IEEEAccess

Multidisciplinary : Rapid Review : Open Access Journal

Moving Target Detection Based on Improved Gaussian Mixture Background Subtraction in Video Images

JUNHUI ZUO[1], **ZHENHONG JIA**[1], **JIE YANG**[2], **AND NIKOLA KASABOV**[3], **(Fellow, IEEE)**

[1]College of Information Science and Engineering, Xinjiang University, Ürümqi 830046, China
[2]Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai 200240, China
[3]Knowledge Engineering and Discovery Research Institute, Auckland University of Technology, Auckland 1020, New Zealand

Corresponding author: Zhenhong Jia (jzhh@xju.edu.cn)

ABSTRACT In recent years, background subtraction techniques have been used in vision and image applications for moving target detection. However, most methods cannot provide fine results due to dynamic backgrounds, noise, etc. The Gaussian mixture model (GMM) is a background modeling method commonly used in moving target detection. The traditional GMM method is vulnerable to noise interference, especially from dynamic backgrounds; thus, its detection performance is not good. Because of the influence of background noise and dynamic effects on moving target detection, we propose a method of moving target detection for dynamic backgrounds based on improved GMM background subtraction. This method can be divided into three stages. First, in the background modeling stage, to facilitate calculation and improve modeling speed, the video frame is blocked, and the background model is reconstructed using the image block averaging method. Second, in the moving target detection stage, the method of combining wavelet semi-threshold function denoising with mathematical morphology closed operation is used for denoising, which effectively eliminates the influence of noise and improves the detection effect. Third, in the background updating stage, the adaptive background updating method is used to update the background to improve detection results. The simulation results show that the improved method can reduce noise and dynamic background interference while improving moving target detection, thereby proving the effectiveness and adaptability of the proposed method.

INDEX TERMS Gaussian mixture model, moving target detection, dynamic background, mathematical morphology, adaptive background updating.

I. INTRODUCTION

In recent years, with the continuous development of computer vision technology. In the computer vision field, moving target detection based on video sequence images is an important research topic. Moving target detection is also the basis of target tracking [1], [2] and behavior understanding [3], [4]. It is widely used for many tasks, such as image processing and intelligent video surveillance. Recently, researchers have proposed many methods for detecting moving targets. A detection method is selected according to the detection scenario. Currently, the main methods of moving target

The associate editor coordinating the review of this manuscript and approving it for publication was Xiao-Yu Zhang.

detection are the optical flow method [5], [6], the interframe difference method [7], [8] and the background subtraction method [9], [10]. The optical flow method is based on the brightness information of the target image. This method has high computational complexity and weak anti-jamming ability; thus, it is not often used. The interframe difference method uses a continuous video frame image to perform a differential operation to extract a moving target. It is highly adaptable to background changes. However, moving target detection by this method produces a cavity phenomenon, which decreases the accuracy of target detection. The main purpose of background subtraction is to establish a background model and then to subtract the current frame image from the established background model to extract the moving

target. Background subtraction relies primarily on generating a stable background model to obtain a complete foreground feature to detect a complete moving target. The Gaussian mixture model (GMM) [11] is the background modeling method most commonly used for background subtraction. However, this method is susceptible to noise interference under dynamic backgrounds, which leads to poor detection performance; thus, further enhancement is needed.

In recent studies on background subtraction, many new and improved methods have been proposed. Chen and Ellis [12] proposed an adaptive GMM that is not susceptible to dynamic background interference; thus, its detection effect is not good under dynamic backgrounds. In [13], fusion-based foreground enhancement is proposed for background subtraction using a multivariate multimode Gaussian distribution. The method is mainly based on foreground enhancement, and a multivariate multimode Gaussian distribution is used for background modeling. The detection time is high in terms of complexity, the method is susceptible to noise interference, and the detection effect is not good. In [14], a pixel-based adaptive segmentation (PBAS) method is proposed. The method can detect the effect under a static background, but under a dynamic background, the detection result contains many noise points. As a result, the detection accuracy is low. In [15], a fast method is proposed for moving target detection in video surveillance images (FM). This method can quickly detect moving targets; however, these moving targets are incomplete, and voids occur, which reduces the detection performance of this algorithm. In [16], a hybrid method of adaptive GMM and BP neural network (AGBP) is proposed to extract foreground targets from complex scenes. Compared with the original GMM algorithm, the detection effect of this method is greatly improved. In addition, the moving target detected under the dynamic background is relatively complete, but noise removal is not very good, as there are still noise points in the detection result. In [17], various background subtraction algorithms based on GMM are comprehensively summarized and compared based on quantitative evaluation indicators to evaluate their detection effects. In [18], a method for initializing the GMM model using the statistical mean and variance is proposed, and the model is updated by the parameter confidence interval. The method has good effect under a static background, but the detection effect is not good under a dynamic background. In [19], an improved background subtraction method is proposed based on GMM, the component method and the speckle marking method. This method improves the accuracy of moving target detection, but the moving target is detected under a dynamic background; there are still noise points. In [20], a background subtraction method based on GMM with color and depth information is proposed. This method has a good detection effect for specific scenes, but it is not suitable for detecting targets under dynamic backgrounds. The target detected contains a many noise points, and the calibration result is very poor. In [21], a time-space-updated GMM is proposed for moving object detection based on quadratic

segmentation. This method improves moving target detection through quadratic segmentation, but it is not very good under dynamic backgrounds. Because noise interference is suppressed, the presence of noise in the detection result causes the accuracy of moving target detection to decrease. In [22], a moving target detection algorithm based on ORB is proposed for dynamic scenes. This method has a good detection effect under dynamic backgrounds, but the moving target detected under dynamic backgrounds is not very complete and contains some noise. In [23], a complex background subtraction method based on a dynamic space-time model is proposed. The moving target detected in the dynamic background is relatively complete, but the time complexity of the method is high, and there is still noise in the detection results. For background subtraction, the visual background extraction algorithm (ViBe) [24] is also commonly used. The method mainly uses the first frame for background modeling, and the modeling time is fast; however, if the first frame contains moving targets, then ghosting will occur in the detection results, which decreases the detection performance of the algorithm. In [25], a moving object detection algorithm based on improved visual background extraction (IPVB) is proposed. Compared with [24], this method has been greatly improved, but the detection effect under dynamic backgrounds is still not very good, as the test results are noisy. In addition to the above methods, there are other methods, such as the neural network-based semantics segmentation method proposed in [26].

To overcome the interference of dynamic backgrounds and noise, this paper proposes an improved method based on the GMM for moving image target detection. The improved version of the traditional GMM method eliminates noise interference, suppresses the influence of the dynamic background and improves the accuracy of moving target detection under dynamic backgrounds. Compared with other methods, the biggest advantage of this method is that, through the three improvements, noise interference under dynamic backgrounds can be eliminated, and the detection effect of the algorithm is greatly improved. Finally, to verify the advantages of the proposed method, we compare the experimental results of the method with simulated results of the comparison algorithm. The simulated results are analyzed subjectively and objectively to verify the effectiveness and superiority of the improved method proposed in this paper.

The main novelties are summarized as follows:

1) In this paper, we propose to segment video frames into blocks in the background modeling stage, and reconstruct the background model by using the average of image blocks, which reduces the computational complexity and speeds up the modeling speed.

2) In this paper, we propose a method of denoising the detected moving targets by combining the wavelet half-threshold and mathematical morphology denoising, which can effectively remove the noise interference and improve the detection effect of the algorithm, especially in the dynamic background.

(3) In this paper, adaptive background updating method is used in background updating stage to effectively reduce the interference of external environment on background updating, which provides a guarantee for more complete detection of moving objects.

II. CONVENTIONAL GAUSSIAN MIXTURE MODEL

For background subtraction methods, the background model is first established. A method commonly used to establish a background model is GMM. GMM is a classical adaptive Gaussian mixture background extraction method proposed by Stauffer *et al.* [11]; it is a background-based modeling method that builds a color distribution model of each pixel according to the time-domain distribution of each pixel in the video to achieve the goal of background modeling. The GMM is the weighted sum of a finite number of Gaussian functions; it can describe the multipeak state of the pixels, and it is suitable for modeling complex backgrounds such as light gradients and swaying trees. With its continuous improvement, this method has become commonly used for background extraction. The main idea of this method is to build a GMM for each pixel in a sequence of frames. In this model, the background is represented with a large weight, and the foreground is represented with a small weight. If the new pixel matches the Gaussian model corresponding to the background, then it will be treated as a background pixel; if it matches a Gaussian model with a small weight, or if there is no matching Gaussian model, then the pixel is considered to be a foreground pixel. Each pixel can be modeled as a mixture of K Gaussian functions as follows:

$$f(x_t) = \sum_{j}^{k} w_{jt} \times \Phi(x_t; \mu_{jt}, \sum_{jt}) \qquad (1)$$

where w_{jt} is the weight of the jth Gaussian in the mixture at time t, and $\Phi(x_t; \mu_{jt}, \sum_{jt})$ is the Gaussian probability density function with mean μ_{jt} and variance $\sum_{jt}$ for the jth distribution at time t, as follows:

$$\Phi(x_t; \mu_{jt}, \sum_{jt}) = \frac{1}{2\pi^{\frac{D}{2}} |\sum_{jt}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\Delta^2\right\} \qquad (2)$$

$$\Delta = (x_t - \mu_{jt})^T \sum_{jt}^{-1} (x_t - \mu_{jt})^{\frac{1}{2}} \qquad (3)$$

Δ represents the Mahalanobis distance.

After modeling, each new pixel is compared to K Gaussian averages. If the new pixel value x_t is within a multiple of the standard deviation from the mean, then a match is found. Mathematically, this step can be defined as follows:

$$x_t \in \Phi(x_t; \mu_{jt}, \sum_{jt}) if |x_t - \mu_{jt}| < T\sigma_{jt} \qquad (4)$$

where T represents a constant multiplier of the standard deviation, and x_t, μ_{jt}, and σ_{jt} are updated according to recursive formulations as follows:

$$w_{jt} = (1 - \alpha)w_{jt} + a$$
$$\mu_{jt} = (1 - \rho)u_{jt-1} + \rho x_t$$

$$\sigma_{jt}^2 = (1 - \rho)\sigma_{jt-1}^2 + \rho(x_t - \mu_{jt})^T(x_t - \mu_{jt}) \qquad (5)$$

where α and ρ are the learning rate and learning factor, respectively. The distributions are put in descending order of w/σ to determine the background, as the background is supposed to consist of one or more distributions with the highest weights and lowest variances. The first B distributions are chosen as the background, and the following holds:

$$B = \arg\min_{b}(\sum_{j=1}^{b} w_{jt} > Th) \qquad (6)$$

where *Th* is a threshold used to determine the minimum amount of data constituting the background.

III. PROPOSED METHOD

Because the Gaussian hybrid modeling method requires each pixel to be modeled separately, the computational complexity of the algorithm is very high. In addition, the method is susceptible to noise from the dynamic background, which results in poor detection. To accelerate background modeling, reduce computational complexity and improve detection under a dynamic background, this paper proposes an improved background modeling method based on GMM. The proposed algorithm is divided into three main phases. In the background modeling stage, the image sequence image 1 is first used to segment the video sequence image 1, and then, the pixel value of the image block is replaced by the average value of each image block pixel. Finally, the image block mean method based on the GMM method is used. In the moving target detection phase, denoising is performed by combining the wavelet half threshold function and the mathematical morphology method to reduce noise, which improves detection. In the background update phase, the background is updated using the adaptive background update method. The following is a specific introduction.

A. BACKGROUND RECONSTRUCTION

In the background modeling process, image blocks are first selected. When the image block is large, the number of blocks to be processed is small; thus, the efficiency will be higher, but the accuracy of target detection will be reduced. If the selected image block is too small, too many blocks will be processed, resulting in lower efficiency and higher computational complexity. On this basis, this article selects image blocks to reconstruct the background. First, the Gaussian distribution in the GMM is initialized, and then the K-Gauss distribution set in the GMM is used to determine whether each pixel of the current image frame matches the initial distribution. If the Gaussian distribution in the current image frame satisfies formulas (7–8), then the pixel is said to match the Gaussian distribution of the GMM.

$$|D_{ij}^{m,k}| \leq \mu \times M_{ij}^{m-1,k} \qquad (7)$$
$$D_{ij}^{m,k} = G_{ij}^{m,k} - M_{ij}^{m-1,k} \qquad (8)$$

Here, $D_{ij}^{m,k}$ represents the absolute distance between the new pixel value and the mean of Kth Gaussian distributions, μ

represents the threshold of deviation, $M_{ij}^{m-1,k}$ represents the standard deviation of the current image frame pixel. $G_{ij}^{m-1,k}$ represents the standard deviation of the current frame, $G_{ij}^{m,k}$ represents the current image frame, i represents the row number of the current image frame, j represents the column number of the current image frame, m denotes the number of image frames, and K represents the number of Gaussian distributions.

In the process of background reconstruction, if the new pixel value matches the kth Gaussian distribution model, then the model parameters are updated. The update method is shown in formulas (9–11):

$$M_{ij}^{m,k} = (1 - p)M_{ij}^{m-1,k} + p \times G_{ij}^{m,k} \qquad (9)$$

$$w_{ij}^{m,k} = (1 - \alpha)w_{ij}^{m-1,k} + \alpha \qquad (10)$$

$$\beta_{ij}^{m,k} = \sqrt{(1-p)(\beta_{ij}^{m-1,k})^2 + p \times (D_{ij}^{m,k})^2} \qquad (11)$$

$$P = \frac{\alpha}{w_{ij}^{m,k}} \qquad (12)$$

where α represents the learning rate determined by the update rate of the background, $w_{ij}^{m,k}$ represents the weights of the current image pixels and p represents the update rate, as defined in formula (12).

If the new pixel value does not match any of the Gaussian distributions, then a new Gaussian distribution is created with the current minimum weight to replace the original Gaussian distribution. The average value of the newly created Gaussian distribution is taken as the average value of the currently observed pixels. The weight is set to the minimum value of the initialization weight, and the standard deviation is set to the maximum value of the initialization weight. The weight updating method is shown in formula (13).

$$w_{ij}^{m,k} = (1 - \alpha)w_{ij}^{m-1,k} \qquad (13)$$

After building a model for each pixel, only part of the Gaussian distribution represents the background, and the rest represents the foreground. First, the K Gaussian distributions are ranked in descending order of priority, and then the first *Bac* Gaussian distributions are selected to build a background model. The mathematical expression is shown in formula (14):

$$Bac = \arg\min_{b}(\sum_{j=1}^{b} w_{jt}^{mk} > \tau) \qquad (14)$$

where τ is a threshold used to determine the minimum amount of data constituting the background. If a static distribution is selected, then the average value of the distribution will represent the background intensity value; otherwise, the background intensity will be expressed by weighting using the prior weights. To better obtain the background, the ownership coefficients in the model are normalized, and its mathematical expression is shown in formula (15).

$$W_{ij} = \frac{w_{ij}}{\sum_{m=1}^{k} w_{mj}}, \quad i = 1, 2, 3 \cdots, k \qquad (15)$$

where w_{ij} represent represents the weight of the current image pixels.

B. MOVING TARGET DETECTION

The main idea of background subtraction is to extract moving targets from video sequence images by using a differential operation between the current image frame and the established Gaussian background model. In background subtraction, if the difference between the mean of the current frame and the Gaussian model is greater than the standard deviation of the Gaussian model by a factor of δ, then the pixel is considered a target pixel; otherwise, the pixel is considered a background pixel. it's the specific expression is shown in formula (16). In the moving target detection phase, especially under a dynamic background, the main problem is noise. In other words, the better the denoising effect, the better the effect of moving target detection and the higher the accuracy.

$$|D_{ij}^{m,k}| > \delta \times M_{ij}^{m-1,k} \qquad (16)$$

Here, $D_{ij}^{m,k}$ represents the absolute distance between the new pixel value and the mean of the Kth Gaussian distribution, δ is the deviation threshold, and $M_{ij}^{m-1,k}$ represents the standard deviation of the current image frame pixel.

1) NOISE PROCESSING

In the process of detecting moving targets, target detection is not ideal due to noise interference. Therefore, denoising is very important for improving the accuracy of moving target detection. At present, commonly used denoising methods include the mean filter denoising method [27], the median filter denoising method [28], the mean filter denoising method combined with a median filter [29], and the wavelet denoising method [30], [31]. In this paper, the method of combining a semisoft threshold function with mathematical morphology is used to denoise the foreground detection image during moving target detection, and a more complete moving target image is obtained. The denoising effect of this method is better than that of wavelet denoising alone. The wavelet threshold denoising method can be divided into three steps. During (1) wavelet decomposition, the appropriate wavelet and the number of decomposition layers are selected to decompose the noise image signal. During (2) wavelet threshold processing, the wavelet coefficients of each layer are quantified by using the selected threshold function and the decomposed threshold. During (3) wavelet reconstruction, the coefficients of signal processing are used to reconstruct the wavelet to obtain the denoised signal. The principle of wavelet denoising is shown in Figure 1. The selection of threshold function is very important, A suitable threshold can achieve a good denoising effect and a good detection effect The model of the detected image with noise is given by formula (17):

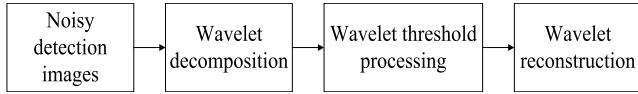$$f_{ij} = x_{ij} + n_{ij} \qquad (17)$$

Noisy detection images → Wavelet decomposition → Wavelet threshold processing → Wavelet reconstruction

FIGURE 1. Principle of Wavelet Denoising.

where f_{ij} represents an image with noise; x_{ij} represents the original image; and n_{ij} represents the 0-mean and σ^2-variance standard Gaussian white noise.

For threshold selection, the most commonly used threshold functions are the soft threshold denoising function and the hard threshold denoising function. The symbolic function is shown in formula (18), the soft threshold function is shown in formula (19), and the hard threshold function is shown in formula (20).

$$\text{sgn}(t) = \begin{cases} 1, & t > 0 \\ -1, & otherwise \end{cases} \qquad (18)$$

$$f(t) = \begin{cases} \text{sgn}(t), & (|t| - \tau), \quad |t| > 0 \\ 0, & otherwise \end{cases} \qquad (19)$$

$$f(t) = \begin{cases} t, & t > 0 \\ 0, & otherwise \end{cases} \qquad (20)$$

Here, t represents the wavelet coefficients, represents the threshold and $f(t)$ represents the wavelet coefficients after threshold processing.

Hard threshold denoising can preserve the local features of image edges, but it will cause image distortion. In the soft threshold denoising method, the wavelet coefficients are compressed, and new wavelet coefficients are used for reconstruction to achieve denoising. However, the soft threshold method also has some drawbacks, e.g., its derivatives are not continuous, which makes it difficult to obtain higher derivatives. Non-negligible mean square error (MSE) will also be generated, resulting in unsatisfactory detection results. To address the defects of the soft and hard threshold functions, an improved semisoft threshold function is proposed. The semisoft threshold function is continuous in the wavelet domain and has continuous high-order derivatives; thus, it has advantages in image denoising tasks. The semisoft threshold function is shown in formula (21).

$$f(t) = \begin{cases} 0, & |t| \leq \tau_1 \\ \text{sgn}(t)\dfrac{\tau_2(|t| - \tau)}{(\tau_2 - \tau_1)}, & \tau_1 < |t| < \tau_2 \\ f(t), & |t| \geq \tau_2 \end{cases} \qquad (21)$$

Here, τ_1 and τ_2 represent thresholds, and wavelet coefficients denote the processing sum of thresholds.

Because of the large noise interference in the dynamic background, there are still some isolated points and local micro-area holes in the target after wavelet denoising. To achieve better detection results, morphological smoothing denoising is carried out based on wavelet denoising. In this paper, the closed operation is used to denoise the moving target, which first expands and then corrodes.
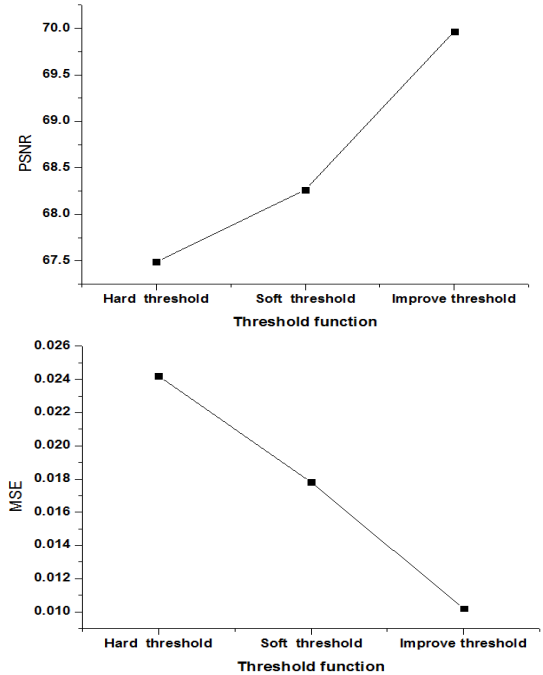

FIGURE 2. Comparison of PSNR and MSE for denoising with different threshold functions.

2) COMPARISON OF DENOISING EFFECTS

In the moving target detection phase, the main task is to denoise the detected target. To determine the main factors affecting wavelet denoising, peak signal-to-noise ratio (PSNR) is used as a basis for comparison. The higher the PSNR, the better the denoising effect. PSNR is defined in (22). MSE is defined in (23).

$$PSNR = 10 \log_{10}\left(\frac{\max |F_r(i,j)|^2}{MSE}\right) \qquad (22)$$

$$MSE = \frac{1}{H \times L} \sum_{i=1}^{H} \sum_{j=1}^{L} (F_r(i,j) - F_r'(i,j))^2 \qquad (23)$$

Here, H and L are used to represent the rows and columns of the image, respectively, $F_r(i,j)$ represents the original image, and $F_r'(i,j)$ represents the denoised image.

To validate the threshold function in the denoising method, simulation experiments are carried out with the hard threshold function, soft threshold function and improved semisoft threshold function. The PSNR and MSE of denoised images are calculated, as shown in Figure 2.

From Figure 2, it can be seen that the PSNR for the improved semisoft function is higher than those for the hard threshold function and the soft threshold function, and the MSE is smaller than that of the improved semisoft function. The results of the simulation experiments are compared in Figure 3.

As show in Figure 3 above, we can see that the semi-soft threshold function has the best denoising effect compared with the hard threshold function and the soft threshold function. In this paper, the method of combining wavelet semi-soft threshold function with mathematical morphology is used to
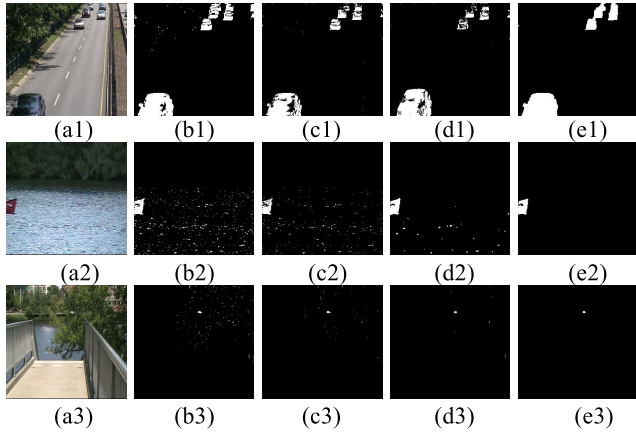
FIGURE 3. Comparison of the denoising effects of different method: (a1) 700th frame; (b1) hard function; (c1) soft function; (d1) semisoft function; (e1) proposed method; (a2) 850th frame; (b2) hard function method; (c2) soft function method; (d2) semisoft function method; (e2) proposed method;(a3) 2205th frame; (b3) hard function method; (c3) soft function method; (d3) semisoft function method;(e3) proposed method.

denoise, which not only eliminates the noise points in the detection results, but also fills in the detection of moving objects with holes, thus achieving the optimization effect, fully proving the superiority of this method in denoising.

C. BACKGROUND UPDATE

Background updating is also very important in moving target detection. To improve the detection results, it is necessary to update the background model. The dynamic nature of the background affects target detection. For example, if there is wind, the leaves of trees in the background will sway, and the test results will be noisy. With the traditional GMM background subtraction method, the background cannot be updated in real time, which leads to ghosting and seriously affects the accuracy of moving target detection. To solve the problem of the traditional algorithm, this paper adopts the adaptive background updating algorithm, which combines the current detection image frame with the background model. The algorithm updates the background model by extracting the current real-time frame so that the detection background can meet the needs of dynamic real-time updating and provides a guarantee for detecting more complete targets. The definition of the adaptive background updating method is as follows (24):

$$B_{s+1}^c = h_s \times B_s^t + (1 - h_s) \times B_s^c \qquad (24)$$

where h_s denotes the update coefficient, and its range is [0,1]; B_s^c denotes the current target frame; and B_s^t denotes the instantaneous target frame, as shown in formula (25):

$$B_s^t = \begin{cases} F_S(i,j), & P^m(i,j) = 0 \\ B_s^t(i,j), & P^m(i,j) = 1 \end{cases} \qquad (25)$$

where $P^m(i,j)$ denotes the binary value after detecting the target; its motion region is 1, and the non-motion region is 0. h_s is determined by the state of the moving target in the
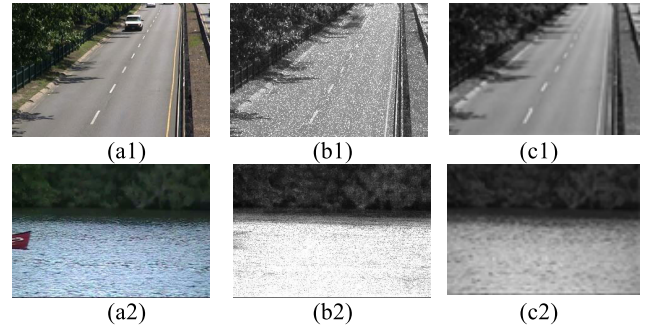

FIGURE 4. Comparing the effect of background updating on the highway dataset and canoe dataset; (a1) 55th frame; (b1) GMM; (c1) proposed method; (a2) 851st frame; (b2) GMM; (c2) proposed method.
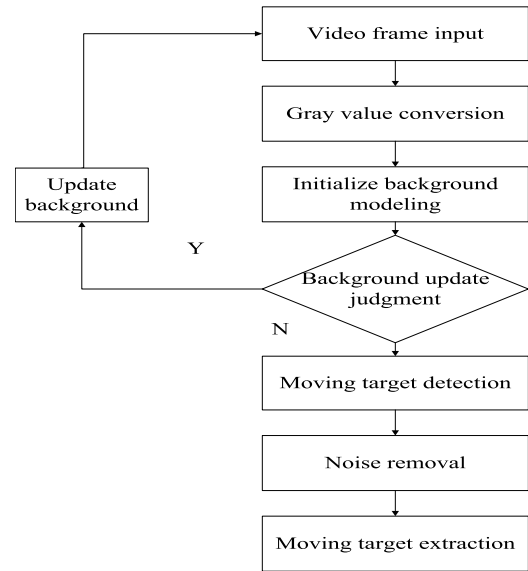

FIGURE 5. Method Flow Chart.

current image frame and the background image frame, and it can be calculated using the following formula (26).

$$h_s = 0.9 \times h_s + 01 \times h\_in \qquad (26)$$

h_in defines the adaptive instantaneous weights of F_n and F_{n-1} in adjacent sequence images.

To illustrate the advantages of background updating, we compare the effect of background updating with that of the original GMM algorithm, as shown in Figure 4.

As shown in Figure 4 above, compared with the background of the traditional GMM method, the background of the proposed method is more complete and close to the real background. This also proves the universality of the adaptive background updating method and provides a powerful condition for improving the accuracy of moving object detection.

D. METHOD FLOW

Background subtraction mainly includes three processes: background model initialization, moving target detection and background updating. In this paper, the background subtraction method is improved compared to the original algorithm.
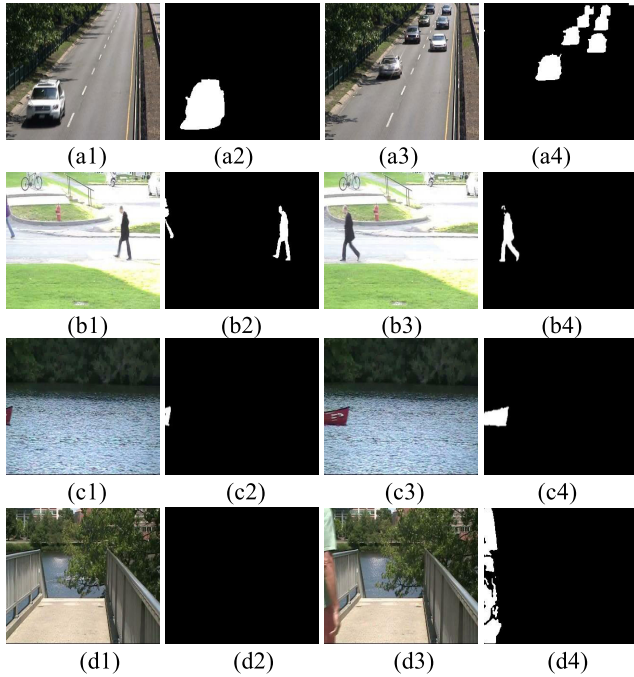
FIGURE 6. Visual results of the proposed method: (a1) 445th frame; (a2) detection result; (a3) 765th frame; (a4) detection result; (b1) 595th frame; (b2) detection result; (b3) 680th frame; (b4) detection result; (c1) 844th frame; (c2) detection result; (c3) 860th frame; (c4) detection result; (d1) 700th frame; (d2) detection result; (d3) 2338th frame; (d4) detection result.

To overcome the interference of noise and the dynamic background, a denoising process is carried out during the detection process. The method mainly includes background modeling, target detection, denoising and background updating, as shown in Figure 5.

IV. EXPERIMENTAL RESULTS

To better illustrate the advantages of the proposed method, we use MATLAB 2014 simulation software for simulation tests. The selection of data sets is very important. For the research of data sets, researchers have also proposed some methods, such as literature [32], which proposes a method to study both unlabeled and unlabeled data sets in a two-way process. In this paper, the pedestrian dataset, highway dataset, canoe dataset and overpass dataset from the change detection dataset were selected as the test datasets for this experiment. The dataset information is shown in Table 1. The detection effect is shown in Figure 6.

A. BACKGROUND MODELING

In the background modeling stage, four test datasets are modeled separately. In GMM background reconstruction, there are many parameters, and different parameters have different values. The definitions and initial values of different parameters in this paper are shown in Table 2. The modeling effect and experimental results are shown in Figure 7.

Under different detection backgrounds, pixels in the same location can be background pixels or foreground pixels.

TABLE 1. Detailed information for the four datasets.

Databases	Number of frame	Size	Background	Moving target
highway	1700	320×240	Roads and trees	moving vehicle
pedestrians	1099	360×240	ground	pedestrians
canoe	1189	320×240	The surface of the sea	moving canoe
overpass	3000	320×240	The tree and overpass	people

TABLE 2. Parameter setting and initialization.

Parameter	Definition	Set value
B	Number of Background Models	3
K	Number of Gauss Distributions	3
μ	Deviation threshold	2.5
α	Learning rate	0.02
σ	standard deviation	15
τ	Foreground threshold	0.3
p	Update rate	0.02

TABLE 3. Comparison of the mean and standard deviation between the GMM method and the proposed in this paper.

		Pixel A1		Pixel A2	
		Traditional GMM	Proposed method	Traditional GMM	Proposed method
pedestrians	M	131.495	116.973	104.1463	125.8162
	σ	9.6562	7.4108	6.4349	5.8498
highway	M	15.0762	118.6870	101.3234	151.7348
	σ	5.6512	5.045	5.4344	4.8492
canoe	M	52.5214	120.8105	167.3792	178.4558
	σ	3.8778	3.0142	3.2158	2.2882
overpass	M	209.3126	223.0824	213.0464	217.0162
	σ	3.8042	3.0125	3.4626	2.6422

To verify that specific pixels in a series of video sequence images are more likely to be background pixels than foreground pixels, two pixels, A1 and A2, are selected in 200 consecutive video sequence images, and then the average and standard deviation of the selected pixels are calculated. Finally, the calculated results are compared with the test results to verify the effectiveness of the proposed method. A comparison between the traditional GMM method and the improved method proposed in this paper, in terms of the mean and standard deviation results, is shown in Table 3.

As seen from Table 3, the mean and standard deviation of the traditional GMM method and the calculation method of the A1 and A2 pixels differ, and the new method is more
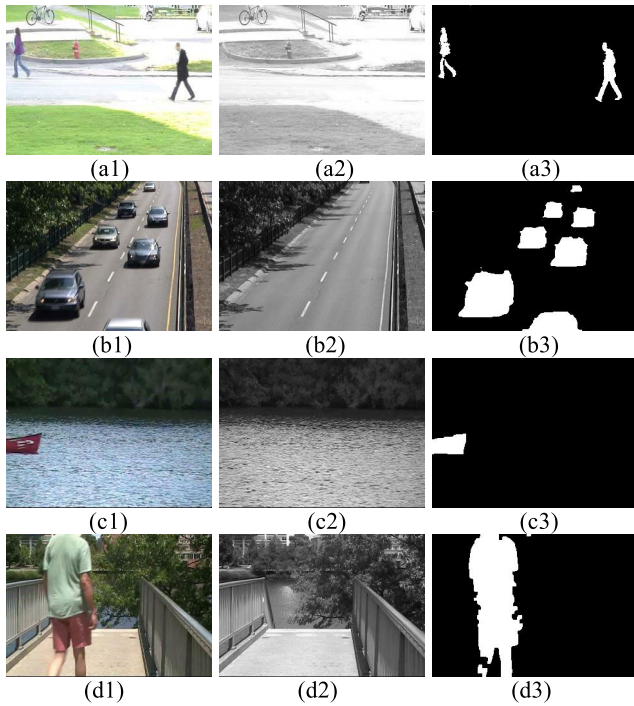
FIGURE 7. Background model and experimental results; (a1) current video frame; (a2) background; (a3) moving target; (b1) current video frame; (b2) background; (b3) moving target; (c1) current video frame; (c2) background; (c3) moving target; (d1) current video frame; (d2) background; (d3) moving target.
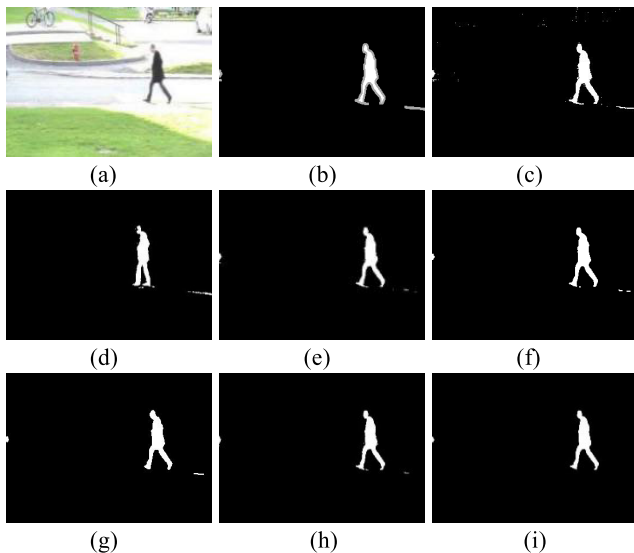

FIGURE 8. Comparative analysis of the test results of the pedestrian dataset: (a) 600th frame; (b) ground truth; (c) GMM; (d) ViBe [24]; (e) PBAS [14]; (f) FM [15]; (g) IPVB [25]; (h) AGBP [16]; (i) proposed method.

stable than the traditional GMM method for background modeling, which further illustrates the advantages of this method over the original GMM method.

B. MOVING TARGET DETECTION

For moving target detection in different scenarios, the detection effect varied due to background changes in the detection

FIGURE 9. Comparative analysis of the test results of the highway dataset: (a) 610th frame; (b) ground truth; (c) GMM; (d) ViBe [24]; (e) PBAS [14]; (f) FM [15]; (g) IPVB [25]; (h) AGBP [16]; (i) proposed method.
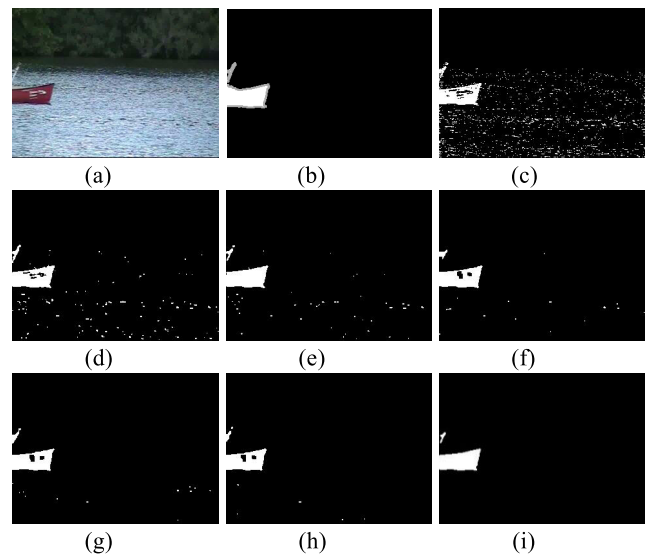
FIGURE 10. Comparative analysis of the test results of the canoe dataset: (a) 865th frame; (b) ground truth; (c) GMM; (d) ViBe [24]; (e) PBAS [14]; (f) FM [15]; (g) IPVB [25]; (h) AGBP [16]; (i) proposed method.

scenarios. The greater the background change of the detection scene, the more vulnerable it is to interference from external factors, and the worse the detection effect. Especially in the dynamic background, it is particularly vulnerable to noise interference. To prove the advantages of the improved method proposed in this paper, we compare and analyze the detection results under a static background and a dynamic background.

1) STATIC BACKGROUND DETECTION RESULTS ANALYSIS

Under the static background, the noise interference is small and the detection effect is better. We first selected two datasets with a static background for simulation experiments. The datasets were pedestrian and highway datasets, and the
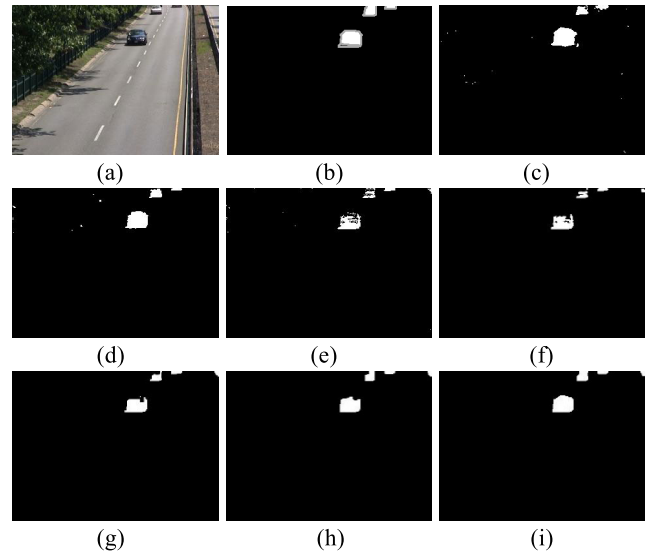
FIGURE 11. Comparative analysis of the test results of the overpass dataset: (a) 2365th frame; (b) ground truth; (c) GMM; (d) ViBe [24]; (e) PBAS [14]; (f) FM [15]; (g) IPVB [25]; (h) AGBP [16]; (i) proposed method.
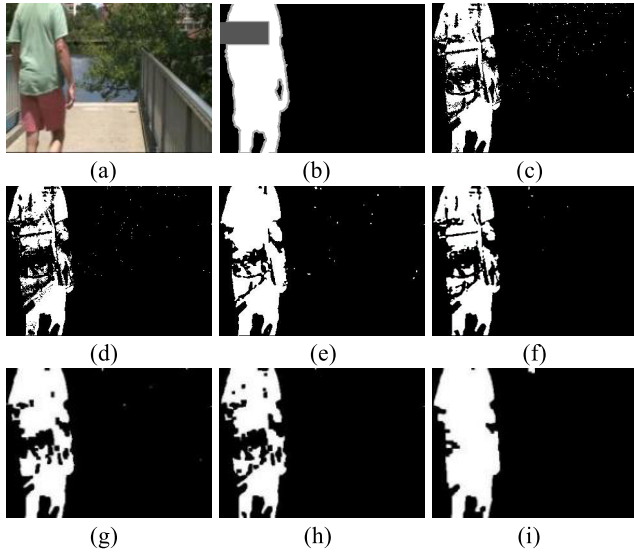
simulation compared the results of GMM, ViBe [24], PBAS [14], FM [15], IPVB [25] and AGBP [16]. The experimental results are shown in Figures 8 and 9.

As shown in Figure 8 and Figure 9, the detection results of the traditional GMM algorithm in a simple background contain noise. The results of other comparison algorithms are better than those of the original GMM algorithm, and the removal of noise is better. However, the drawbacks of each method can be seen from the detection results: some of the results are incomplete and hollow. In contrast, the moving target detected by the proposed method has no noise points and is relatively complete. The results show that, under a static background, the proposed method has certain advantages in moving target detection.

2) DYNAMIC BACKGROUND DETECTION RESULTS ANALYSIS

Under a dynamic background, there will be many noise points in the results of moving target detection, which will affect the accuracy of moving target detection. The traditional GMM method does not perform well under a dynamic background. ViBe [24], PBAS [14], FM [15], IPVB [25] and AGBP [16] are greatly improved methods compared with the traditional GMM algorithm, but the detected moving targets still have holes and noise. Because the proposed method combines wavelet threshold denoising with mathematical morphology closed operation denoising to remove the noise points in the detection results under dynamic backgrounds. The dynamic background datasets are the canoe dataset and the overpass dataset. The detection results of the proposed method are compared with the results of the other algorithm, as shown in Figure 10 and Figure 11.

As seen from Figure 10 and Figure 11, there are many noise points in the detection effect image of the traditional GMM algorithm under the dynamic background. ViBe [24],

PBAS [14], FM [15], IPVB [25] and AGBP [16] remove many noise points, but that there are some noise points remain, and the detected targets are incomplete. Unlike the other algorithms, the proposed method in this paper completely removes noise and detects relatively complete targets In terms of overall analysis, the improved method proposed in this paper performs better than the comparison algorithms under dynamic backgrounds.

3) OBJECTIVE EVALUATION AND ANALYSIS

This paper evaluates and analyses from subjective and objective aspects. Subjectively, this algorithm has certain advantages for moving target detection. To better illustrate the advantages in terms of moving target detection accuracy, this paper objectively analyzed the effect of moving target detection. In the objective evaluation of moving target detection, the pixel features in the moving target detection results are divided into two categories. These two categories are the moving target pixels and the background pixels, and they are recorded as positive and negative, respectively. To illustrate the advantages of the improved algorithm, the indicators [33] Re (recall) (27), Sp (specificity) (28), FNR (false negative rate) (29), FPR (false positive rate) (30), Pr (Precision) (31), F-measure (32) and PCC (percentage of correct classification) (33) were used for evaluation. Recall and precision are shown in Figure 12, and F-measure and PCC are shown in Figure 13. F-measure is the weighed harmonic average of recall and precision [33]. The remaining indicators, i.e., specificity, false positive rate, and false negative rate, are shown in Table 4.

$$Re = \frac{tp}{fn + tp} \quad (27)$$

$$Sp = \frac{tn}{fp + tn} \quad (28)$$

$$FNR = \frac{fn}{fn + tp} \quad (29)$$

$$FPR = \frac{fp}{fp + tn} \quad (30)$$

$$Pr = \frac{tp}{fp + tp} \quad (31)$$

$$F - measure = \frac{2 \times Re \times Pr}{Re + Pr} \quad (32)$$

$$PCC = \frac{tp + tn}{tp + tn + fn + fp} \quad (33)$$

Here, tn denotes the number of pixels correctly classified as negative pixels, i.e., background pixels that are correctly classified as background pixels. fn denotes the number of pixels incorrectly classified as negative pixels, i.e., foreground pixels incorrectly classified as background pixels. tp represents the number of pixels correctly classified as positive pixels, i.e., foreground pixels correctly classified as foreground pixels. fp represents the number of pixels incorrectly classified as positive pixels, i.e., background pixels incorrectly classified as foreground pixels.
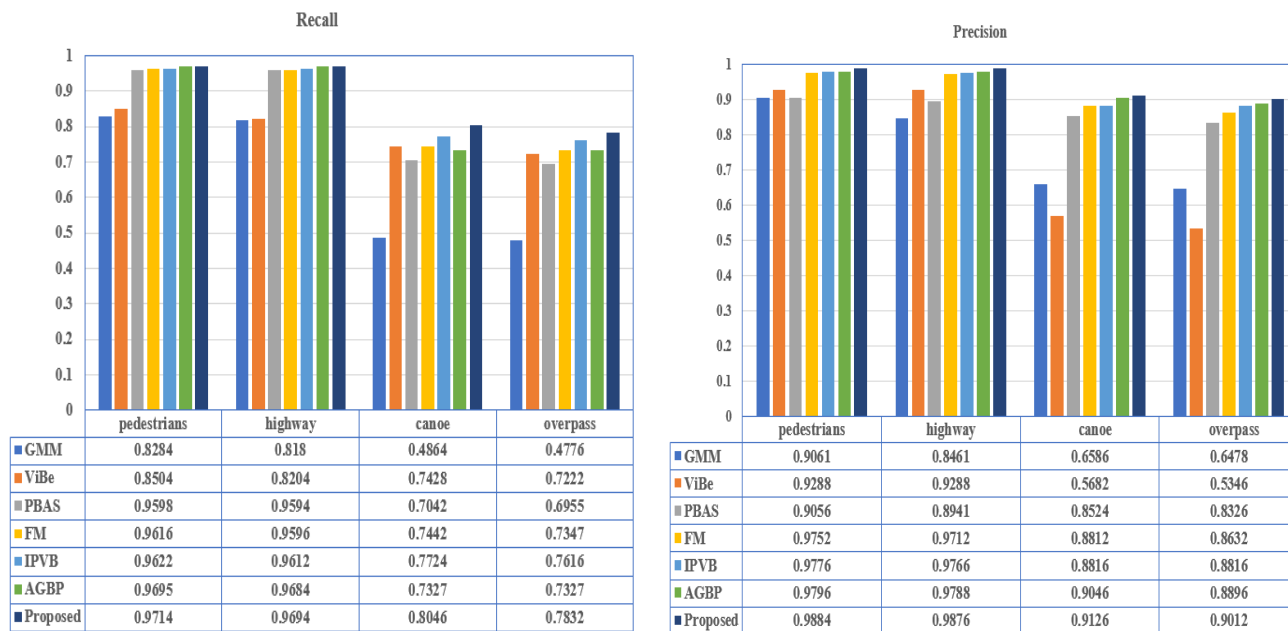
Recall

	pedestrians	highway	canoe	overpass
GMM	0.8284	0.818	0.4864	0.4776
ViBe	0.8504	0.8204	0.7428	0.7222
PBAS	0.9598	0.9594	0.7042	0.6955
FM	0.9616	0.9596	0.7442	0.7347
IPVB	0.9622	0.9612	0.7724	0.7616
AGBP	0.9695	0.9684	0.7327	0.7327
Proposed	0.9714	0.9694	0.8046	0.7832

Precision

	pedestrians	highway	canoe	overpass
GMM	0.9061	0.8461	0.6586	0.6478
ViBe	0.9288	0.9288	0.5682	0.5346
PBAS	0.9056	0.8941	0.8524	0.8326
FM	0.9752	0.9712	0.8812	0.8632
IPVB	0.9776	0.9766	0.8816	0.8816
AGBP	0.9796	0.9788	0.9046	0.8896
Proposed	0.9884	0.9876	0.9126	0.9012

FIGURE 12. The comparison of the recall and precision results of our proposed method and the other methods.

F-measure

	pedestrians	highway	canoe	overpass
GMM	0.8655	0.8318	0.5745	0.5498
ViBe	0.8879	0.8712	0.6439	0.6144
PBAS	0.9319	0.9256	0.7712	0.7579
FM	0.9684	0.9654	0.8069	0.7938
IPVB	0.9698	0.9688	0.8234	0.8172
AGBP	0.9745	0.9736	0.8096	0.8036
Proposed	0.9798	0.9784	0.8552	0.8381

PCC

	pedestrians	highway	canoe	overpass
GMM	0.9456	0.9436	0.9024	0.9018
ViBe	0.9524	0.9518	0.9188	0.9164
PBAS	0.9578	0.9558	0.9256	0.9247
FM	0.9616	0.9604	0.9428	0.9412
IPVB	0.9698	0.9676	0.9464	0.9438
AGBP	0.9745	0.9736	0.9576	0.9544
Proposed	0.9812	0.9804	0.9614	0.9602

FIGURE 13. Comparison of the F-measure and PCC results of our proposed method and other methods.

In Figures 12 and 13, the evaluation indexes of recall rate, precision, F-measure and PCC of the proposed algorithm are higher than those of the compared methods, This also shows that the proposed method has more effective pixels and fewer invalid pixels than the contrast algorithm. That is to say, the detection effect of the proposed method is the best, and the accuracy of target detection is also high. Which further proves the effectiveness of the method proposed in this paper and its advantages in moving target detection.

From the statistical results in Table 4, because the traditional GMM method is susceptible to background changes and the detection results are greatly disturbed by noise, the recall rate is the low, and the missed detection rate and false detection rate are high. Because PBAS [14], FM [15], IPVB [25] and AGBP [16] are less affected by dynamic background changes, their performance evaluation indexes are better than those of the traditional GMM. The improved method proposed in this paper eliminates the influence of

TABLE 4. Objective evaluation and analysis of performance indicators of the algorithms.

	pedestrians dataset			highway dataset			canoe dataset			overpass dataset		
	Sp	FPR	FNR	Sp	FPR	FNR	Sp	FPR	FNR	Sp	FPR	FNR
GMM	0.9968	0.0046	0.1768	0.9948	0.0052	0.182	0.9856	0.0262	0.6228	0.9838	0.0162	0.5224
ViBe [24]	0.9984	0.0018	0.1626	0.998	0.002	0.1796	0.9862	0.0204	0.3724	0.9896	0.0104	0.2778
PBAS[14]	0.9976	0.0026	0.0365	0.997	0.003	0.0406	0.9982	0.0111	0.3246	0.9989	0.0011	0.3045
FM [15]	0.9988	0.0024	0.0326	0.9978	0.0028	0.0348	0.9986	0.0114	0.2878	0.9984	0.0016	0.2673
IPVB[25]	0.9992	0.0021	0.0292	0.9984	0.0024	0.0302	0.9981	0.0126	0.2516	0.998	0.0024	0.2384
AGBP[16]	0.9985	0.0016	0.0226	0.9987	0.0018	0.0238	0.9984	0.0122	0.2726	0.9984	0.0016	0.2673
Proposed method	0.9995	0.0009	0.0126	0.9992	0.0012	0.0156	0.9996	0.0094	0.2212	0.9992	0.0009	0.2012

noise on the detection results and reduces the target void phenomenon; thus, the detected moving target is relatively complete. According to the comprehensive analysis, the number of effective pixels detected by the improved method proposed in this paper increases, and the number of invalid pixels decreases. Therefore, the objective performance evaluation index is better than the performance index of the contrast algorithm.

V. CONCLUSION

To improve the moving target detection accuracy of the GMM algorithm and reduce its susceptibility to noise interference under dynamic backgrounds, this paper proposes a background subtraction method based on improved GMM. The method can detect moving targets under a dynamic background in a video image, and the detection performance is better than that of the original GMM algorithm. The method consists of three stages. In the background modeling stage, the image block mean method is used to build the background model. The second stage is moving target detection. To remove the influence of noise on the detection results, denoising based on a wavelet semisoft threshold function and mathematical morphology denoising method is used to eliminate noise interference. In the background updating stage, the adaptive background updating method is used to update the background. Through a simulation experiment, the improved method is subjectively and objectively better than the compared algorithms, which verifies the effectiveness and adaptablity of the method.

REFERENCES

[1] T. Liu, X. Cao, and J. Jiang, "Visual object tracking with partition loss schemes," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 3, pp. 633–642, Mar. 2017.

[2] T. S. Waykole and Y. K. Jain, "Detecting and tracking of moving objects from video," *Int. J. Comput. Appl.*, vol. 81, pp. 23–28, Nov. 2014.

[3] M. Devanne, S. Berretti, P. Pala, H. Wannous, M. Daoudi, and A. D. Bimbo, "Motion segment decomposition of rgb-d sequences for human behavior understanding," *Pattern Recognit.*, vol. 61, pp. 222–233, Jun. 2017.

[4] X. Y. Zhang, H. Shi, C. Li, K. Zheng, X. Zhu, and L. Duan, "Learning transferable self-attentive representations for action recognition in untrimmed videos with weak supervision," in *Proc. AAAI Conf. Artif. Intell. (AAAI)*, 2019, pp. 9227–9234.

[5] Y. Xin, J. Hou, L. Dong, and L. Ding, "A self-adaptive optical flow method for the moving object detection in the video sequences," *Optik*, vol. 125, pp. 5690–5694, Oct. 2014.

[6] P. Han et al., "An object detection method using wavelet optical flow and hybrid linear-nonlinear classifier," *Math. Problems Eng.*, vol. 2013, no. 1, pp. 1–14, 2013.

[7] Y. H. Cheng and J. Wang, "A motion image detection method based on the inter-frame difference method," *Appl. Mech. Mater.*, vols. 490–491, pp. 1283–1286, Jan. 2014.

[8] M. Wan, G. Gu, E. Cao, X. Hu, W. Qian, and K. Ren, "In-frame and inter-frame information based infrared moving small target detection under complex cloud backgrounds," *Infr. Phys. Technol.*, vol. 76, pp. 455–467, May 2016.

[9] A. Akula, N. Khanna, R. Ghosh, S. Kumar, A. Das, and H. K. Sardana, "Adaptive contour-based statistical background subtraction method for moving target detection in infrared video sequences," *Infr. Phys. Technol.*, vol. 63, pp. 103–109, Mar. 2014.

[10] M. Kaushal and B. S. Khehra, "BBBCO and fuzzy entropy based modified background subtraction algorithm for object detection in videos," *Appl. Intell.*, vol. 47, no. 4, pp. 1008–1021, 2017.

[11] C. Stauffer, and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2, Jun. 1999, pp. 23–25.

[12] Z. Chen and T. Ellis, "A self-adaptive Gaussian mixture model," *Comput. Vis. Image Understand.*, vol. 122, pp. 35–46, May 2014.

[13] T. Akilan, Q. M. J. Wu, and Y. Yang, "Fusion-based foreground enhancement for background subtraction using multivariate multi-model Gaussian distribution," *Inf. Sci.*, vols. 430–431, pp. 414–431, Mar. 2018.

[14] M. Hofmann, P. Tiefenbacher, and G. Rigoll, "Background segmentation with feedback: The pixel-based adaptive segmenter," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Providence, RI, USA, Jun. 2012, pp. 38–43.

[15] R. Zhang, X. Liu, J. Hu, K. Chang, and K. Liu, "A fast method for moving object detection in video surveillance image," *Signal, Image Video Process.*, vol. 11, no. 5, pp. 841–848, 2016.

[16] X. Ou, P. Yan, W. He, Y. K. Kim, G. Zhang, X. Peng, W. Hu, J. Wu, and L. Guo, "Adaptive GMM and BP neural network hybrid method for moving objects detection in complex scenes," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 33, no. 2, 2019, Art. no. 1950004.

[17] K. Goyal and J. Singhai, "Review of background subtraction methods using Gaussian mixture model for video surveillance systems," *Artif. Intell. Rev.*, vol. 50, no. 2, pp. 241–259, 2018.

[18] X. Chen, C. Xi, and J. Cao, "Research on moving object detection based on improved mixture Gaussian model," *Optik*, vol. 126, pp. 2256–2259, Oct. 2015.

[19] D. K. Yadav, "Efficient method for moving object detection in cluttered background using Gaussian mixture model," in *Proc. Int. Conf. Adv. Comput. Commun. Inform. (ICACCI)*, Sep. 2014, pp. 943–948. doi: 10.1109/ICACCI.2014.6968502.

[20] Y.-M. Song, S.-J. Noh, J. Yu, C.-W. Park, and B.-G. Lee, "Background subtraction based on Gaussian mixture models using color and depth information," in *Proc. Int. Conf. Control Automat. Inf. Sci. (ICCAIS)*, Dec. 2014, pp. 132–135. doi: 10.1109/ICCAIS.2014.7020544.

[21] Z. Wang, C. Cheng, and J. Yang, "Scene division-based spatio-temporal updating mixture Gaussian model for moving target detection," in *Proc. 7th Int. Conf. Softw. Comput. Appl.*, Kuantan, Malaysia, 2018, pp. 169–172. doi: 10.1145/3185089.3185127.

[22] W. H. Luo, "A moving object detection algorithm based on ORB under dynamic scene," *Appl. Mech. Mater.*, vols. 602–605, pp. 1638–1641, Aug. 2014.

[23] L. Lin, Y. Xu, X. Liang, and J. Lai, "Complex background subtraction by pursuing dynamic Spatio-temporal models," *IEEE Trans. Image Process.*, vol. 23, no. 7, pp. 3191–3202, Jul. 2014.

[24] O. Barnich andM. Van Droogenbroeck, "ViBe: A universal Background Subtraction Algorithm for Video Sequences," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1709–1724, Jun. 2011.

[25] W. Huang, L. Liu, C. Yue, and H. Li, "The moving target detection algorithm based on the improved visual background extraction," *Infr. Phys. Technol.*, vol. 71, pp. 518–525, Jul. 2015.

[26] X. Zhu, X. Zhang, X.-Y. Zhang, Z. Xue, and L. Wang, "A novel framework for semantic segmentation with generative adversarial network," *J. Vis. Commun. Image Represent.*, vol. 58, pp. 532–543, Jan. 2019.

[27] H. Yu, J. Gao, and A. Li, "Probability-based non-local means filter for speckle noise suppression in optical coherence tomography images," *Opt. Lett.*, vol. 41, no. 5, pp. 994–997, 2016. doi: 10.1364/OL.41.000994.

[28] L. A. Aranda, P. Reviriego, and J. A. Maestro, "Error detection technique for a median filter," *IEEE Trans. Nucl. Sci.*, vol. 64, no. 8, pp. 2219–2226, Aug. 2017.

[29] A. Roy, J. Singha, L. Manam, and R. H. Laskar, "Combination of adaptive vector median filter and weighted mean filter for removal of high-density impulse noise from colour images," *IET Image Process.*, vol. 11, no. 6, pp. 352–361, 2017.

[30] H. Om and M. Biswas, "A generalized image denoising method using neighbouring wavelet coefficients," *Signal Image Video Process.*, vol. 9, pp. 191–200, Jan. 2015.

[31] D. Mukherjee, Q. M. J. Wu, and T. M. Nguyen, "Multiresolution based Gaussian mixture model for background suppression," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 5022–5035, Dec. 2013.

[32] X.-Y. Zhang, S. Wang, and X. Yun, "Bidirectional active learning: A two-way exploration into unlabeled and labeled data set," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 12, pp. 3034–3044, Dec. 2015.

[33] N. Goyette, P.-M. Jodoin, F. Porikli, J. Konrad, and P. Ishwar, "Changedetection.net: A new change detection benchmark dataset," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Providence, RI, USA, Jun. 2012, pp. 1–8.

[34] S. Brutzer, B. Höferlin, G. Heidemann, "Evaluation of background subtraction techniques for video surveillance," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 1937–1944.

JUNHUI ZUO received the B.S. degree from the Lanzhou University of Technology, China, in 2017. He is currently pursuing the M.S. degree with the Department of Information Science and Engineering, Xinjiang University. His research interest includes moving object detection.

ZHENHONG JIA received the B.S. degree from Beijing Normal University, Beijing, China, in 1985, and the M.S. and Ph.D. degrees from Shanghai Jiao Tong University, Shanghai, China, in 1987 and 1995, respectively. He is currently a Professor with the Autonomous University Key Laboratory of Signal and Information Processing, Xinjiang University, China. His research interests include digital image processing and photoelectric information detection and sensors.

JIE YANG received the Ph.D. degree from the Department of Computer Science, Hamburg University, Germany, in 1994. He is currently a Professor with the Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, China. His major research interests include object detection and recognition, data fusion and data mining, and medical image processing.

NIKOLA KASABOV (M'93–SM'98–F'10) received the M.S. degree in computing and electrical engineering and the Ph.D. degree in mathematical sciences from the Technical University of Sofia, Sofia, Bulgaria, in 1971 and 1975, respectively. He is currently the Director and the Founder of the Knowledge Engineering and Discovery Research Institute and a Professor of knowledge engineering with the School of Computing and Mathematical Sciences, Auckland University of Technology, Auckland, New Zealand. His major research interests include information science, computational intelligence, neural networks, bioinformatics, and neuroinformatics.

● ● ●