected and the background model, it is determined whether each pixel is a front sight or a background point. First, calculate the distance between the pixel value of the pixel m in the current frame to be detected and the M background samples in the background model of the pixel, and obtain the judgment result of each background sample.

\[
M^c_i(m) = \begin{cases} 
1 & |I^c_i(m) - B^c_i(m)| \leq T \\
0 & \text{otherwise} 
\end{cases} \tag{4}
\]

\[
l^c_i(m) = \begin{cases} 
1 & |I^c_i(m)| \leq T \\
0 & \text{otherwise} 
\end{cases} \tag{5}
\]

To ensure that the background model adapts to the changing actual scene, the background model update process needs to update the pixel information of the current frame to be detected into the background model after the change. The SACON algorithm background model update adopts a selective update strategy, by defining the foreground count matrix (Time Out Map, TOM) to select pixels to be updated.

\[
ToM^c_i(m) = \begin{cases} 
\text{TOM}_{i-1}(m) + 1 & |I^c_i(m)| \leq 1 \\
0 & \text{otherwise} 
\end{cases} \tag{6}
\]

The SACON algorithm can solve the ghost problem by counting the matrix TOM. However, if the fixed threshold t TOM is selected too large, the ghost exists for a long time, and if the fixed threshold TOM is selected small, the slow-moving target is updated into the background model to cause pollution. This article has made improvements based on the counting matrix TOM idea and adopted a probability update strategy based on the foreground counting matrix TOM, which can quickly and accurately solve the ghost problem.

The spatial filtering algorithm achieves the purpose of removing noise in the input spatial domain through a certain spatial transformation. It operates about the input airspace and recalculates the pixel values in the output airspace based on the pixel values in a certain neighborhood around the pixel values of the input airspace. If the size of the neighborhood is m \times n, m, n coefficients are required. These coefficients are arranged in a matrix called a filter, template, or kernel. Filtering templates are required for filtering. Different filtering templates can be used to implement different filtering algorithms.

The Gaussian mixture model takes Y historical gray values of pixels at (x, y) in the video sequence as input:

\[
\{ Y_1, Y_2, \ldots, Y_T \} = \{ I \{ x, y, \ i : 1 \leq t \leq T \} \} \tag{7}
\]

The pixel points in the background space are represented by a mixed probability distribution of H Gaussian models. Formula (8) can express the probability distribution model of pixel points:

\[
\rho (X_t | \beta_t) = \sum_{i=1}^{h} w_{i,t} \ast \phi (X_t | \beta_{i,t}) \tag{8}
\]

The mean shift method is a non-parametric method based on the density gradient rise. The algorithm calculates the offset vector value at this point and moves to the mean position of offset and iteratively converges to the target position. The mean drift method is defined as:

\[
M_k(x) = \frac{1}{k} \sum_{i=1}^{h} (x_i - x) \tag{9}
\]

In the formula: lk represents a spherical surface with radius h, which meets the set of y points in formula (10):

\[
l_k = \{ y : (y-x)^T (y-x) \leq k^2 \} \tag{10}
\]

The affine motion model not only satisfies these basic requirements but is also much simpler than the projection model. Without considering the difference in depth of field, the complexity of space segmentation, the effect of space segment estimation, and the real-time performance of the algorithm are combined [34]. In this article, a six-parameter affine transformation model is chosen to characterize space segmentation. The model can describe the translation, rotation, and small-scale zoom of the background. It is suitable for practical applications and has relatively high accuracy. In this article, the six-parameter affine model is expressed as follows:

\[
\begin{pmatrix}
x_h \\
y_h
\end{pmatrix} = \begin{pmatrix}
a & b \\
e & f
\end{pmatrix} \begin{pmatrix}
x_{h-1} \\
y_{h-1}
\end{pmatrix} \tag{11}
\]

It is further converted into the following linear equation system:

\[
\begin{align*}
x_h &= ax_{h-1} + by_{h-1} + e \\
y_h &= dx_{h-1} + fy_{h-1} + c
\end{align*} \tag{12}
\]

Use the established Gaussian pyramid in the airspace to complete the purpose of detecting moving targets in different scale-spaces. This method uses a Gaussian blur to obtain the scale space. The calculation is as follows:

\[
M(x, y, \partial) = H(x, y, \partial) \ast K(x, y) \tag{13}
\]
When the size of the two-dimensional template is H, it becomes:

\[
H(x, y, \partial) = \frac{1}{2\pi \sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \quad (14)
\]

\[
H(x, y, \partial) = \frac{1}{2\pi \sigma^2} e^{-\frac{((x-m)^2+(y-n)^2)}{2\sigma^2}} \quad (15)
\]

When calculating local extreme points in DOG scale space, it is not enough to compare each pixel with 8 adjacent pixels on the same scale. To ensure the accuracy of the calculation, it is usually necessary to compare with 2 \times 9 pixels on the adjacent scale. Only when the point value is greater than those compared pixels, the point is judged as a preliminary feature point and the next step of screening.

**B. DESIGN OF REAL-TIME DETECTION OF MOVEMENT CHANGES**

The actual environmental background of the video surveillance system application is constantly changing. There are light gradients or sudden rain and snow weather. The introduction of the background model update mechanism can increase the robustness of the algorithm. Therefore, the background model update has become an important part of various background difference algorithms. The background model update mechanism is to update the changed background pixel information into the background model to avoid detecting the new background as a moving target during the next detection process. The background model of each pixel in this algorithm can be updated, but the updated probability of the background model of each pixel is different. In the previous section, the method of the update rate of the adaptive background model is introduced in detail. When the background model of pixel m is determined to be updated, multiple backgrounds in the background model of the pixel should be selected for updating? Most scholars’ research results can be summarized into two strategies, first-in-first-out strategy, and random selection strategy, as shown in Figure 2.

Figure 2 is a first-in-first-out update strategy. The background samples with the longest survival time are selected from the N background samples of the pixel background model for updating. Figure 2 is a random selection update strategy. Both the ViBe algorithm and the SuBSENSE algorithm adopt this strategy. The random selection update strategy is to randomly select a background sample from N background samples to update, because the probability of each background sample being updated is the same, ensuring the smoothness of the life cycle of all background samples. However, neither of the above two strategies can guarantee that the selection of updated background samples is the most appropriate. This article uses an update strategy based on the effectiveness of background samples, as shown in Figure 2, where the area size of each background sample circle represents the background sample’s size. For the size of validity, the background sample with the lowest validity is selected for update. For example, the fourth background sample in the figure has the lowest validity, so the fourth background sample is selected for update.

The most important thing in the update strategy based on the validity of the background sample is to calculate the validity of the background sample. The size of the background sample validity represents the degree of reliability of the background sample in the background sample set. Given the constantly changing background area in the actual scene, the changed pixel information needs to be updated into the background model. When the background sample determined to be a background point is more effective, it is not conducive to quickly update the background model of the pixel in the changing area. The update rate of the adaptive background sample model in the previous section avoids updating the moving target information into the background model, so it is determined that the background samples of the former scenic spot are more effective than other background samples. In the algorithm, the background samples determined to be the former scenic spot also get a high weight [35]–[37], but the algorithm’s background sample validity calculation is too complicated, and there may be multiple background samples with the lowest validity at the same time.

In this article, the parameters that affect the improvement of the prediction effect of various anisotropic backgrounds are first evaluated for analysis. Then, 5 frames of different signal-to-noise ratio airspaces are selected from the two collected scenes, for experiment and comparative analysis of each the modeling effect of this background modeling method is to use a mid-infrared camera for spatial acquisition. The spectral response band of the camera is in the range of 3.58~4.77um, the size of each pixel is 20um, and the equivalent temperature difference of the camera noise is 20mK. The target size in scene A was 4 \times 4, and the target size in scene B was 9 \times 9. At the same time, to evaluate the background modeling effect of various background modeling methods on the airspace, especially for the non-stationary edge background modeling effect, the paper adheres to two principles to select different airspaces. On the one hand, select airspaces with different signal-to-noise ratios. Experiments to verify the adaptability of the improved algorithm in this article to different signal-to-noise ratio space domains. On the
other hand, the selected space domain contains not only more noise, but also areas with non-stationary edge contours. Whether it has good modeling ability when smooth edge contour area; finally, the enhancement algorithm proposed in the paper is used to make a comparative analysis of the airspace before and after enhancement.

III. SPATIAL SEGMENTATION ANALYSIS OF MOVING TARGETS

A. SPACE SEGMENTATION DESIGN

Airspace segmentation has gone through a long development process, and it has been continuously improved with the change of technology. Before deep learning was widely used, to complete the task of understanding the airspace, quite a few airspace segmentation methods were developed. During this period, disciplines, or knowledge such as digital airspace processing, topology, mathematics, etc. were all the principles or tools used for airspace segmentation, and many original methods of segmentation were born. Although, after the continuous development of deep learning and the continuous enhancement of computer hardware, the artificial field has brought great innovation, the feature-based spatial segmentation method is no longer sufficiently competitive. However, the development of any industry and field is a process of continuous accumulation, and in this process, the ideas of genius will never be outdated, and the motivation and entry point of any solution may inspire the latecomers. Only to understand the development process and direction of the technology, and the root cause of temporarily eliminating each technology. We can discover the shortcomings of the existing technology and the problems that can now be solved on this basis. The key is that many seemingly obsolete technologies often appear in new frameworks with new looks. And regain vitality, which is inspiring for this work. Therefore, although the task of this article is based on the study of deep space segmentation algorithms, this article will still explore the advantages and disadvantages of traditional space segmentation algorithms. These algorithms can also be applied to deep learning frameworks in some cases.

In addition to the threshold segmentation method, the edge detection method is also a common segmentation method. Many edge detection algorithms are implemented based on the gray-scale function derivation of the spatial domain and matching specific edge models in the spatial domain. The accuracy of the spatial segmentation results obtained by the edge detection algorithm greatly depends on the results of post-processing. In the face of complex airspace, edge detection is arguably the easiest way to think and use according to normal thinking. Space segmentation based on edge detection is also one of the most studied methods. The basis is that the junction is always accompanied by dramatic pixel color changes, as shown in Figure 3.

The advantage of the spatial segmentation algorithm based on edge detection is that the edge positioning is more accurate than other methods and the calculation speed is fast. The main problem it faces is that due to the limitations of the edge detection algorithm, many broken or scattered edges will have a greater impact on the results. Moreover, if the airspace is more complex, many fragmented edges are prone to appear. Their presence will cause serious errors in the detailed segmentation of the airspace. Due to the above difficulties, the spatial segmentation algorithm based on edge detection can usually only be performed by edge point detection, rather than spatial segmentation. This means that after acquiring the edge point information of the airspace through the airspace segmentation algorithm based on edge detection, the airspace still needs subsequent processing or the use of related algorithms to realize the real space segmentation.
This dilemma also means that research on edge selection and refinement is very important to improve the accuracy of spatial segmentation based on edge detection.

The threshold segmentation method is too simple, the division is too rough, and it is susceptible to noise to produce broken edges. Edge detection requires a lot of post-processing. These difficulties make the region-based spatial segmentation method begin to appear currently. The spatial segmentation method based on the region can be divided into region growing method and split and merge method. The area growth method starts from a certain point or certain points, expands itself continuously according to certain rules, and finally realizes a continuous segmentation effect. The quality of the rules for expanding itself is the key to determining the segmentation effect. The split and merge method is different from the growth method. It first divides airspace into many small blocks and achieves the purpose of airspace segmentation by combining small blocks into large categories.

In the region growth algorithm, its accuracy is determined by the starting point, the rules for expanding itself, and the rules for stopping expansion. At present, the research on the area growth algorithm mainly focuses on these three directions. The merging rules of the region splitting and merging algorithm enable it to effectively identify some tiny objects and deal with some messy airspace. However, because of its huge initial area and many merge rules, the algorithm is relatively complex, the calculation is too large, and the segmentation speed may be slow. Also, the process of splitting may lead to over-segmentation, that is, many segmentation areas have no practical significance, and additional constraints are required to obtain good segmentation results.

As the deep learning network gets deeper, a very important difficulty is that the deeper the deep learning network, the gradient will become smaller and smaller during the transmission process. When the network is deep to a certain degree, the gradient may even disappear. The worsening effect of the increasing error is a side effect of a continuous deepening of all deep networks, and it is also the reason deep networks cannot deepen and increase complexity indefinitely. In this process, the disappearance of the gradient will make training impossible. Res Net has made a lot of contributions to a deeper research of deep learning networks and has become another benchmarking work in the development of deep learning network frameworks.

B. STEP DESIGN AND EVALUATION ANALYSIS

The algorithm flow chart based on the number of adaptive background samples is shown in Figure 4.

The background model is initialized, using the first frame of space to initialize the background model for each pixel. Taking pixel m as an example, the initialization decision threshold $m$, the learning rate of the background model update rate and the foreground count matrix are in this pixel value $TOM(m)$, also, use the neighboring pixel information to calculate the number of initial background samples (N, establish M background samples for the pixel background model, and use the spatial consistency principle of pixels. Select the domain pixels to initialize the background samples, and finally, initialize the background sample validity $V$. For foreground and background points classification, the SACON algorithm is used to detect the front sight based on the idea of sample consistency, and the detection result of each pixel of the current frame to be detected is obtained. The pixel value
of the pixel m of the current frame to be detected is recorded as \(I(m)\), and the distance between \(I(m)\) and all background samples of the pixel is calculated. If the distance is less than the decision threshold \(T\), the number of background samples is greater than or equal to the threshold \(T\). The pixel point \(m\) is finally determined as the background point, otherwise, it is the front sight, and the validity of the background sample \(V\) at which the distance from \(I(m)\) is greater than the determination threshold \(T\) is increased by 1.

In the post-processing of the spatial domain, median detection is used to filter the detection result of step two to obtain the final binarized detection result. Morphological processing can be added when there is a hollow phenomenon in the detection result. The background sample is updated in the background model, and the background model update rate \(1m/s\) is used to determine whether the background model of each pixel needs to be updated. At the same time, to solve the ghost problem, the value of TOM\((m)\) is calculated according to the foreground count matrix. Whether a pixel is in a ghost area, when the pixel \(m\) satisfies any of the two conditions, the background sample with the smallest validity in the background sample set is selected for update. The background model parameters are updated, and the learning threshold of the update decision threshold \(T\) and the background model update rate is adaptively adjusted through a feedback system. Also, when the binary airspace detection result of the pixel \(S\) is in front for scenic spots, the foreground count matrix adds 1 to the pixel value TOM\((m)\), otherwise, TOM\((m)\) is re-initialized to 0.

The number of adaptive background samples and the number of fixed frames is recalculated by the adaptive background sample number algorithm based on pixel complexity. When the number of background samples decreases, the background samples are deleted from small to large according to the validity of the samples. When the background samples increase, new background samples are added and initialized.

To verify the effectiveness of the proposed moving target detection algorithm, choose to test on a server with multiple graphics cards. The hardware configuration of the server host and the software version number used in the test are listed in Table 1. The computer operating system installed on the server is Ubuntu 18.04 LTS [38].

| Hardware | model          |
|----------|----------------|
| CPU      | Xeon Xeon-E3   |
| GPU      | NVIDIA TitanXP 16G4 card |
| RAM      | 64G            |
| storage  | 1T SSD         |
| software | version        |
| TensorFlow | 1.12.1     |
| OpenCV   | 4.1.4          |
| Python   | 3.7.3          |

The experimental data mainly comes from the surveillance video taken on the spot. The shooting scene is a road intersection, and the detected moving targets are passing cars, trucks, and buses. A total of 4 surveillance videos from 2 locations were collected. The original surveillance video has a resolution of 1980 × 1080. Only the road part of the video contains moving targets. Therefore, after obtaining the original video, the areas with no moving targets in each surveillance video are cropped. Narrow the target search range can also reduce the amount of calculation. The video frame resolution after cropping becomes 875 × 815 (width 875 pixels, height 815 pixels), the video is split into consecutive video frames, and an average of 14485 video frames are split into each video segment.

The video material of the experiment was taken by a surveillance camera at a high place. The shooting time was in the morning. The vehicle target in the scene changed in color, shape, size, angle, lighting conditions, etc. Because the surveillance camera was shot at a high place, therefore, the size of the moving target to be detected is small. The moving objects in this video scene include pedestrians and people riding electric vehicles in addition to vehicles, but due to the problem of shooting angle, such moving objects are smaller in size, and effective features cannot be extracted in the feature extraction stage, so in the moving object detection stage actively filter out these moving objects whose pixel area is too small, and only detect large-sized moving vehicles. Besides, there are no other moving targets or dynamic backgrounds in the shooting scene, so the video can be used as experimental data to detect moving targets in the video. The video analysis module is the core part of the system in this article, and it is also the intelligent embodiment of the power grid video monitoring behavior analysis system. This system feeds back to the monitoring staff by understanding the staff behavior. To achieve an intelligent effect. However, the system’s understanding of work behavior is only superficial, and it can only be judged by the staff’s behavior, and cannot have other communication with the staff. However, for the monitoring staff, it greatly eases the pressure of their monitoring, can avoid real-time monitoring, can effectively help the monitoring personnel to complete the monitoring work. Alleviate the huge pressure of real-time monitoring by monitoring personnel.

IV. RESULTS ANALYSIS

A. ANALYSIS OF REAL-TIME MOTION DETECTION RESULTS

To use objective evaluation indicators to verify the effectiveness of the fusion color and LFSP texture feature shadow elimination algorithm, the experimental process uses the bungalows, cube, and people in shade videos in the dataset Cnet that contains artificially marked shadow areas. The above videos include indoor and outdoor scenes and scenes. There are static shadows and moving shadows, and the area and shape of the shadow area are irregular, which poses a serious challenge to the shadow elimination algorithm.

Figure 5 is the shadow detection result of the fused color and LFSP texture feature shadow elimination algorithm in this article. From left to right are the experimental results.
of data set people in shade frame 578, cubicle frame 5008, and bungalows frame 358. Figure 5 is the video input; Figure 5 is the result of manual labeling in the CDnet dataset. The gray area is the shaded area manually marked. The gray shaded area (gray value 250) and the white motion target area are used as candidate target areas during this experiment; Figure 5 is the shadow detection result obtained by the color feature in the HSV color space. From the detection result, the method based on the color feature detects the shadow area almost completely, but at the same time, some moving targets are also detected as shadows, and there is a false detection. Figure 5 is the shadow detection result obtained by using LBP and LFSP texture features. The detection result contains complete shadows, but because the texture features in people in shade and bungalows video scenes are not obvious, there is a large area of false detection in the moving target area. The shadow detection results based on the LFSP texture feature are greatly reduced in the cubicle scene. The experiment proves that the improved LFSP texture in this article reduces the errors generated in the moving target area compared with the LBP operator.

To illustrate the background modeling effect of the improved anisotropic algorithm in this article, five images with different signal-to-noise ratios in scenes A and B were selected as the experimental objects, and they were combined with TDLMS, adaptive Butterworth filtering, improved Top-Hat, improved Bilateral filtering, improved gradient reciprocal weighted filtering and anisotropic filtering for comparative analysis. At the same time, to evaluate the effect of background modeling of different algorithms, three evaluation indicators are used to evaluate the background of different algorithms, including mean square error (MSE), structural similarity (SSIM), and signal-to-noise ratio gain (GSNR). Five frames with different signal-to-noise ratio images in two scenes are shown in Figure 6.

The comparison of MSE, SSIM, and GSNR of different algorithms shows that the improved anisotropic background modeling method in scene A has the best effect, followed by the improved gradient reciprocal weighting, in which the MSE obtained by the improved algorithm in scene A is less than 9.8, SSIM is greater than 0.995, GSNR reaches 11.87, the MSE obtained by the improved algorithm in scenario B is less than 10.42, SSIM is greater than 0.961, and GSNR reaches 10.74. From the three indicators of the above two scenarios, it is seen that compared with the traditional Background modeling algorithm, this article improves the algorithm to achieve better modeling results. To intuitively show the modeling effects of different algorithms, an image with a signal-to-noise ratio of 12.48dB in scene A is taken as an example for analysis and description. The background modeling method of different algorithms and the corresponding difference map and three-dimensional map are shown in Figure 7.

It can be seen from the above figure that the background obtained by the TDLMS background modeling method is blurred, and the target in the difference map only retains part of the target signal, and the energy is relatively weak; the adaptive Butterworth filter background modeling method is more adaptable to the non-stationary background area poor,
the difference map contains more noise; the anisotropic background modeling method simply averages the diffusion function in the four directions in the neighborhood, which makes it impossible to remove the edge contour area in the difference map; the improved Top-Hat takes advantage of the combination of internal and external structural elements, makes full use of structural elements of different scales, and effectively performs background modeling, but it still has more noise in the difference map after filtering scenes with low SNR; based on the full use of the pixel spatial position and pixel gray similarity, the bilateral filter effectively preserves the target signal, but its difference map still retains some edge noise; improved gradient reciprocal weighted filtering can effectively model most of the background is modeled, but its effect on modeling isolated noise is poor, and this part of the isolated noise is retained in the difference map; while the background obtained by the improved anisotropic background modeling method in this article greatly preserves smooth background and edge contour area, effectively extract the target signal in the difference map, which is helpful to improve the target detection rate.

B. ANALYSIS OF ALGORITHM PERFORMANCE RESULTS

To further explore the performance of these algorithms, combined with the receiver operating curve (Receiver operating curve, ROC) to evaluate the performance of these algorithms in three scenarios. These algorithms include TDLMS, adaptive Butterworth filtering, anisotropy, improved Top-Hat, improved bilateral filtering, improved gradient reciprocal, pipeline filtering, and the algorithm proposed in this article. The ROC curves of the three scenes are shown in Figure 8. Pd represents the detection rate, and Pf represents the false alarm rate. NTDT is the number of detected real targets, and NFDT is the number of detected false alarm targets. NT is the total number of real targets in the image, and NP is the total number of targets detected in the image.

As can be seen from the ROC curves of the three scenarios in Figure 8, the detection effect obtained by the algorithm proposed in this section is better, the gradient is reciprocal second, and the anisotropy, improved Hop-Hat, and improved bilateral filtering are obtained by these three algorithms. The detection rate is better than the pipeline filtering algorithm, but because the pipeline filtering algorithm is based on binary image detection, the effect is better than the traditional detection algorithm. It can be seen from Figure 8 that when the false alarm rate Pf = 0.016, the detection rate of the algorithm proposed in this section is higher than 95.27%, while other algorithms are less than 85.58% in Figure 8, when the false alarm rate Pf = 0.034, the detection rate of the algorithm proposed in this section is higher than 95.25%, and under the same circumstances, the detection rate of improved gradient reciprocal, improved Hop-Hat, bilateral filtering and anisotropy are all higher than 80.24%, and the detection rate of other algorithms is Pd.

Both algorithms are lower than 73% in Figure 8, the detection rate obtained by the other eight detection algorithms is lower than the method proposed in this section. For example, when Pf = 0.0025, the detection rate of the algorithm proposed in this section is 88%, improving the gradient detection rate of reciprocal Pd is less than 85.54%, while other methods are less than 80%. It is also found that in the three scenes A, B, and C, as the signal-to-noise ratio decreases, the lower the detection rate obtained during the extraction process, the higher the false alarm rate. Experiments show that the algorithm proposed in this section achieves better detection results than other algorithms in low SNR scenes.

Comparing the processing results, the color image fusion method based on spatial frequency and blur transformation proposed in this article produces more natural colors. The resulting visual effect is better, the contrast is higher, more fully display scene information, and objectively evaluate the overall index better than other methods, as shown in Figure 9.

There are sufficient data in real life to prove that the existence of furniture and other objects has a spatial relationship with the boundary, and transfer learning is feasible, which is also the fundamental basis of network design in this article. In the network construction, the jump connection structure that has proved to be quite effective has also
been incorporated into the network of this article, as shown in Figure 10.

The framework learns a classifier that can fill the gap between 37 channels of semantic features and 5 types of specific room structure semantic tags. The jump connection structure organizes the network into the shape of a spatial pyramid, fusing multi-scale context information from deep to shallow, from thick to thin, and from local to global.

C. ANALYSIS OF SPATIAL SEGMENTATION RESULTS

Figure 11 shows the comparison results between the DMOSM algorithm and five traditional segmentation algorithms (EM, PCNN, X-DoG, FCM, and K-Means). From the comparison of the segmentation results, the PCNN and X-DoG algorithms have a poor segmentation effect on the material, and there is an obvious under-segmentation situation.

The two algorithms have insufficient segmentation of the circular impact crater, but these two algorithms achieve the perforated area. The EM algorithm is not good at segmenting circular impact craters. The algorithm over-divides circular impact craters. Compared with FCM, K-Means, and DMOSM algorithms, EM, PCNN, and X-DoG algorithms have obvious under-segmentation, that is, the ring-shaped impact crater is not obvious. From the above comparison, we can see that the DMOSM algorithm in this chapter has good segmentation performance in the circular impact crater formed by the impact damage and the perforated area formed in the impact center, with rich details and clear edges. Explained the advantages of multi-objective evolutionary optimization in this chapter.

At the same time, Figure 12 shows the E index data of the nine algorithms used in this article. Analyzed from the index, the DMOSM algorithm in this article has the smallest $E$ index 4.1839 in the airspace and has the best segmentation effect.

In this group of comparative experiments, the saliency map generator GS with an input image size of 128*128 was selected as the benchmark model using the data set MSRA10K all data labels for fully supervised training as the benchmark model, and the data with different labeling rate PI was used Set MSRA-10K training Saliency GAN containing 4 subnetworks, contains 2 subnetworks saliency graph generation adversarial network and contains only one subnetwork. A saliency map generator that only performs fully supervised learning. Figure 13 uses the included sub-network modules to represent different methods. The marking rate PI is sampled from 0% to 100%. The indicators with the best performance at each mark rate are marked in bold. As the labeling rate rises, the performance of saliency map generation and saliency map generation against the network rises rapidly. When the labeling rate is 20% and 40%, the image generation adversarial network in Saliency GAN can bring significant performance improvement, indicating that the image generation adversarial network can indeed learn effective image features. When the labeling rate exceeds 50%, Saliency GAN can already achieve comparable performance to the benchmark model trained with 100% labeling rate data.

When detecting moving targets in the video, the Gaussian mixture model is improved by combining the two-frame difference method. The improved model can extract the moving target foreground with more complete contours. For the
problem that occlusion is likely to occur between targets of different sizes in the moving target detection stage, the ratio of the contour area of the foreground of the moving target to the convex hull area is calculated to determine whether occlusion occurs between the targets. If an occlusion occurs, the candidate targets generated by the RPN network are used the frame locates the coordinate position of the occluded target. In the stage of recognizing moving targets, for the problems of weak manual feature extraction ability, complicated multi-target classification module, and difficulty in algorithm training convergence, it is proposed to use a convolutional neural network to complete multi-tasks of moving target feature extraction, classification and coordinate regression. Experiments show that the moving target detection algorithm can accurately and efficiently detect moving targets in the video under complex scenes where the size, shape, color, and lighting conditions of the target change, and the shooting angle are overhead, as shown in Figure 14.

V. CONCLUSION

With the development of science and technology, computer vision has also received a lot of attention. Moving object detection and tracking plays an important role in the field of computer vision. Detecting and tracking moving objects in a dynamic background is a hot spot that has attracted many people’s research interest. In practical applications, the tracking environment is complex and changeable, such as changes in lighting conditions, target deformation, target occlusion, and noise interference. Therefore, it is difficult to solve the problem in the field to study algorithms that can track targets stably under complex conditions. The detection and tracking algorithm proposed in this article is to detect the target area by spatial domain segmentation graphic compensation difference method, extract its features and input it into the multi-feature fusion particle filter algorithm to achieve continuous tracking of moving targets. Without knowing the position and size of the target, it can directly detect and track the moving target. In the case of space segmentation, a differential algorithm with space segmentation compensation is used to detect moving targets in a dynamic background. The detection experiment of single and multiple moving targets under dynamic background shows that the direct inter-frame difference method misinterprets the background as a moving target. The detection results contain a lot of noise, which is not suitable for moving target detection under dynamic background. The weight of each feature can be adaptively assigned according to the similarity between the target template and the candidate template. Through multiple experiments, the improved target tracking algorithm effectively overcomes the effects of background lighting changes, target occlusion, target color similar to the background or other interfering colors on target tracking reliability in video image sequences, and make target tracking anti-occlusion ability, adaptability to background movements and complex changes, its performance is better. And effectively overcome the shortcomings of the particle filter tracking algorithm based on a single feature.

REFERENCES

[1] Y. Jiang, J. Wang, Y. Liang, and J. Xia, “Combining static and dynamic features for real-time moving pedestrian detection,” Multimedia Tools Appl., vol. 78, no. 3, pp. 3781–3795, Feb. 2019.
[2] F.-C. Gheseu, B. Georgescu, Y. Zheng, S. Gribic, A. Maier, J. Hornegger, and D. Comaniciu, “Multi-scale deep reinforcement learning for real-time 3D-landmark detection in CT scans,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 41, no. 1, pp. 176–189, Jan. 2019.
[3] M. Capogrosso, F. B. Wagner, J. Gandar, E. M. Moraud, N. Wenger, T. Milekovic, P. Shkorbatova, N. Pavlova, P. Musienko, E. Bezard, J. Bloch, and G. Courtine, “Configuration of electrical spinal cord stimulation through real-time processing of gait kinematics,” Nature Protocols, vol. 13, no. 9, pp. 2031–2061, Sep. 2018.
[4] K. Zhao, J. Deng, and D. Cheng, “Real-time moving pedestrian detection using contour features,” Multimedia Tools Appl., vol. 77, no. 23, pp. 30891–30910, Dec. 2018.
[5] R. Weitschat and H. Aschemann, “Safe and efficient human–robot collaboration part II: Optimal generalized human-in-the-loop real-time motion generation,” IEEE Robot. Autom. Lett., vol. 3, no. 4, pp. 3781–3788, Oct. 2018.
[6] Z. Hou, Y. Dong, J. Xiang, X. Li, and B. Yang, “A real-time QRS detection method based on phase portraits and box-scoring calculation,” IEEE Sensors J., vol. 18, no. 9, pp. 3694–6702, Mar. 2018.
[7] T. Long, T. Zeng, C. Hu, X. Dong, L. Chen, Q. Liu, Y. Xie, Z. Ding, Y. Li, Y. Wang, and Y. Yang, “High-resolution radar real-time signal and information processing,” China Commun., vol. 16, no. 2, pp. 105–133, Mar. 2019.
[8] A. J. Alzahrani and I. Khan, “Robust approach of edge detection in videos using spatial-temporal features,” Int. J. Comput. Sci. Issues, vol. 15, no. 6, pp. 42–48, Nov. 2018.
[9] G. Maulucci, F. Di Giacinto, C. De Angelis, O. Cohen, B. Daniel, C. Ferreri, M. De Spirito, and S. Sasson, “Real time quantitative analysis of lipid storage and lipolysis pathways by confocal spectral imaging of intracellular micropolarity,” Biochimica et Biophysica Acta (BBA)-Mol. Cell Biol. Lipids, vol. 1863, no. 7, pp. 783–793, Jul. 2018.
**RUI YUAN** was born in Henan, China, in 1993. She received the bachelor’s and master’s degrees from Zhengzhou University, in 2015 and 2018, respectively. She is currently pursuing the Ph.D. degree with Pukyong National University, South Korea. She has published a total of three articles. Her research interests include sports training and sports psychology.

**ZHENDONG ZHANG** was born in Beijing, China, in 1961. He received the bachelor’s degree from Wuhan Sports University, in 1983. He currently works with Zhengzhou University. He has published a total of 12 articles on CSSCI. He hosts a general project of the China National Social Science Foundation. His research interests include sports training and social sports.