

Research Article

Ant_ViBe: Improved ViBe Algorithm Based on Ant Colony Clustering under Dynamic Background

Yingying Yue ^{1,2}, Dan Xu ¹, Zhiming Qian,³ Hongzhen Shi,¹ and Hao Zhang¹

¹School of Information Science and Engineering, Yunnan University, Kunming 650091, China

²School of Mathematics and Information Technology, Yuxi Normal University, Yuxi 653100, China

³School of Information Science and Technology, Chuxiong Normal University, Chuxiong 675000, China

Correspondence should be addressed to Dan Xu; danxu@ynu.edu.cn

Received 3 July 2020; Revised 6 August 2020; Accepted 11 August 2020; Published 7 September 2020

Guest Editor: Jiayi Ma

Copyright © 2020 Yingying Yue et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Foreground target detection algorithm (FTDA) is a fundamental preprocessing step in computer vision and video processing. A universal background subtraction algorithm for video sequences (ViBe) is a fast, simple, efficient and with optimal sample attenuation FTDA based on background modeling. However, the traditional ViBe has three limitations: (1) the noise problem under dynamic background; (2) the ghost problem; and (3) the target adhesion problem. In order to solve the three problems above, ant colony clustering is introduced and Ant_ViBe is proposed in this paper to improve the background modeling mechanism of the traditional ViBe, from the aspects of initial sample modeling, pheromone and ant colony update mechanism, and foreground segmentation criterion. Experimental results show that the Ant_ViBe greatly improved the noise resistance under dynamic background, eased the ghost and targets adhesion problem, and surpassed the typical algorithms and their fusion algorithms in most evaluation indexes.

1. Introduction

FTDA [1] means to detect moving targets from the video sequence, achieve foreground pixel segmentation and location, provide robust and stable preprocessing results for subsequent advanced tasks such as target tracking, behavior identity, and gesture recognition. FTDA includes traditional methods [2–7] and deep learning methods [8–10]. Deep learning approach requires more supervision information, manual labeling costs are high, and a lot of repeated training is required when the monitoring scene changes, while traditional methods are unsupervised and with good scenario migration capability. This paper focuses on the traditional method in the field of FTDA. There are three types of traditional FTDA: interframe difference method, optical flow method, and background modeling method [11, 12]. The background modeling method is the most commonly used FTDA in surveillance applications for fixed cameras.

The background modeling method uses the difference between the current frame and the background model to

detect the foreground target. If there is a certain degree of difference between the pixels or other features in the same position, the pixel will be classified as foreground pixel. By further processing of the foreground pixels, more information such as position, shape, and size of foreground target can be obtained. This method has fast calculation speed and high robustness and can extract relatively complete foreground targets under normal circumstances. The key steps of background modeling method are the establishment and updating of the background model. However, obtaining a clean background is difficult because of various interference factors in the actual monitoring scene, which leads to various proposed background modeling methods by creative researchers in recent years. Stauffer and Grimson proposed GMM [2] based on statistical models, which regards all gray values of pixels in the video sequence as a random process and assumes that the appearance of gray values follows Gaussian distribution. Z. Zivkovic and F. van der Heijden. proposed a KNN [3] algorithm based on clustering model, in which standard K -

Means algorithm is used to obtain the mean and variance of each classification through data training, and foreground pixels are segmented according to the difference between pixel gray value and the clustering center. Kim et al. proposed codebook [4] based on clustering, in which background model codebook is obtained by background learning; foreground pixels are segmented according to whether the pixel value is in the corresponding codeword. The above three algorithms have complicated background models and long modeling time, so Barnich and Van Droogenbroeck proposed ViBe [5], respecting the spatio-temporal continuity of the video, establishing a sample set for each pixel by random neighborhood sampling, and using the intersection numbers between pixel R -radius and sample as a basis for segmentation. Combined with an effective random update mechanism, ViBe is a fast, simple, efficient, and with optimal sample attenuation FTDA. Inspired by the theory of control systems, Hofmann et al. proposed an improved ViBe [6], which improves the robustness of the model by dynamically adjusting the threshold and updating rate of each pixel. St-Charles et al. proposed the SuBSENSE [7], which improves feature expression ability by combining color features and texture features and dynamically adjusts its internal parameters using a pixel-level feedback mechanism. Although ViBe and its improved algorithm have made great progress, there are still three limitations (see Figure 1) as follows:

- (1) The noise problem under dynamic background: when there is frequent disturbance in the background, such as swaying branches, water sparkling in the sun, and fountains, a large area of noise will be included in detection result.
- (2) The ghost problem: when the change rate of the foreground target and the background update rate do not match, the detected target is larger than the actual target, and there is smearing in the direction of motion; or when there is a foreground target in the initial frame, the foreground pixels may enter the initial sample set, causing the original foreground target area to be erroneously detected as the foreground and maintained for a period of time.
- (3) Target adhesion problem: the target detected by the ViBe is slightly larger than the actual target, so the targets tend to stick together under the environment of dense targets.

There are two main reasons for the above three limitations. One is because the number of samples stored in the background model (the default is 20) in ViBe is too small, which is only local samples in video sequence. Increasing the number of samples can improve the anti-interference ability, but will reduce the efficiency of the algorithm too. The other is because the segmentation of the foreground and background is only based on intersection between the current pixel neighborhood and sample set, which cannot reflect the change of the global background. In order to improve these limitations, ant colony clustering algorithm is introduced in the proposed Ant_ViBe in this paper without increasing the

complexity, which extends local modeling ViBe to global modeling algorithm.

The main work of Ant_ViBe includes the following: (1) In order to solve the noise problem under dynamic background, the ant colony clustering algorithm (ACCA) is introduced into FTDA area for the first time. Background modeling based on ACCA is established under traditional ViBe framework, which is the foundation of the following main work. (2) In order to solve the ghost problem, a dual-stream update mechanism based on samples and pheromone is proposed, which extends the traditional ViBe update mechanism from local to global. In addition, enlarged sample update range is applied to dual-stream update mechanism. Such an update mechanism can promote quick adaption of background model to the motion state of the foreground target. (3) In order to solve the problem of target adhesion, a nested foreground and background segmentation mechanism is established, taking the historical segmentation statistics of pixels as important consideration.

The typical algorithms, such as ViBe, GMM, KNN, KDE, dense optical flow method, and their fusion algorithms, are compared with Ant_ViBe. The experimental results show that the Ant_ViBe eased the ghost problem and targets adhesion problem, greatly improved the noise resistance under dynamic background, and surpassed the typical algorithms and their fusion algorithms in most evaluation indexes.

2. Related Work

In the past several years, various FTDA have been proposed to build robust and flexible background model which can be utilized in surveillance scenarios with different challenges. To model the variance in video sequences more effectively, probabilistic approaches are adapted. One of the most widely used probabilistic model is the Gaussian Mixture Model (GMM) [2], which uses a mixture of Gaussians to model each pixel instead of modeling all pixel values as one distribution. Kaewtrakulpong and Bowden [13] modified the update equation of [2] for improving accuracy and proposed a shadow detection scheme based on existing GMM. In [6], Zivkovic constantly used a constantly adapted number of Gaussian distributions of the GMM for each pixel.

As for nonparametric approaches, Barnich et al. proposed ViBe [5], where current pixel value is compared to its closest sample within the collection of samples. Combined with an effective random update mechanism, ViBe is robust against small camera movements and noise. Van Droogenbroeck and Paquot [14] proposed several modifications to ViBe by adjusting some parameters. St-Charles et al. [15] combined color intensities and Local Binary String Pattern (LBSP) for detecting camouflaged objects and handling illumination variations. They also proposed PAWCS [16] by using color and texture information as good representational models. Bianco et al. [17] used a Genetic Programming to select and combine best approaches from existing methods then applied post-processing technique to determine final labels.

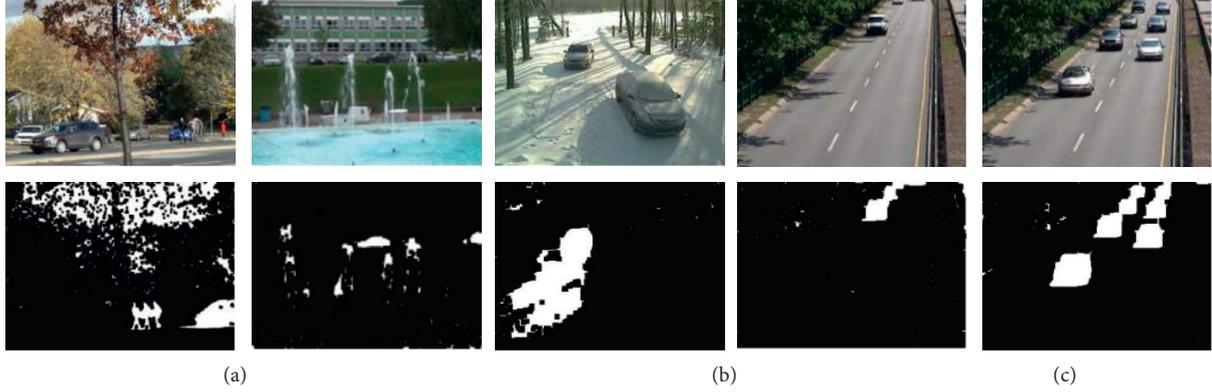


FIGURE 1: The noise, ghost, and targets adhesion problems in ViBe. (a) The noise problem. (b) The ghost problem. (c) Target adhesion problem.

In this work, ACCA is introduced and integrated into traditional ViBe for more effective model of the variance in video sequences. Our method, which is called Ant_ViBe, can improve the noise, the ghost, and the target adhesion problems of traditional ViBe without increasing the complexity. Section 2.1 will focus on the basic steps of traditional ViBe (basic algorithms of Ant_ViBe). Section 2.2 will illustrate the Basic theory of ACCA and the motivation to introduce ACCA into background modeling of ViBe.

2.1. Traditional ViBe. ViBe [5] regards object detection as a binary classification problem. The key to the problem is to determine whether a pixel belongs to a background point or a foreground point. Firstly, ViBe needs to model the background (model representation), which determine the robustness of the algorithm. Then, the background model needs to be initialized. Finally, according to some discriminant principles, the foreground targets detection result is obtained by classification of every pixel, and the background model should be updated according to the segmentation result finally. As mentioned above, ViBe includes following basic steps (see Figure 2).

2.1.1. Background Initialization. Background initialization refers to the process of initializing the parameters in background model. ViBe uses single frame to initialize the background model initialization. For a pixel, considering the spatial distribution characteristics (adjacent pixels have similar pixel values), randomly select the pixel values of its neighbors as its model sample values: $M_0(x) = \{V_0(y) | y \in NG(x)\}$. $M_0(x)$ represents the sample value of pixel x at the initial frame ($t=0$); $NG(x)$ is the neighbor points of pixel x . This initialization method has a sensitive response to noise, a small amount of calculation, and fast speed. The disadvantage is that it introduces ghost area.

2.1.2. Model Representation. Model representation refers to what form of model is used to represent the background, which essentially determines the ability of the algorithm to process the background. In the ViBe, the background model

stores a sample set for every pixel. $p(x)$ is the pixel value at pixel x ; $M(x) = \{V_1, V_2, \dots, V_N\}$ is the background sample set (sample set size is N) at pixel x ; and V_i is the value of the sample i , and i has a value range from 1 to N .

2.1.3. Foreground Target Detection. Foreground target detection is the process of classifying pixels according to the foreground and background segmentation strategy and finally obtaining the foreground target area with some postprocessing operations. Calculate the distance between the new pixel value and each sample value in the sample set. If the distance is less than the threshold, the number of similar sample points $W(x)$ increases. $SR(p(x))$ is an area with x being the center and R the radius for the determination of similar distances (see Figure 3). If $W(x)$ is greater than the threshold, the new pixel point is considered as the background. As shown in the following equation, two parameters are involved: the number of sample sets is N ; the threshold of $W(x)$ is m :

$$W(X)[\{SR(p(x)) \cap \{V_1, V_2, \dots, V_N\}\}] > m. \quad (1)$$

2.1.4. Background Updating. The function of background updating is to promote the background model to adapt to the changes in the monitoring scene. The ViBe uses a conservative updating strategy (the foreground pixels are not involved in the updating), foreground pixels counting method (if a pixel is detected as the foreground for N consecutive times, it is updated as a background point; for example, a car parked for a long time can be converted into a background), and a stochastic diffusion updating mechanism (each background pixel can randomly update its own sample value and neighbor point's sample value).

2.2. ACCA. ACCA [18] is a clustering algorithm based on ant colony foraging. In the ant colony's foraging behavior, each ant has no information on the food source in advance. After finding the food, the ants will release a volatile secretion to the environment (called pheromone, this substance will gradually evaporate and disappear over time, and

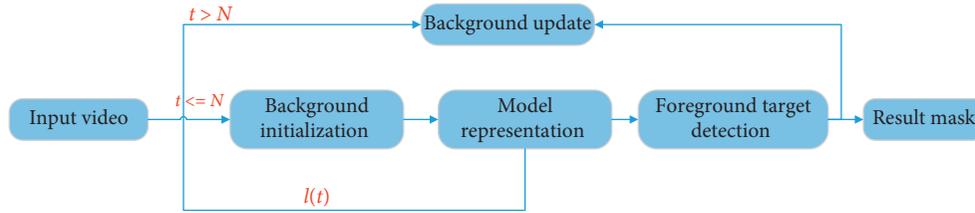


FIGURE 2: Basic steps of ViBe algorithm.

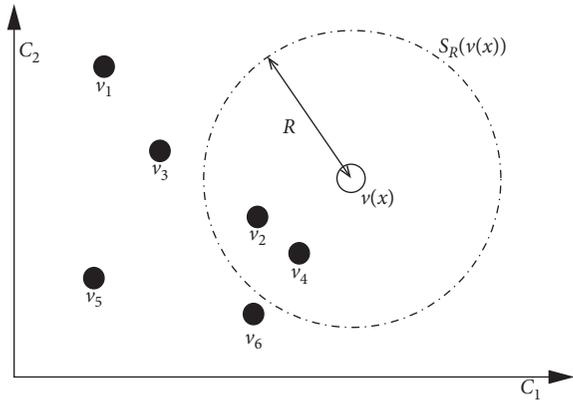


FIGURE 3: Distance between the current pixel and sample.

the size of the pheromone concentration indicates the distance of the path) to attract the rest of the ants. The ants choose path according to a certain probability, so that the paths are diversified. If there is a better path, more ants will be attracted to this path over time. In analysis based on ant colony foraging behavior, the data are regarded as ants with different attributes, and the clustering results are regarded as food sources. Each ant moves with a certain probability, gathers on different food sources, and finally achieves clustering (see Figure 4).

The combination of ACCA and FTDA: in FTDA, the background pixels will basically remain stable in time series, and the situation where the pixels become foreground is contingent and transient. As time goes, the number of image frames (ants) is increasing, and more and more ants will concentrate on the position of the background pixels. At the beginning, the background pixels are spotted then slowly clustered and stabilized, and then the foreground target is segmented (see Figure 5). The pheromone of ACCA is global (from the start frame to current frame), while the sample set of ViBe is based on partial information (the n default number of samples is 20). Ant_ViBe will use both local and global information to realize nested foreground and background segmentation.

Aiming at the three limitations of ViBe, this paper introduces ACCA into the ViBe framework. Without increasing computational complexity, a new Ant_ViBe algorithm based on global background modeling is proposed.

3. Improved Ant_ViBe Algorithm

The main work of Ant_ViBe includes the following: (1) background modeling based on ACCA under traditional

ViBe framework: definition of concept and parameters, initializing ant colony parameters, constructing pheromone matrix, and constructing objective function, (2) a dual-stream updating mechanism based on pheromone and sample with enlarged sample range, and (3) nested foreground and background segmentation mechanism. The rest of this section is organized as follows. Background modeling based on ACCA is illustrated in Section 3.1, dual-stream updating mechanism is showed in Section 3.2, and nested segmentation mechanism is shown in Section 3.3. The flowchart of Ant_ViBe is shown below (see Figure 6).

3.1. Background Modeling Based on ACCA. Because Ant_ViBe in this paper introduces the ACCA to foreground target detection problem for the first time, we need to model the background based on ACCA and integrate it into the ViBe framework. Background modeling based on ACCA includes four steps as follows.

3.1.1. Definition of the Concept and Parameters of ACCA and FTDA. In order to apply ACCA to the ViBe, the concept and parameter of the ACCA and FTDA is defined, and the corresponding relationships are listed (see Table 1).

3.1.2. Initializing Ant Colony Parameters. In Ant_ViBe, the frame S is regarded as an ant. The ant exists on the time axis t and increases with time. Set the size of the frame image to $M \times N$. Each ant needs to determine each pixel of the current frame based on all previous frame pheromone attributions. Each pixel has two types of classification: background and foreground. The output result of the Ant_ViBe is a binary image (antimage) and target localization. Ant S classifies $M \times N$ pixels into 2 categories, and each ant corresponds to a solution set.

3.1.3. Constructing Pheromone Matrix. Let the pheromone at the sample point $P[i][j]$ be $antmat[i, j, k]$, $k=0$ represents the background pheromone, and $k=1$ represents the foreground pheromone. For example, $antmat[i, j, 0]$ represents pixel $P(i, j)$ belonging to the background pheromone, and $antmat[i, j, 1]$ represents pixel $P(i, j)$ belonging to the foreground pheromone. The pheromone is initially assigned a value of 0 (see Table 2).

3.1.4. Objective Function. The objective function value of each frame (ant) is the sum of the distances of pixels

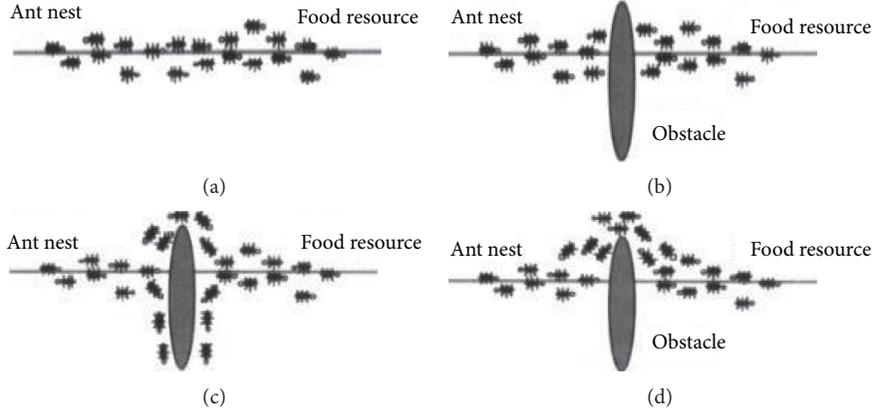


FIGURE 4: Clustering algorithm for ant colony foraging.

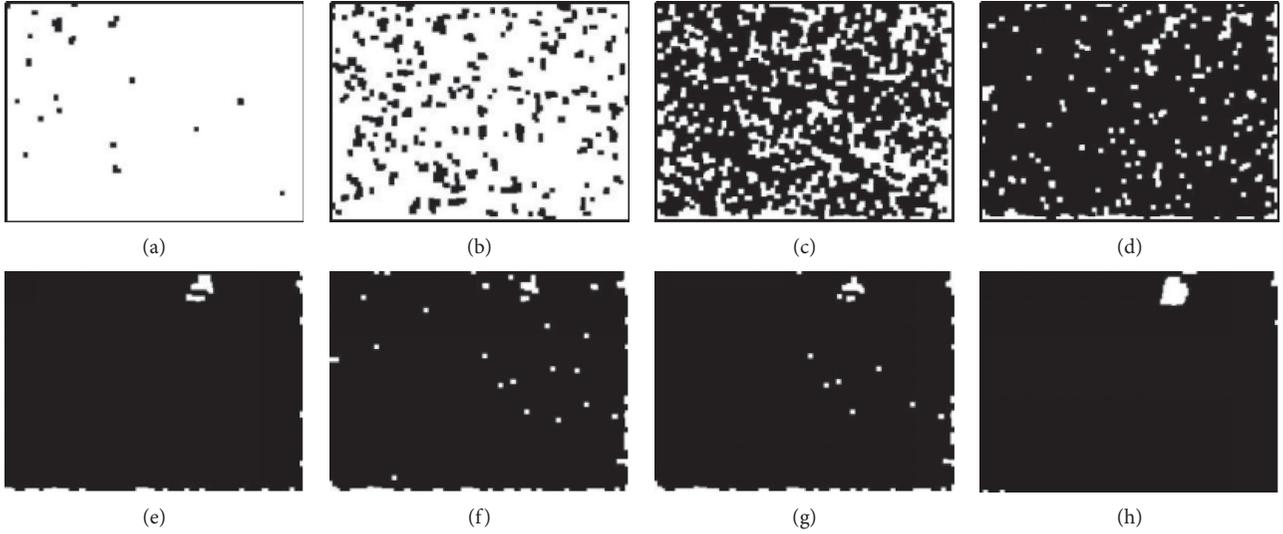


FIGURE 5: The combination of ACCA and FTDA.

(samples) on the time axis to the cluster center, as shown in (2). Let J^t be the objective function value. There are 2 pattern classifications $|S_j, j = 1, 2|$ (foreground and background). $X[i, j]^t$ is the gray value at the pixel point $P(x, y)$ at time t . C is the current frame number:

$$J^t = \sum_{k=1}^2 \sum_{i=1}^M \sum_{j=1}^N \sum_{t=1}^C W_k \|X[i, j]^t - C_{ij}\|, \quad (2)$$

$$C_{ij} = \frac{\sum_{t=1}^C W_j X[i, j]^t}{\sum_{t=1}^C W_j}, \quad (3)$$

$$W_k = \begin{cases} 1 & X[i, j]^t \in S_j \\ 0 & X[i, j]^t \notin S_j. \end{cases} \quad (4)$$

3.2. Dual-Stream Background Model Update Mechanism.

In the Ant_ViBe, there are two ways to update the background model: one is the pheromone updating, and the other is the sample set updating. In addition, in order to improve the adaptation speed of the background model to the motion state of the foreground target, the update range is expanded when each pixel updates its field randomly.

3.2.1. Enlarging Sample Update Range. When there is frequent and small-scale interference in the background, a large area of noise will appear in ViBe, as shown in Figure 1. Therefore, in Ant_ViBe, the random sampling and updating range of 8 neighborhoods is changed to 16 neighborhoods (see Figure 7). In this way, the pixel values in the 16 neighborhoods can enter the background model, and the new pixel values at the same position when the branches swing are likely to be P in 16 neighborhoods. Increasing sampling and updating neighborhood will make P more likely to become the background sample, reduce the foreground misjudgment, and improve the accuracy of the algorithm.

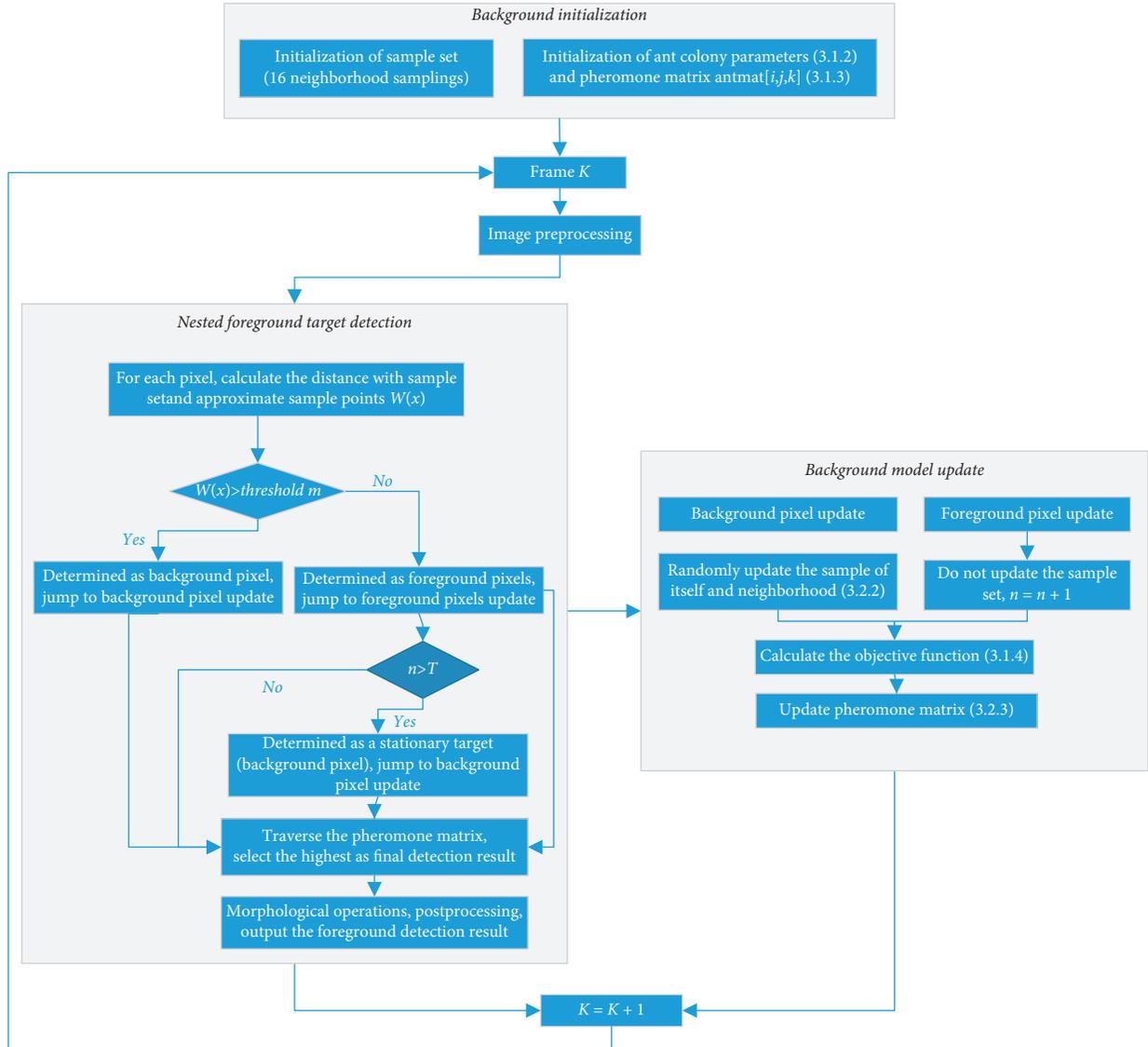


FIGURE 6: Flowchart of Ant_ViBe.

TABLE 1: Parameter mapping of ACCA and Ant_ViBe.

ACCA	Ant_ViBe
Ant	Image frame
Number of ants	Number of frames
Number of clusters	2 classifications: foreground and background
Ant current path	Segmentation result of current frame
Sample	Pixel

3.2.2. Sample Set Update Strategy. In Ant_ViBe, the random update strategy of the traditional ViBe algorithm is used. The update neighborhood is changed from 8 to 16. Each background point has a probability of $1/\varphi$ to update its own model sample and its neighbors. Updating neighbors takes advantage of the spatial propagation characteristics of pixel values, and the background model gradually diffuses outward, which facilitates the faster elimination of the ghost area. When the count of foreground pixels goes up to threshold T , it becomes the background, and there is a probability of $1/\varphi$ to update its

own sample. When selecting the sample to be replaced, randomly select a sample to update, which ensures a smooth lifecycle of the sample value. Due to a random update strategy, the probability that a sample value is not updated at time t is $(N-1)/N$; assuming that the time is continuous, then the probability that the sample value will remain after the time of dt is shown in the following equation. This indicates that whether a sample value is replaced in the model is independent of time t , and a random strategy is appropriate:

$$p(t, t + dt) = \left(\frac{N-1}{N}\right)^{(t+dt)-t}. \quad (5)$$

3.2.3. Pheromone Matrix Update. The update of pheromone includes three processes: volatilization, enhancement, and decreasing. The volatilization process attenuates the pheromone at a certain rate, simulating the process of volatilization of the natural ant colony with time. This volatilization process

TABLE 2: Initialization of pheromone matrix.

Sample\category	Background	Foreground
$P(0, 0)$	0	0
$P(0, 1)$	0	0
$P(0, 2)$	0	0
...	0	0
$P(M, N)$	0	0

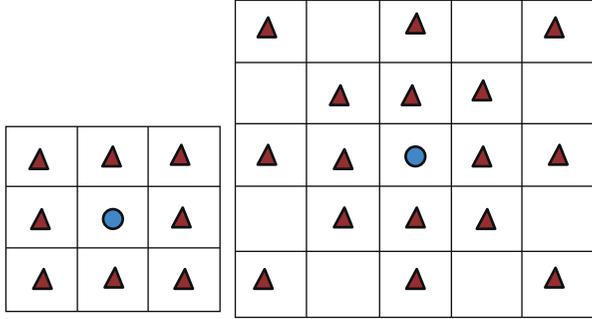


FIGURE 7: Enlarged sampling and updating range.

guarantees the continuous update of the algorithm in the time series and is realized by the volatilization parameter. The processes of pheromone enhancement and decreasing mean that when a pixel is classified as background, the corresponding column of background in pheromone matrix is increased, while the corresponding column of foreground is decreased. Similarly, when the pixel is classified as foreground, add the pheromone to the column corresponding to foreground in pheromone matrix, and reduce the pheromone to the column corresponding to background, as shown in equations (6) and (7). At the same time $\%AV = (\sum_{j=1}^N AV_j/N)$, the strength of the enhancement and weakening process is controlled by the objective function J^t on the time axis. When J^t is large, the update range is small, and vice versa. J^t ensures that the background model can be adaptively updated according to the current frame changes, as shown in (8). In this way, global motion that cannot be achieved by a single ant can be achieved, and global segmentation result statistics can be established for each pixel in the time series without increasing the computational complexity, which provides a more robust basis for subsequent segmentation of foreground and background:

$$\text{antmat}[i, j, 0]^{(t+1)} = (1 - \rho)\text{antmat}[i, j, 0]^{(t)} + \sum_{s=1}^t \Delta\tau_{ij}^s, \quad (6)$$

$$\text{antmat}[i, j, 1]^{(t+1)} = (1 - \rho)\text{antmat}[i, j, 1]^{(t)} + \sum_{s=1}^t \Delta\tau_{ij}^s, \quad (7)$$

$$\Delta\tau_{ij}^s = \begin{cases} \frac{Q}{J^t} & P(i, j) \in \text{background} \\ -\frac{Q}{J^t} & P(i, j) \in \text{foreground} \end{cases} \quad (8)$$

In formula (8), ρ is the volatilization parameter, S is the frame, t is the current frame, Q is the pheromone concentration parameter, and J^t is objective function value of ant S .

3.3. Nested Foreground Target Detection. Foreground target detection in Ant_ViBe is nested. The inner layer uses the traditional ViBe to compare the similarity between the current pixel and the sample set and then determines whether the pixel is foreground or background. This inner layer determination is based on the local information of the sample set. For each frame, the global parameter pheromone is updated in the outer layer according to the judgment result of the inner layer. According to the pheromone matrix at current frame, the one with the highest pheromone is selected as the final judgment result, as shown in the following equation:

$$\text{antimage}[i, j] = \begin{cases} 255, & \text{antmat}[i, j, 0] \leq \text{antmat}[i, j, 1], \\ 0, & \text{else.} \end{cases} \quad (9)$$

Ant_ViBe's nested foreground target detection mechanism effectively combines local and global information. With the passage of time, the background pixels will accumulate more and more ants, and the white area will be gradually separated. This white area is the foreground target to be extracted by the algorithm.

4. Experimental Results and Analysis

The experiments in this paper were carried out on the dataset CDnet2014 [19]. There are 53 different video sequences in CDnet2014 totally, which are divided into 11 categories (baseline, camera jitter, bad weather, dynamic background, intermittent object motion, low frame rate, night videos, PTZ, shadow, thermal, and turbulence). Each category contains from 4 to 6 video sequences, and each video sequence contains from 900 to 7000 frames. Spatial resolutions of video frames vary from 320x240 to 720x576 pixels. CDnet2014 contains different challenging scenarios which is appropriate for measuring the robustness of algorithm. In experiments, video sequences with large background interference, prone to ghost area, and intermittent targets (baseline, dynamic background, intermittent object motion, and camera jitter) are selected for testing the performance of Ant_ViBe.

All the experiment parameter settings are listed in Table 3. N is the size of sample set, R is the radius for the determination of similar distances as shown in Figure 3, and m is a threshold for the number of approximate sample points. $CLASS_NUM$ represents the number of categories of pixel classification. In this paper, pixels are divided into two types: background and foreground. ρ is the volatilization parameter, and Q is the pheromone concentration. T is a threshold, which is explained in Section 3.2.2. When the number of times that a certain pixel is continuously detected

TABLE 3: Default values of experimental parameters.

Parameters	N	R	m	CLASS_NUM	ρ	Q	T
Default values	20	20	2	2	0.02	0.1	50

as the foreground exceeds T , the pixel classification will be changed from the foreground to the background, such as a parked car.

4.1. Qualitative Analysis. In order to test the performance of Ant_ViBe under the three problems of dynamic background, ghosting, and target adhesion, four sets of comparative experiments were carried out in this section. Sections 4.1.1 and 4.1.2, respectively, are the comparisons of typical algorithms and their fusion algorithms. Section 4.1.3 is the comparison of antinoise performance, and Section 4.1.3 is the comparison of ability to eliminate ghost area. This paper focuses on the traditional method without deep learning in the field of FTDA, so four typical traditional algorithms with relatively good performance (GMM, KNN, Flow, and ViBe) are selected for comparison experiments with Ant_ViBe.

4.1.1. Comparison of Typical Algorithms. Comparison of typical algorithms under multiple data streams in CDnet 2014 is shown above (see Figure 8). The vehicles detected by the ViBe algorithm in (b) column are stuck together, while the vehicles detected by Ant_ViBe are independent, which solves the problem of target adhesion. There are slight swings in the branches on the upper left corner of the (c) column. The other four algorithms had noise to some degree, and Ant_ViBe basically eliminated the effect of noise and obtained ideal detection results. There is a large area of the branches swinging in the (d) column. The detection results of the other four algorithms contained a large area of noise, while Ant_ViBe eliminated the noise interference, showing good noise resistance performance. In the (e) column, there are intermittently moving vehicles, which change from a stationary state to a moving state. The other four algorithms all have a large area of ghost area, but Ant_ViBe has a significant performance in eliminating it. The fountain in the (f) column is easily detected as foreground target due to its movement by the other four algorithms, especially the optical flow method, while Ant_ViBe succeeds in integrating the fountain area into the background to accurately detect the real moving target in the scene.

4.1.2. Comparison of Fusion Algorithms. Under the dynamic background of the fountain, the fusion algorithms of the four classic algorithms ViBe, flow, GMM, and KNN will detect the fountain as foreground target in different degrees. Our Ant_ViBe eliminates the influence of the fountain well, and the target detection performance under dynamic background is better than the fusion algorithm (see Figure 9).

4.1.3. Comparison of Antinoise Performance. Experiment in this part uses the “fall” and “canoe” video streams in the “Dynamic background” category of the CDnet dataset. There are frequent background disturbances, such as tree disturbances and sparkling water. It is very suitable for testing the antinoise performance of the Ant_ViBe algorithm. Ant_ViBe can remove frequent branches and water waves to obtain a relatively clean foreground target (see Figure 10).

4.1.4. Comparison of Ability to Eliminate Ghost Area. This experiment uses the “Winter Driveway” video stream in the “Intermittent Object Motion” category. The stationary car in the video reverses between 1800 frame and 2080 frame; its state changes from background to moving foreground target. Ant_ViBe can alleviate the ghost problem: the ghost area is relatively small and the duration time of ghost is shorter than other algorithms (see Figure 11).

4.2. Quantitative Analysis. With the combination of the ACCA, Ant_ViBe can realize more robust and flexible background modeling based on the historical global information of each pixel. False targets can be removed effectively by Ant_ViBe under dynamic background, and the accuracy of foreground segmentation is improved correspondingly. In order to show the better performance of Ant_ViBe under dynamic background compared to traditional algorithms, six typical and widely used traditional background modeling algorithms with relatively good performance (GMM, KNN, PSP-MRF, KDE, Bayesian Background, and ViBe) are selected for quantitative analysis in Section 4.2.2, and the comparison experiments for quantitative analysis are conducted on the “Dynamic Background” categories in CDnet2014. Because our Ant_ViBe is improved under the framework of traditional ViBe, the comparison of Ant_ViBe and ViBe is performed in Section 4.2.3 to show the better segmentation performance of our algorithm compared to ViBe. The evaluation metrics used in the quantitative analysis are introduced in detail in Section 4.2.1.

4.2.1. Evaluation Metrics. In the field of FTDA, there are mainly seven evaluation indexes (recall, precision, F-measure, specificity, FPR, FNR, and PWC) commonly used. These seven indexes are calculated by formulas (10) to (16). In the formulas given as follows, there are four basic variables: TP, TN, FP, and FN. True Positive (TP) represents the number of pixels correctly classified as foreground. True Negative (TN) represents the number of pixels correctly classified as background. False Positive (FP) represents the number of pixels that are incorrectly classified as foreground. False Negative (FN) represents the number of pixels that are incorrectly classified as background:

$$\text{recall} = \frac{TP}{TP + FN}, \quad (10)$$

$$\text{precision} = \frac{TP}{TP + FP}, \quad (11)$$

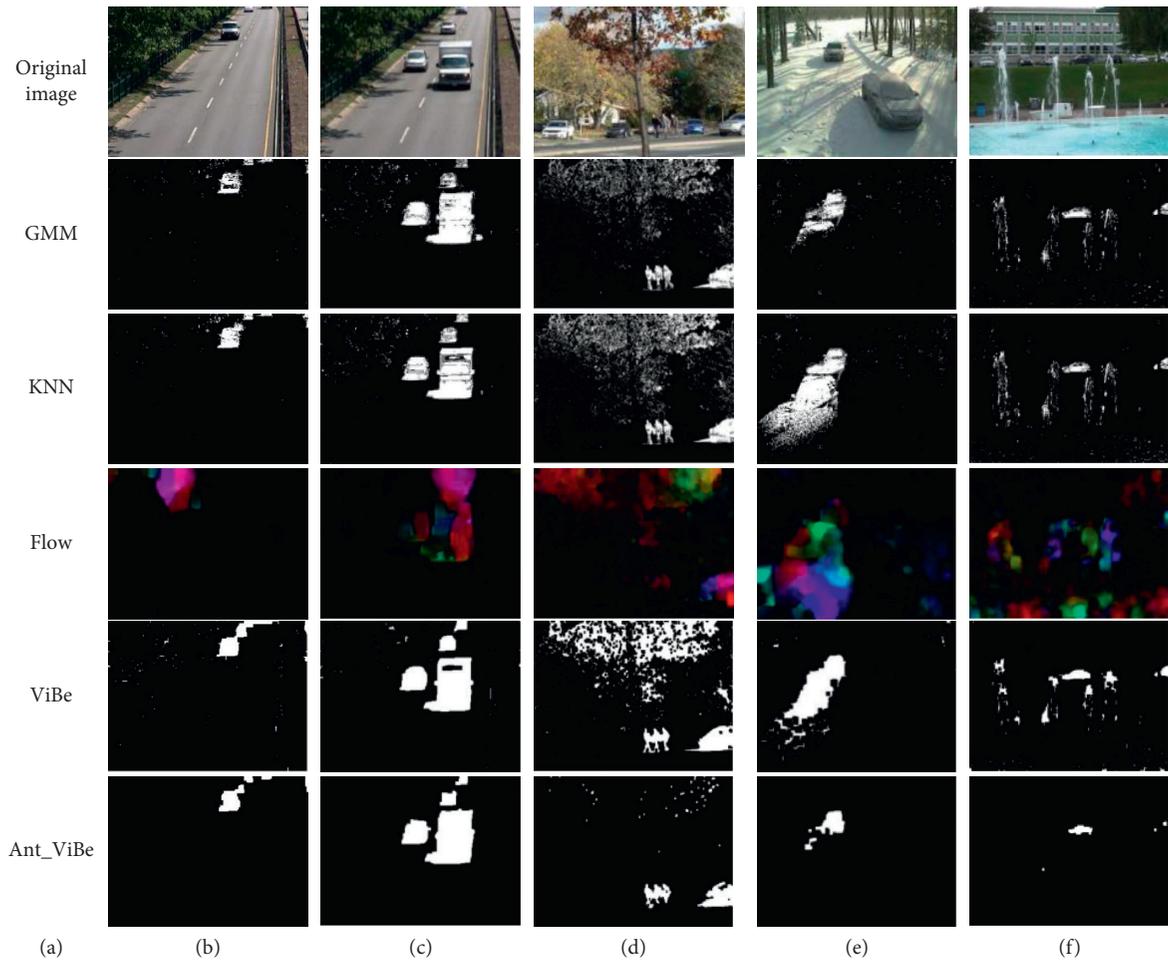


FIGURE 8: Comparison of typical FTDA.

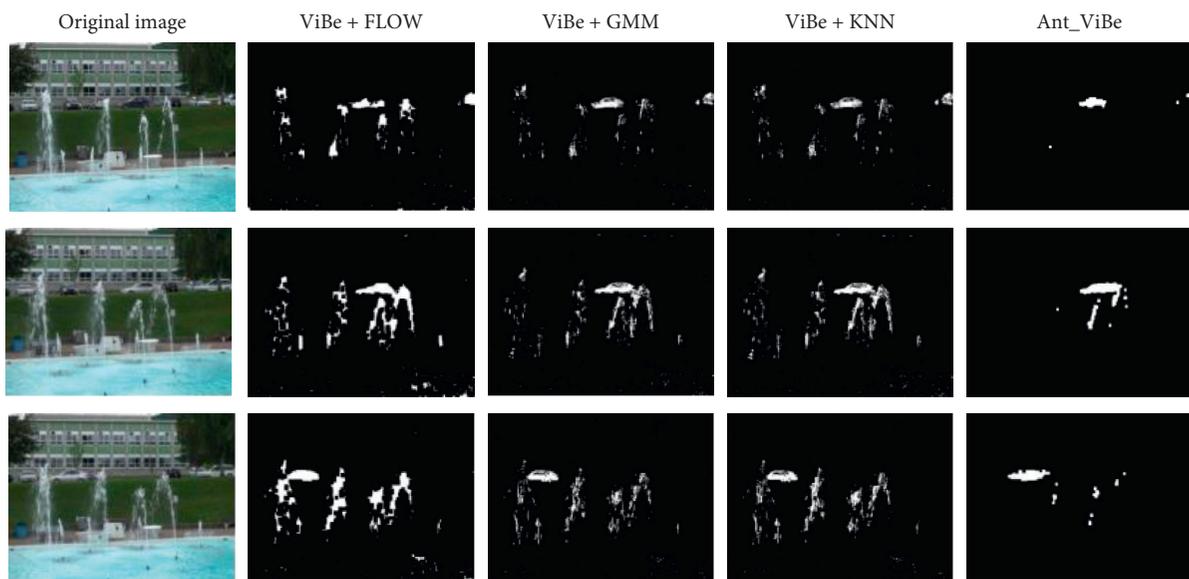


FIGURE 9: Comparison of fusion algorithm and Ant_ViBe.

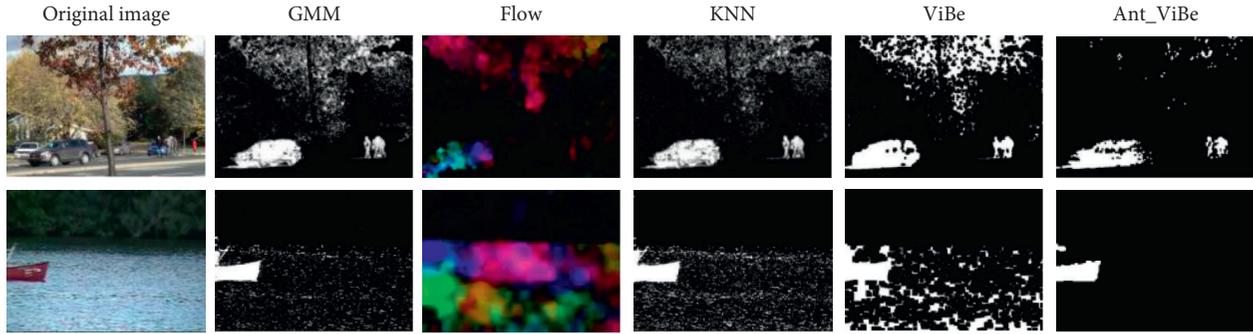


FIGURE 10: Comparison of noise resistance.

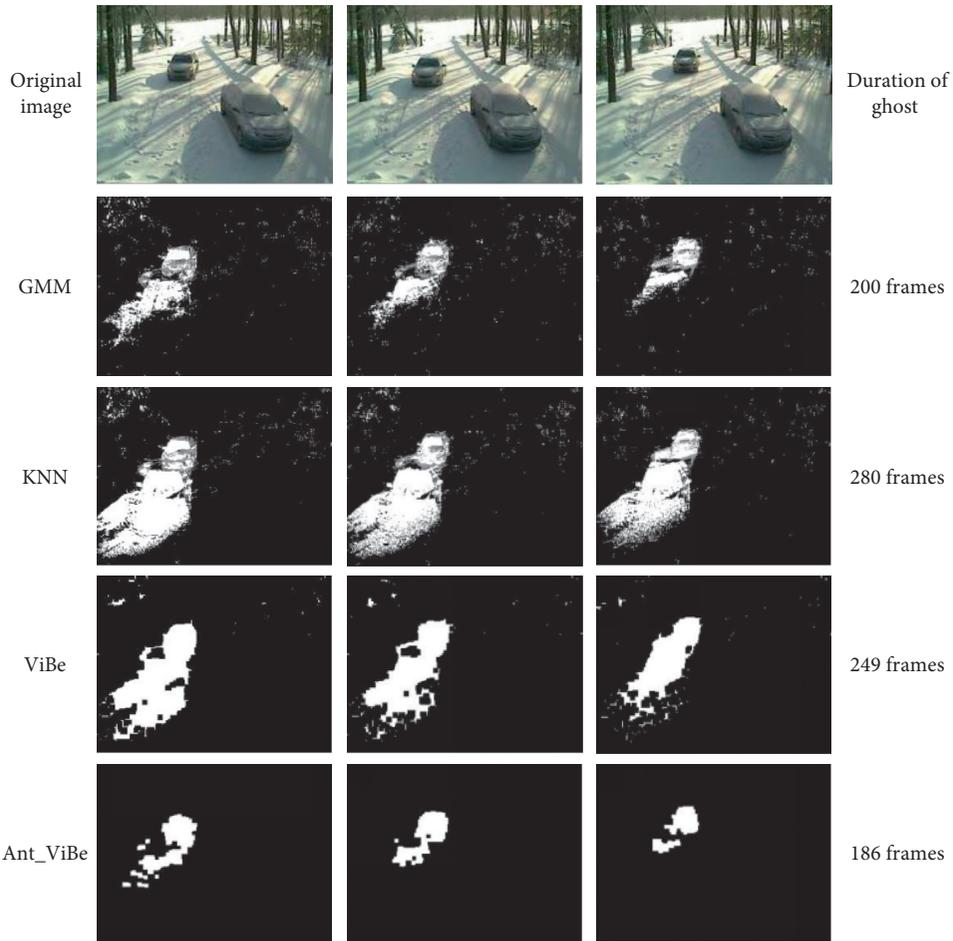


FIGURE 11: Comparison of the ability to eliminate ghost area.

$$F - \text{measure} = 2 * \frac{\text{recall} * \text{precision}}{\text{recall} + \text{precision}}, \quad (12)$$

$$\text{FNR} = \frac{\text{FN}}{\text{TP} + \text{FN}}, \quad (15)$$

$$\text{specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \quad (13)$$

$$\text{PWC} = 100 * \frac{\text{FN} + \text{FP}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}}, \quad (16)$$

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}}, \quad (14)$$

In quantitative analysis, recall represents the percentage of the number of foreground pixels correctly detected to the number of all foreground pixels in the benchmark result

TABLE 4: Comparison of Ant_ViBe and traditional FTDA under dynamic background.

Method	Recall	Specificity	FPR	FNR	PWC	F-measure	Precision
KNN [3]	0.8047	0.9937	0.0063	0.1953	0.8059	0.6865	0.6931
PSP-MRF [8]	0.8955	0.9859	0.0141	0.1045	1.4514	0.6960	0.6576
KDE [15]	0.8401	0.9908	0.0092	0.1599	1.1501	0.6016	0.5413
GMM [2]	0.8344	0.9896	0.0104	0.1656	1.2083	0.6330	0.5989
Bayesian Background [20]	0.5962	0.9917	0.0083	0.4038	1.2427	0.5369	0.6898
ViBe [5]	0.7222	0.9896	0.0104	0.2778	1.2796	0.5652	0.5346
Ant_ViBe	0.6067	0.9969	0.0031	0.3933	0.8644	0.5821	0.7149

The best index values of seven algorithms are shown in bold.

image. Precision is foreground segmentation accuracy, and specificity is background segmentation accuracy. F-measure is the trade-off between recall and precision. FPR is the proportion of background pixels incorrectly marked as foreground, and FNR is the proportion of foreground pixels incorrectly marked as background. PWC represents the error rate. The value range of evaluation indexes above is between 0 and 1. For recall, precision, F-measure, and specificity, the value close to 1 represents better performance of algorithm. But for FPR, FNR, and PWC, the value close to 0 represents better performance.

4.2.2. Comparison of Ant_ViBe and Traditional FTDA.

In order to show the better performance of Ant_ViBe under dynamic background, six typical and widely used traditional background modeling algorithms with relatively good performance (GMM, KNN, PSP-MRF, KDE, Bayesian Background, and ViBe) are selected for comparison experiments conducted on the “Dynamic background” categories in CDnet2014. The detection performances of seven algorithms above are listed in Table 4. In Table 4, Ant_ViBe gets relatively good segmentation performance under dynamic background and performs best in three of the seven indexes, namely, specificity, FPR, and precision, second only to the KNN algorithm in PWC index.

From Section 4.2.1, we know that precision is foreground segmentation accuracy, specificity is background segmentation accuracy, and FPR is the proportion of background pixels incorrectly marked as foreground. Because Ant_ViBe uses a nested foreground and background segmentation mechanism based on sample sets and pheromone, when segmenting, not only the similarity between the current pixel and the local sample set is considered, but also the historical pheromone statistical information of the pixel. This segmentation mechanism ensures that the accuracy of detected foreground pixels is improved. Therefore, Ant_ViBe performs best in the precision index (0.7149). Considering the fact of video sequences that foreground pixels are less than background pixels on the time axis, Ant_ViBe introduces ACCA to make most pixels aggregate towards the background pixels over time. This kind of background modeling based on ACCA ensures that the accuracy of background pixels is improved. At the same time, the proportion of background pixels that are incorrectly marked as foreground is reduced. Therefore, Ant_ViBe performs best in specificity index (0.9969) and FPR index (0.0031). Due to the

improved modeling, update, and segmentation mechanism, the index which represents the error rate is second only to the KNN algorithm. FNR is the proportion of foreground pixels incorrectly marked as background. Because the pixels of the foreground target are transient, and the accumulation of pheromone on the pixels is small, the probability that the foreground pixels are classified as the background (FNR) increases, and the recall is in the middle level.

4.2.3. Comparison of Ant_ViBe and ViBe. Ant_ViBe is an improved algorithm based on ViBe, so the comparison experiments of Ant_ViBe and ViBe on “Camera Jitter” and “Dynamic Background” categories in CDnet2014 with large background jitter and noise interference are conducted (see Table 5). Improved background modeling, updating, and segmentation mechanisms promote better performance of the Ant_ViBe. It can be clearly seen from the table that under Dynamic Background, except the recall, Ant_ViBe has improved in other indexes more than ViBe, especially having remarkable effect in precision and PWC. On the “Camera Jitter” video stream, Ant_ViBe has improved on specificity, FPR, PWC, F-measure, and precision.

4.3. Comparison of Detection and Location Results. After foreground and background segmentation, postprocessing, the Ant_ViBe locates the target through connected domain recognition method (see Figure 12). In connected domain recognition method, we use the same selection criteria to mark the targets whose connected domain area is greater than 100 and less than 30000.

In Figure 12, there is movement of branches in the (b) and (c) columns. The previous three algorithms detected the leaves as targets in different degrees. But Ant_ViBe can accurately locate the cars without being influenced by the dynamic background. In the (d) and (e) columns, GMM and KNN detected the targets with some false detections. ViBe can accurately detect the target, but with larger object region than the actual target. The detection and location of Ant_ViBe is more accurate than the previous three algorithms. In the (f) column, although there is sunlight reflection on the water surface, Ant_ViBe can effectively eliminate the influence of dynamic background, but the small target is not effectively detected. This is a problem that needs further research in the future.

TABLE 5: Comparison of Ant_ViBe and ViBe on “Camera Jitter” and “Dynamic Background”.

Video	Method	Recall	Specificity	FPR	FNR	PWC	F-measure	Precision
Camera Jitter	ViBe	0.7112	0.9694	0.0306	0.2888	4.0150	0.5995	0.5289
	Ant_ViBe	0.6789	0.97709	0.0230	0.3211	3.3552	0.6317	0.6123
Dynamic Background	ViBe	0.7222	0.9896	0.0104	0.2778	1.2796	0.5652	0.5346
	Ant_ViBe	0.6067	0.9969	0.0031	0.3933	0.8644	0.5821	0.7149



FIGURE 12: Comparison of detection and location results.

5. Conclusions

Aiming at the noise, ghost, and target adhesion problems of traditional ViBe, this paper proposed Ant_ViBe, which introduces ant colony clustering algorithm and integrates it into the traditional ViBe framework from the aspects of data modeling, initializing ant colony parameters, constructing pheromone matrix and objective function, and updating ant colony and pheromone, and extended the ViBe based on local modeling to a global modeling algorithm. The experimental results show that the Ant_ViBe eased the ghost and targets adhesion problem, greatly improved the noise resistance under dynamic background, and surpassed the typical algorithms and their fusion algorithms in most evaluation indexes. However, Ant_ViBe needs several frames to establish an accurate and stable background model, which is related to the time characteristics of the ACCA. In addition, Ant_ViBe cannot effectively process shadows. This is the bottleneck of Ant_ViBe, and we will further our work in an extension including the two aspects.

Data Availability

The experimental data used to support the findings of this study have been deposited in the Figshare repository (<https://doi.org/10.6084/m9.figshare.12616796.v1>).

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported in part by the Yunling Scholars Program of Yunnan Ten Thousand Talents Plan and in part by the Donglu Scholars Program of Yunnan University. This work was also supported in part by Double First-Class Construction Project of Yunnan University and was supported in part by the Science Research Fund Project of Yunnan Provincial Department of Education, named “Key Technology Research for Target Detection and Tracking in Complex Monitoring Scenarios” under Grant 2018JS423.

References

- [1] T. Bouwmans, "Traditional and recent approaches in background modeling for foreground detection: an overview," *Computer Science Review*, vol. 11-12, pp. 31-66, 2014.
- [2] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Fort Collins, CO, USA, June 1999.
- [3] Z. Zivkovic and F. van der Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," *Pattern Recognition Letters*, vol. 27, no. 7, pp. 773-780, 2006.
- [4] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, "Real-time foreground-background segmentation using codebook model," *Real-time Imaging*, vol. 11, no. 3, pp. 172-185, 2005.
- [5] O. Barnich and M. Van Droogenbroeck, "ViBe: a universal background subtraction algorithm for video sequences," *IEEE Transactions on Image Processing*, vol. 20, no. 6, pp. 1709-1724, 2011.
- [6] M. Hofmann, P. Tiefenbacher, and G. Rigoll, "Background segmentation with feedback: the pixel-based adaptive segmenter," in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition Workshops*, IEEE, Washington, DC, USA, pp. 38-43, June 2012.
- [7] P.-L. St-Charles, G.-A. Bilodeau, and R. Bergevin, "SuB-SENSE: a universal change detection method with local adaptive sensitivity," *IEEE Transactions on Image Processing*, vol. 24, no. 1, pp. 359-373, 2015.
- [8] M. Braham and M. Van Droogenbroeck, "Deep background subtraction with scene-specific convolutional neural networks," in *Proceedings of the IEEE International Conference on Systems, Signals and Image Processing*, pp. 1-4, Bratislava, Slovakia, May 2016.
- [9] M. Babaee, D. T. Dinh, and G. Rigoll, "A deep convolutional neural network for video sequence background subtraction," *Pattern Recognition*, vol. 76, 2017.
- [10] Y. Wang, Z. Luo, and P. M. Jodoin, "Interactive deep learning method for segmenting moving objects," *Pattern Recognition Letters*, vol. 96, 2016.
- [11] T. Bouwmans, S. Javed, M. Sultana, and S. K. Jung, "Deep neural network concepts for background subtraction: a systematic review and comparative evaluation," *Neural Networks*, vol. 117, 2019.
- [12] H. Sajid and S. C. Sen-Ching, "Universal multimode background subtraction," *IEEE Transactions on Image Processing*, vol. 26, no. 7, 2017.
- [13] P. KaewTraKulPong and R. Bowden, "An improved adaptive background mixture model for real-time tracking with shadow detection," *Video-based Surveillance Systems*, vol. 1, pp. 135-144, 2002.
- [14] M. Van Droogenbroeck and O. Paquot, "Background subtraction: experiments and improvements for ViBe," in *Proceedings of the Computer Vision and Pattern Recognition Workshops (CVPRW)*, Providence, RI, USA, June 2012.
- [15] P.-L. St-Charles, G.-A. Bilodeau, and R. Bergevin, "Subsense: a universal change detection method with local adaptive sensitivity," *IEEE Transactions on Image Processing*, vol. 24, no. 1, pp. 359-373, 2015.
- [16] P.-L. St-Charles, G.-A. Bilodeau, and R. Bergevin, "A self-adjusting approach to change detection based on background word consensus," in *Proceedings of the Applications of Computer Vision (WACV), 2015 IEEE Winter Conference*, Waikoloa, HI, USA, January 2015.
- [17] S. Bianco, G. Ciocca, and R. Schettini, "How far can you get by combining change detection algorithms?" in *Proceedings of the International Conference on Image Analysis and Processing*, Trento, Italy, September 2017.
- [18] G. Hojat, K. Kamran, and M. S. Helfroush, "An improved feature selection algorithm based on graph clustering and ant colony optimization," *Knowledge-Based Systems*, vol. 159, 2018.
- [19] N. Goyette, P. M. Jodoin, F. Porikli et al., "Changetection.net: a new change detection benchmark dataset," in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition Workshops*, Providence, RI, USA, June 2012.
- [20] F. Porikli and O. Tuzel, "Bayesian background modeling for foreground detection," in *Proceedings of the ACM Visual Surveillance and Sensor Network*, San Diego, CA, USA, September 2005.