

An Improved ViBe Algorithm of Moving Target Extraction for Night Infrared Surveillance Video

Zhiqiang Feng^{1,2}, Xiaogang Wang^{1,2*}, Zhongfan Yang^{1,2}, Shaojie Guo^{1,2}, and Xingzhong Xiong^{1,2}

¹School of Automation and Information Engineering, Sichuan University of Science and Engineering,

²Artificial Intelligence Key Laboratory of Sichuan province

Yibin, 644000 China

[e-mail: Jonathan_fzq@163.com; wxg_zf@163.com; yang56535@163.com; fortunate7_gs@163.com;
xzxiong@suse.edu.cn]

*Corresponding author: Xiaogang Wang

*Received August 6, 2021; revised September 6, 2021; accepted November 1, 2021;
published December 31, 2021*

Abstract

For the research field of night infrared surveillance video, the target imaging in the video is easily affected by the light due to the characteristics of the active infrared camera and the classical ViBe algorithm has some problems for moving target extraction because of background misjudgment, noise interference, ghost shadow and so on. Therefore, an improved ViBe algorithm (I-ViBe) for moving target extraction in night infrared surveillance video is proposed in this paper. Firstly, the video frames are sampled and judged by the degree of light influence, and the video frame is divided into three situations: no light change, small light change, and severe light change. Secondly, the ViBe algorithm is extracted the moving target when there is no light change. The segmentation factor of the ViBe algorithm is adaptively changed to reduce the impact of the light on the ViBe algorithm when the light change is small. The moving target is extracted using the region growing algorithm improved by the image entropy in the differential image of the current frame and the background model when the illumination changes drastically. Based on the results of the simulation, the I-ViBe algorithm proposed has better robustness to the influence of illumination. When extracting moving targets at night the I-ViBe algorithm can make target extraction more accurate and provide more effective data for further night behavior recognition and target tracking.

Keywords: moving target extraction, illumination change, adaptive threshold, image entropy, regional growth.

This work was supported in part by National Natural Science Foundation of China (grant no. 61902268), Graduate innovation fund of Sichuan University of Science & Engineering (grant no. y2021072), Sichuan Science and Technology Program (grant no. 2018JY0197, 19ZDZX0037, 2019YFSY0045, 20ZDYF0919, 21ZDYF4052, 2020YFH0124, 2021YFSY0060)

1. Introduction

In the field of public night video surveillance, active near-infrared cameras are the mainstream surveillance equipment, and active near-infrared refers to wavelengths in the range of 0.78 to 2.0 microns. The active near-infrared camera collects reflected infrared light for night imaging. Active infrared imaging at night is susceptible to light interference. When a moving target is close to the infrared camera, it will block part of the infrared light emitted by the infrared fill light, causing illumination changes, resulting in brighter imaging of the moving target and darker background imaging. The closer the moving target is to the infrared camera, the more severe change of illumination. In addition, street lamps and incandescent lamps will also emit infrared light that affects infrared imaging.

Moving target detection is a hot research point in the field of computer vision. Traditional moving target detection algorithms can be divided into temporal difference algorithm [1], optical flow algorithm [2], motion energy map [3], and background subtraction algorithm [4]. The temporal difference algorithm is the simplest moving target extraction algorithm that the foreground contour is extracted by the absolute difference between the images before and after the fixed number of frames. The typical inter-frame algorithm is the three-frame difference algorithm. In the optical flow algorithm, the instantaneous velocity of the pixel movement in the sequence image is used to calculate the movement information of the target between adjacent frames. The motion energy map is used to detect moving objects by describing how the object moves and the position where the motion changes in space. The ViBe algorithm [5], a commonly used background subtraction method, first establishes a background model for the image, then subtracts the background model from the current frame image to obtain the moving target, and continuously updates the background model. The ViBe algorithm has the characteristics of a small amount of calculation and relatively complete extraction of moving targets, but it still has defects such as sensitivity to light and easy to produce ghosts and holes.

Aiming at the phenomenon that the ViBe algorithm is prone to ghosting and sensitive to light changes, many researchers have proposed a series of improvements [6-17]. [6] proposed the use of TOM (Time of map) mechanism to suppress ghost images. [7] proposed a ghost suppression algorithm based on visual saliency. [8] Proposed an estimation and difference of Gaussian (DOG) method to suppress ghost images. [9] proposed a method based on ant colony clustering to suppress ghosts. [10] proposed a background model that uses the Spatio-temporal information of pixels to initialize the background model to suppress ghost images. [11] proposed a background subtraction algorithm based on parameters and samples. [12] proposed an improved ViBe algorithm that introduces Weber's law of adaptive distance threshold to eliminate the influence of illumination changes on target extraction. [13] proposed an image preprocessing method. Establish a brightness reference image, and then use the image intensity normalization method of histogram statistics to adaptively adjust the brightness of the current frame due to changes in illumination back to the brightness level of the reference image. [14] proposed a ViBe moving target detection algorithm based on Lab color space. The improved color difference formula has low sensitivity to changes in illumination. [15] proposed a moving target detection method based on the fusion of color features and XCS-LBP (eXtended Center Symmetric-Local Binary Pattern) based on confidence. [16] proposed an adaptive ViBe algorithm based on dispersion coefficient and spatial consistency factor. [17] proposed an improved ViBe algorithm based on adaptive local median texture (ALMT) features. However, these improved algorithms have inaccurate extraction of moving targets under severe illumination changes and misjudge the background

as the foreground target.

Therefore, in view of the severe illumination changes caused by the moving target approaching the camera in the active infrared imaging video at night, this paper proposes a I-ViBe algorithm for the extraction of moving targets in the night infrared surveillance video. The I-ViBe algorithm mainly includes five parts: 1) Use the ViBe algorithm to extract the foreground of the image of a fixed number of frames in the video interval, and realize the judgment of the illumination change by calculating the change of the pixel point of the foreground image. 2) Use the ViBe algorithm to extract foreground images when there is no light change. 3) The ViBe algorithm with adaptive segmentation threshold is used when the illumination changes are small. 4) When the illumination changes drastically, an improved region growing algorithm based on image entropy is introduced to extract foreground images. 5) Use the visually significant characteristics of the first frame of video to eliminate ghost images. 6) Eliminate false pixels of foreground and noises generated by the ViBe algorithm. Through experimental data testing, the improved algorithm has a good performance in extracting moving targets from night infrared surveillance video with changes in illumination.

2. ViBe Algorithm and Experimental Analysis

2.1 ViBe

The ViBe algorithm is a foreground target extraction algorithm based on a sample model that stores a background sample set including the historical pixel value of the pixel and the pixel value of its domain for each pixel. Then the new pixel is compared with the sampled value in the sample set to determine whether it is a background pixel.

2.1.1 Initialize the Background Model

The ViBe algorithm uses a single frame of video image to initialize the background model that according to the spatial distribution characteristics of similar pixel values of adjacent pixels, it randomly selects its neighboring pixels as its model sampling value, for a pixel. The formula for the initializing background model is defined as

$$M^0(x) = \{v^0[y|y \in N_G(x)]\} . \quad (1)$$

According to (1), $M^0(x)$ is the initialized background model, $v^0(y)$ is the pixel value of the selected pixel in the first frame of the video, $N_G(x)$ is the set of domain pixels of pixel point x in the first frame of video, y is the pixel value of the pixel selected from $N_G(x)$.

2.1.2 Classification of Video Pixel

The detection of the foreground pixel is regarded as a classification problem of the foreground pixel and the background pixel. The background model stores a sample set for each pixel of image, and compares each pixel of the next frame with the sample set to determine whether these pixels are the pixel of the foreground. Let $v(x)$ be the pixel value at the pixel point x in the image, $M(x)$ be the background model of the pixel point x , and the empirical value N is taken as 20. The background model formula is defined as

$$M(x) = \{v_1, v_2, \dots, v_N\} . \quad (2)$$

Define a sphere $S_R[v(x)]$ with $v(x)$ as the center and R as the radius in the two-dimensional color space. If the number of background sample points in the intersection of the ball $S_R[v(x)]$ and the background sample set is greater than or equal to the minimum matching number threshold β , namely

$$\text{count}\{S_R[v(x)] \cap \{v_1, v_2, \dots, v_N\}\} \geq \beta \quad (3)$$

$$\text{or } \text{count}\{\text{dist}\{v(x), v_n(x)\} > R\} \geq \beta. \quad (4)$$

Where the pixel point $M(x)$ is judged to be a background point, otherwise it is a foreground pixel. $n = 1, 2, \dots, n$, $v_n(x) \in M(x)$, $\text{dist}\{v(x), v_n(x)\}$ is the Euclidean distance between the pixel value and the sample of $M(x)$, and R is foreground segmentation threshold.

2.1.3 Updating background model

The ViBe algorithm conducts to update background model with strategies of memoryless updating, random sub-sampling, and spatial consistency. 1) Memoryless update strategy: When the background model needs to be updated, a new background pixel randomly replaces a sample value in the sample set of the pixel. 2) Random sub-sampling update strategy: update the background model according to a certain probability of $1/\varphi$, where φ is the time sub-sampling factor. 3) Spatial consistency update strategy: When a pixel is considered to be the foreground for M consecutive times, the pixel is used to update the background model.

2.2 ViBe Experimental Analysis

A set of night infrared surveillance videos with change in illumination was used to verify the ViBe algorithm. The two frames of (a) and (b) in Fig. 1 have no illumination changes, and the two frames of (c) and (d) have illumination changes. The verification results of the ViBe algorithm are shown in (e), (f), (g), and (h). When there is no illumination change, the ViBe algorithm can completely extract the moving target, as shown in (e) and (f), but when there is a change in illumination, the ViBe algorithm misjudges a large number of background pixels as moving targets, as shown in (g) and (h). Because when there is a change in illumination, the pixel value of a large number of background pixels has changed dramatically. The difference between the pixel value of the pixel that has undergone a huge change and the pixel value in the background model sample is greater than the foreground segmentation threshold R , which makes the ViBe algorithm misjudge the background pixels as the foreground pixels. From the results of moving target extraction, there is a ghosting phenomenon, which is due to the moving target in the first frame of the video, which causes the moving target to be regarded as the background when the background model is established. In addition, there are interference noise points in the extracted image. When the ViBe algorithm is applied to night infrared surveillance video, there are ghosts and noises in the extracted moving target images, and when there are changes in the light in the video, the ViBe algorithm will cause serious misjudgments. This caused serious obstacles to the next step of processing.

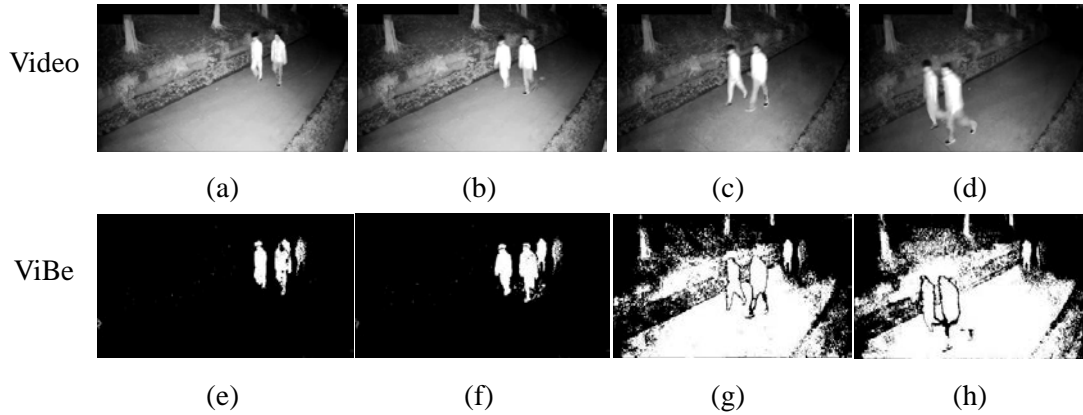


Fig. 1. Using ViBe algorithm to extract moving targets

3. I-ViBe Algorithm

Based on the above experimental verification and analysis, this paper proposes a moving target extraction algorithm I-ViBe that resists changes in illumination. First, establish a background model for the video, and use the ViBe algorithm to extract foreground moving targets in fixed-interval frames. Then, according to the change in the number of foreground pixels of the two frames of images at fixed intervals, it is judged whether there is a change in the illumination of the video frames between the two frames. Finally, different moving target extraction methods are adopted according to the change of illumination, and ghosts and noise are suppressed at the same time. The main steps of the I-ViBe are as follows.

3.1 Establishing an Initial Background Model of Video

The first k frames of the video are used to establish the video background model. For a frame of video, a pixel is randomly selected from the 8 areas of the pixel as a sample in the background model of the pixel. The formula for establishing the background model is defined as follows.

$$M(x) = \{y_i(x) | y_i(x) \in N_{AG}(x) | i = 1, 2, \dots, N\} \quad (5)$$

$$N_{AG}(x) = \frac{1}{k} \sum_{i=1}^k f_i \quad (6)$$

Where f_i is i -th frame image of the input video sequence, the empirical value k in this paper is taken as 20, and N is taken as 20.

3.2 Eliminate Ghost

The I-ViBe is based on the initial background model using the visual saliency characteristics of the image, a method of saliency target detection based on frequency domain information to eliminate ghosts. From the perspective of information theory, an image consists of two parts [18, 19], namely:

$$H(\text{Image}) = H(\text{Innovation}) + H(\text{PriorKnowledge}) \quad (7)$$

Where $H(\text{Innovation})$ is the foreground part that should be paid attention to, and

$H(\text{PriorKnowledge})$ is the background part that should be suppressed. Paper [8] establishes a counter $f_{count}(x)$ for each pixel. If pixel x is the foreground pixel and is salient at the same time, the counter $f_{count}(x)$ is increased by 1. If $f_{count}(x)$ is greater than the threshold T_{count} , the point is judged to be a background point, and the background model is updated.

$$f_{count}(x) = \begin{cases} f_{count}(x) + 1 & , f(x) = 1 \text{ \& } S(x) = 1 \\ 0 & , \text{ else} \end{cases} \quad (8)$$

Where $S(x)$ denotes that the pixel point x has a visual salient characteristic, and T_{count} is an empirical value.

3.3 Judging the Degree of Light Change

According to the actual test situation, the illumination change can be divided into three situations, namely no illumination change, small illumination change, and severe illumination change. The reference object of the illumination change is the background model of the video. Because the ViBe algorithm is sensitive to changes in illumination, the ViBe algorithm is used to sample the foreground objects extracted from the video sequence every 5 frames, and then the illumination change is judged according to the changes in the foreground pixels of two adjacent samples. The judgment process is as follows.

$$C_{fore}^i = \sum_{x=1, y=1}^{M \times N} f_{ViBe}^i(x, y) \quad (9)$$

Where $f_{ViBe}^i(x, y)$ = foreground pixel, $i = 1, 6, 11, \dots$, the size of i -th frame of the video is $M \times N$, C_{fore}^i represents the number of foreground pixels of the sampled frame, and the $f_{ViBe}^i(x, y)$ is the foreground image extracted from the i -th frame of video by the ViBe algorithm.

$$L^{i-5} = \begin{cases} = 2, & C_{fore}^i - C_{fore}^{i-5} \geq \text{max} \\ = 1, & \text{max} > C_{fore}^i - C_{fore}^{i-5} \geq \text{mid}, i = 6, 11, 15K \\ = 0, & \text{else} \end{cases} \quad (10)$$

Where $i > 5$, the value of L^{i-5} represents the lighting change from the $i-5$ th frame of video to the i -th frame video, when the value of L^{i-5} is 2, it means that the lighting changes drastically, 1 means that the light change is small, and 0 means that there is no light change. Max and mid are the empirical thresholds of light change.

3.4 Adaptive Foreground Target Extraction

3.4.1 No Illumination Change

For video frames with no illumination changes, the Vibe algorithm is directly used to extract the foreground target of the video, and the obtained moving target image is denoted as $f_v^t(x)$.

3.4.2 Small Changes in Illumination

For video frames with small lighting changes, the adaptive foreground segmentation threshold R^{i-5} is used on the basis of the ViBe algorithm. The formula of R^{i-5} is defined as

$$R^{i-5} = R \cdot [1 + \mu(C_{fore}^i - C_{fore}^{i-5} - \text{mid})], \quad i = 6, 11, 16, L \quad (11)$$

Where R^{i-5} is defined as the foreground segmentation threshold used from the $i-5$ th frame to the i -th frame, μ , an adaptive factor, is an empirical value, and R is defined as the foreground segmentation threshold of the ViBe algorithm. The value of R^{i-5} is determined by the change of illumination. The larger the change of illumination, the larger the value of R^{i-5} , to realize adaptive selection of the segmentation threshold. The resulting moving target image is denoted as $f_v^t(x)$.

3.4.3 Severe Changes in Illumination

For video frames with severe lighting changes, the method of region growth is adopted to achieve foreground segmentation. The specific implementation steps of the algorithm are as follows.

a) Randomly take n sample points from the background model, and calculate the difference between the pixel points of the current frame and the n sample points, and then take their average as the difference image.

$$f_D^i(x) = \frac{1}{n} \sum_{j=1}^n [f^i(x) - y_j(x)] \quad , \quad y_j(x) = \text{Random}(M(x)), j = 1, 2, \dots, n \quad (12)$$

Where $f_D^i(x)$ is the difference image between the i -th frame image and the background model, j is a random number between 1 and N , the number of j is n , n is an empirical value, $y_j(x)$ is the j th sample point in the background model $M(x)$ corresponding to the pixel point x , and $f^i(x)$ is the pixel value of the pixel point x in the i -th frame image in the video.

b) Gamma transformation is performed on the image $f_D^i(x)$, so that darker pixels are transformed into lighter pixel values [20]. The formula for gamma transformation is defined as

$$s = cr^\gamma \quad (13)$$

C and γ are constants. In this algorithm, c is taken as 1, r is the gray level of the input image, s is the gray level of the output image after transformation, and the image after gamma transformation is denoted as the $f_{DL}^i(x)$.

c) Obtain the region growth threshold h from the image entropy. The one-dimensional information entropy of the image [21, 22] can represent the distribution characteristics of the image gray level, that is, reflect the average amount of information in the image, the formula is as follows.

$$S_f^i = \sum_k^L P_k \log_2 P_k \quad (14)$$

Where L is the number of gray levels of the image, and P_k represents the proportion of pixels with a pixel value of k in the image $f_{DL}^i(x)$ to the total pixel points of the image. The $f_{DL}^i(x)$ image with larger illumination changes and the $f_{DL}^i(x)$ image with smaller changes have a large difference in one-dimensional image entropy. The experimental test results are shown in Fig. 2, where the video frame in (a) is relatively less affected by lighting that the image entropy value is 2.6723 and the video frames in (b), (c), and (d) are greatly affected by illumination, that images entropies are 3.1879, 3.259, and 5.6903, respectively. Therefore, the entropy S_b is introduced as the segmentation threshold that determines the segmentation range of the image. The calculation formula is as follows.

$$h = b \cdot [1 + (S_b - S_f^i) / MAX_s] \quad (15)$$

Where S_b is the empirical value, MAX_S is the maximum image entropy, the maximum entropy of the gray image is 8, and the minimum is 0, b is the reference threshold for region growth segmentation. The growth seed point was selected according to the entropy of the image.

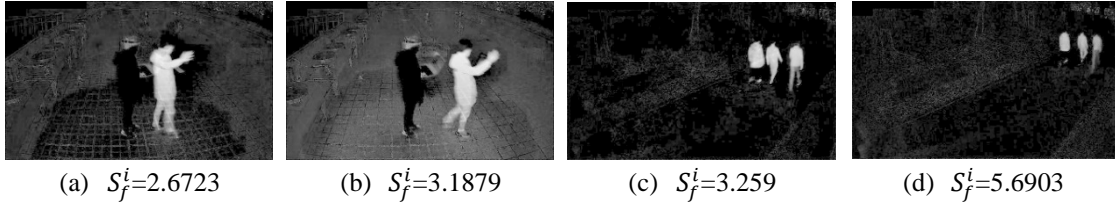


Fig. 2. Comparison chart of one-dimensional image information entropy

d) Select the initial growth seed points according to the entropy value. Perform average filtering on the gamma-transformed image to eliminate interference points and other noises caused by light changes. The filtered image is denoted as the $f_{DLA}^i(x)$, where i is the i -th frame image, and then the growth seed pixels are selected from the image $f_{DLA}^i(x)$ according to the entropy value. If $S_f^i < S_b$, adopt method 1 to select the growing seed pixels. If $S_f^i > S_b$, method 2 is adopted to select the pixel points of the growth seed. In method 1, since the entropy value of the image is relatively small, the intensity of the illumination change of the image is relatively small, and the gray levels distribution of the image after the gamma transformation are not uniform that the gray levels of the entire image are distributed in two areas, which are the gray values near the background pixels and the moving target pixels. The gray value of the moving target pixels are relatively large, so the pixel value with the largest number of pixels is found from the distribution area divided by the larger gray value in $f_{DLA}^i(x)$. Then select H pixels from the pixel value and neighboring pixel values as initial seed points. In method 2, the entropy value of the image is relatively large, that is, the intensity of the illumination change is relatively large, it is small that the pixel value of the background pixels of the image after gamma transformation and the pixel value of the target point, and the pixel value of the moving target Larger. Therefore, in the $f_{DLA}^i(x)$, starting from the pixel value of 250 to the direction of 0, search for H pixel points as initial seed points to avoid interference points with large pixel values.

e) Using the region growing algorithm [23] to extract the foreground target of the image. Region growth, a bottom-to-up segmentation algorithm, starts to grow from a group of seed pixels with the same attributes, and merges all the pixels with the same attributes as the seed pixels. As long as the appropriate seed pixels and segmentation threshold are selected, a better target advance effect can be achieved. The foreground image segmented by the region growing algorithm is denoted as $f_{out}^i(x)$.

f) Eliminate false foreground pixels and noises caused by the algorithm [7]. When the Vibe updates the background model, it misjudges the front spot as a background spot and generates a wrong spot; and the camera shake and noise can also cause noise. The foreground image $f_v^t(x)$ extracted by the ViBe algorithm is filtered, that is, the critical value m is introduced according to the number of foreground pixels contained in the 24 areas of each pixel. If the total number of foreground pixels contained in the 24 areas of a certain foreground pixel is less than m , then it is judged that the point is a foreground pixel or noise. The final foreground image is denoted as $f_{out}^i(x)$, and the empirical value m is 3 after experimental testing.

4. Experimental Results

Next, we will verify and analyze the I-ViBe algorithm by constructing outdoor night infrared surveillance video data. It is mainly divided into the following parts. The judgment of the degree of influence of the video on the change of light is verified, the effect of the algorithm in suppressing ghosts is verified, and the extraction effect and noise suppression effect of the running target when subjected to different degrees of light and shadow are verified.

The experiment was carried out on a hardware device with an i7-4510 CPU and 8G memory, and the experiment running environment was MATLAB2016b. The experimental test data uses the actual collected data, and the collection equipment uses the Tyrannical wireless network camera, BTA-c6625JA-WBS3-4MM. The pixel size of the experimental test video is 360*640. Due to the built-in parameter settings of the capture device, the time watermark of the recorded video will appear in the upper left corner of the recorded video, so pre-processing was performed before the test, namely, the watermark was replaced with pixels value of 0.

The parameters in the I-ViBe algorithm are defined as follows, namely, using the average value of the first 20 frames of video images to establish the background model, the number of background samples $N=20$, the foreground segmentation threshold $R=20$, the small matching number threshold $\beta=2$, the time sub-sampling factor $\phi=16$, the ghost point count $T_{count}=30$, adaptive factor $\mu=0.0002$, gamma transformation parameter $\gamma=0.4$, random selection of background model sample points $n=5$, benchmark entropy $S_b=3.0$, region growth benchmark segmentation threshold $b=0.3$, The 24-field threshold of false foreground pixels or noise $m=3$, The number of initial seed pixels $M=2000$. The following experiments are conducted to verify and analyze the calculations. The experimental results are the real test results of the I-ViBe algorithm in this paper.

4.1 Experiment of Light Change Judgment

To express the lighting changes of the video frame image, a two-dimensional curve is established in the experiment, the ordinate is the total number of foreground pixels, and the maximum is the total pixels of the image. The abscissa is the number of frames of the video, and the two-dimensional graph is established by the ratio of the abscissa and the ordinate to 1:1000. Through experimental analysis, the most suitable basis for judging the serious change of light is that the slope of the curve is greater than 60° , and the light change between the slope of 60° and 30° is the most appropriate. When the slope is 60° , the value is increased by 8660 pixels (max=8660) every 5 frames, and when the slope is 30° , the value is increased by 2000 pixels (mid=2000) every 5 frames.

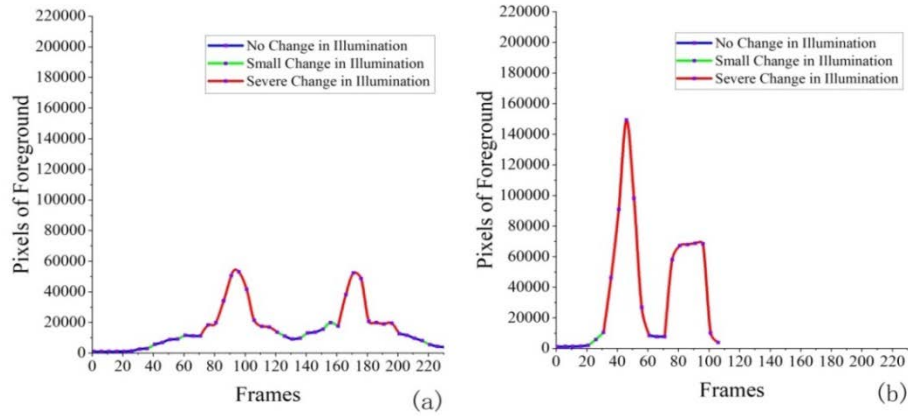


Fig. 3. Illumination curve of video 1 and video 2

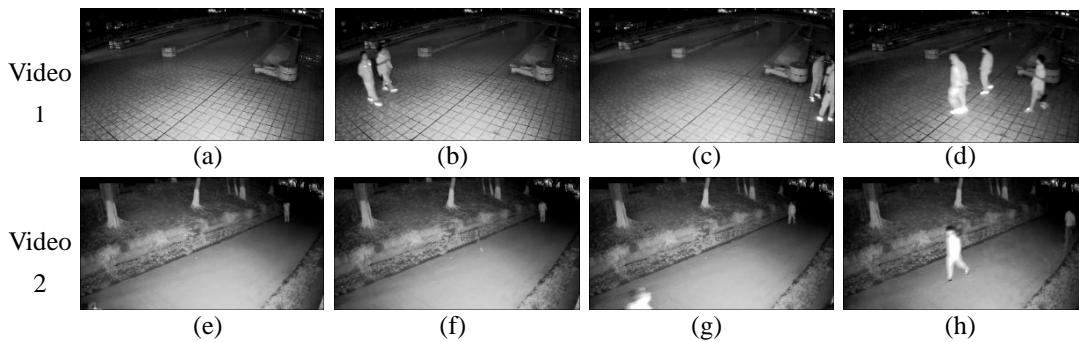


Fig. 4. Comparison of lighting changes between video 1 and video 2

In **Fig. 3**, (a) is the illumination change judgment diagram of video 1, and (b) is the illumination judgment diagram of video 2. In **Fig. 4**, (a), (b), (c), (d) are the first, 55th, 125th, and 170th frames of video 1, respectively. In **Fig. 4**, (e), (f), (g), (h) are the first, 18th, 30th, and 85th frames of video 1, respectively. (b) and (f), (c) and (g), and (d) and (h) respectively correspond to three situations of no light change, small light change and severe light change. There are two situations in which it will be judged that the illumination change is small, which are the sudden appearance of multiple or large moving targets in the video. The pixel point of the background pixel value changes slightly and cannot be detected by the naked eye. When the illumination changes sharply, the pixel value of the video background pixel can be seen by the naked eye that the pixel value has undergone a huge change.

4.2 The Experiment of Ghost Elimination

The ghost image appears in the moving target extraction image because the moving target appears in the first frame of the image used for the ViBe algorithm to establish the background model. The ghost elimination experiment was carried out on video 3. Because the I-ViBe algorithm only uses the ViBe to extract the motion foreground in the case of non-light changes, and the algorithm of suppressed ghost is performed when the ViBe algorithm is used, the test video 3 is selected as the video without light changes. When the light changes drastically, only when the moving target overlaps the ghost image will it have an impact, and the overlapped

part will be treated as the background.

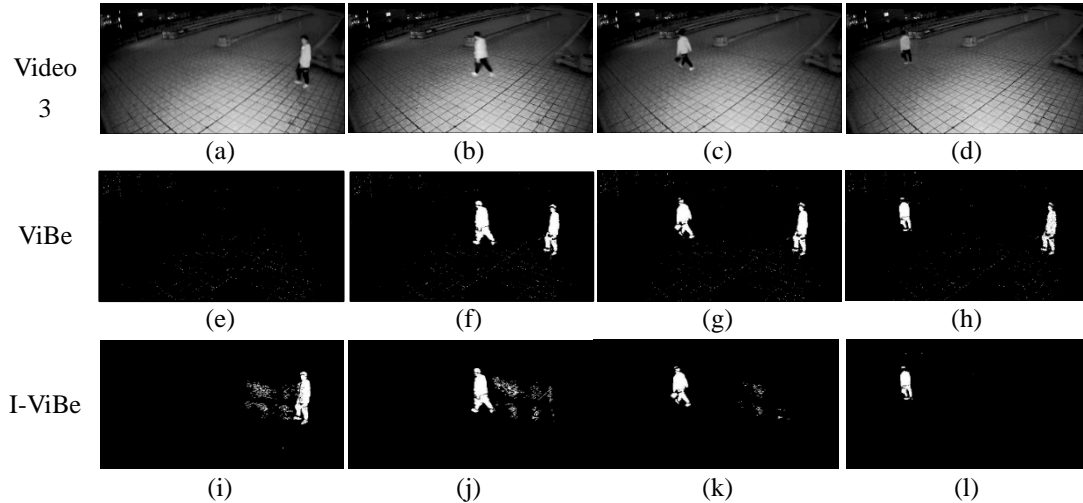


Fig. 5. Experiment of ghost elimination

Judging from the experimental results in **Fig. 5**, the I-ViBe algorithm has a good effect on the suppression of ghosts. (a), (b), (c), and (d) are the first and 24th, 43th, and 60th frames of video 3, respectively. (e), (f), (g), (h) are the test results of the the ViBe algorithm, and (i), (j), (k), (l) are the results of the I-ViBe algorithm. There is no moving target in (e) because the ViBe algorithm uses the first frame to build the background model, and the moving target in the first frame is regarded as the background. (f), (g), (h) almost contain the complete ghost image, because the ViBe algorithm often requires more than 100 frames to eliminate the ghost. The result of (i) is that because the average value of the first 20 frames of images is used to establish the background model, the ghost area generated is relatively large. In (k), it can be clearly seen that most of the ghost images have been eliminated, and (l) shows that the ghost images have been completely eliminated. By comparison, it is easy to know that the I-ViBe algorithm has a better ability to suppress ghosts.

4.3 The Experiment of Moving Target Extraction

The illumination change experiment was carried out in video 2. The illumination change judgment curve of video 2 is shown in (b) of **Fig. 3**. **Fig. 6** shows the result of extracting moving targets in video 2 using the ViBe algorithm, [13], [14], [18] and the I-ViBe algorithm. (a), (b), (c), and (d) are the 26th, 43th, 65th and 89th frame images of video 2, respectively. (a) shows a small change in light, (b) shows a sharp change in light, (c) shows no change in light, and (d) shows a sharp change in light. In (a1), since a moving target suddenly enters the video shooting area, the current frame is judged to have a small change in illumination. From the perspective of the final moving target extraction effect, the algorithm [18] performs the worst, and the other algorithms perform better. (a2) and (a4) are situations where the illumination changes drastically. Judging from the extracted results, the I-ViBe algorithm performs best. Although the algorithm [18] overcomes the illumination change to a certain extent, it only extracts the contours of part of the moving target, and there is a large amount of interference noise at the same time. This is because the algorithm [18] is only suitable for high-speed moving target extraction. ViBe, [13], and [14] perform poorly when the light changes

drastically. (a3) is the case of no light changes. From the extraction results, the I-ViBe algorithm, algorithm [14], and ViBe algorithm perform best, and algorithms [13] and [18] perform poorly.

From the comparison results of the algorithm test, it can be seen that the I-ViBe algorithm is better than other algorithms in the three situations of no light change, small light change, and drastic light change.

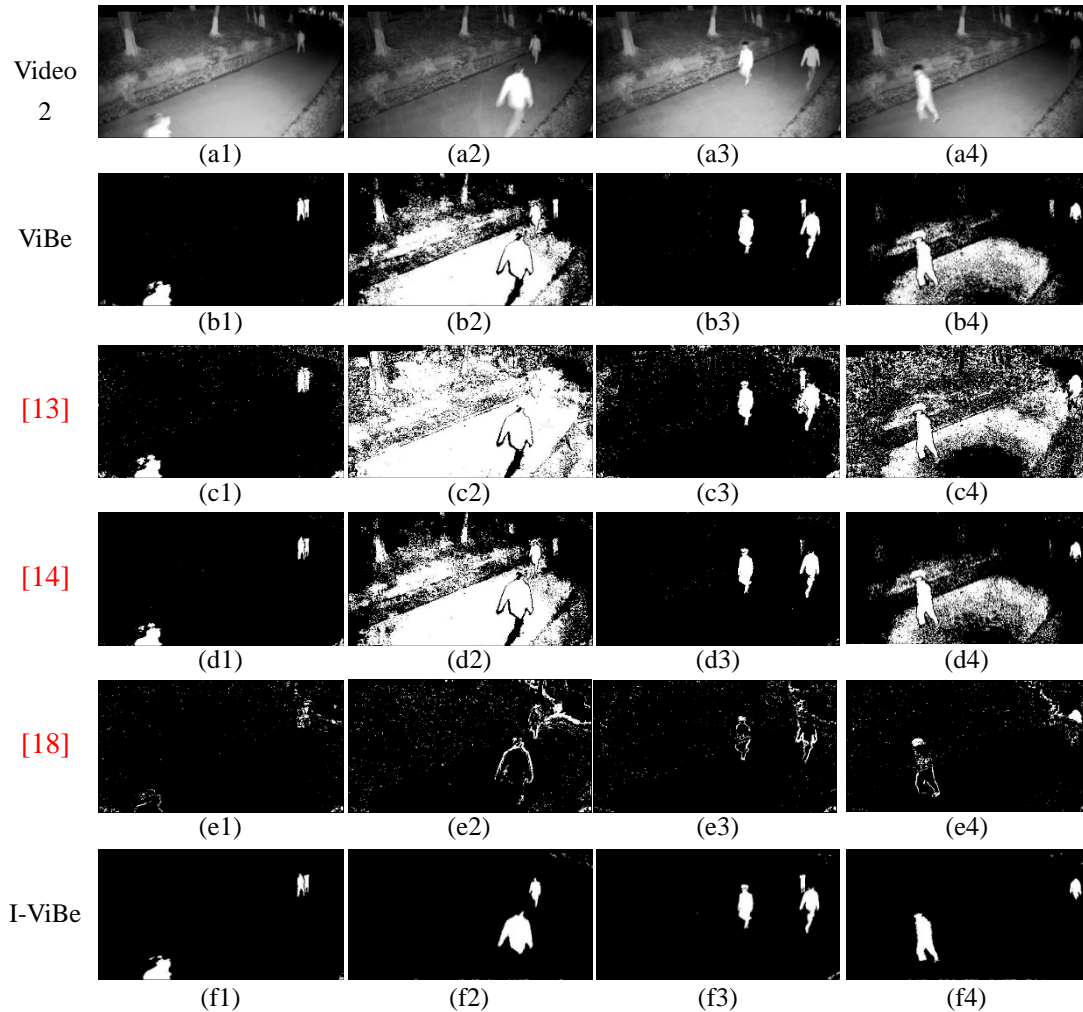


Fig. 6. Comparison of moving target extraction results

4.4 Evaluate Metrics

For moving target detection, it can be regarded as a two-classification problem. Percentage of Correct Classification (PCC), Precision, and Recall can be used evaluate the performance of the algorithm. The higher value of the evaluation metrics presents the better quality of algorithm. The following four parameters are used in the evaluation metrics: the number of true positives (TP) is correctly identified as positive samples, the number of true negatives (TN) that are correctly identified as negative samples, the number of false negatives (FN) that are incorrectly identified as negative samples and the number of false positives (FP) that are incorrectly identified as positive samples. The formulas of the PCC, the Precision, and the

Recall are defined as follows:

$$PCC = \frac{TP+TN}{TP+FP+TN+FN}, \quad (16)$$

$$Precision = \frac{TP}{TP+FP}, \quad (17)$$

$$\text{and } Recall = \frac{TP}{TP+FN}. \quad (18)$$

Video 2 is tested using evaluation metrics. The three evaluation metrics are shown in Fig. 7 (a), (b), and (c) correspond to the three parameters of the PCC, Precision and Recall, respectively. In (a), the I-ViBe algorithm outperforms ViBe, [13], [14] and [18] when the video has severe lighting changes, and [13] has the worst performance. In (b), because the moving target in the first 20 frames of video 2 only has slight movement, the effect of the five algorithms is poor. The lighting changes in the 31st to 61st frames and 71 to 101 frames of the video are huge. The I-ViBe algorithm has good performance, with an average accuracy of more than 0.75, and algorithm [13] and algorithm [18] perform the worst. In (c), when the video illumination changes greatly, the Recall metric of the I-ViBe algorithm performs relatively poorly. The reason is that when the illumination changes drastically, the I-ViBe algorithm cannot accurately extract the edge contour of the moving target. Algorithm [18] has the worst performance, because [18] only extracts the contour of the moving target.

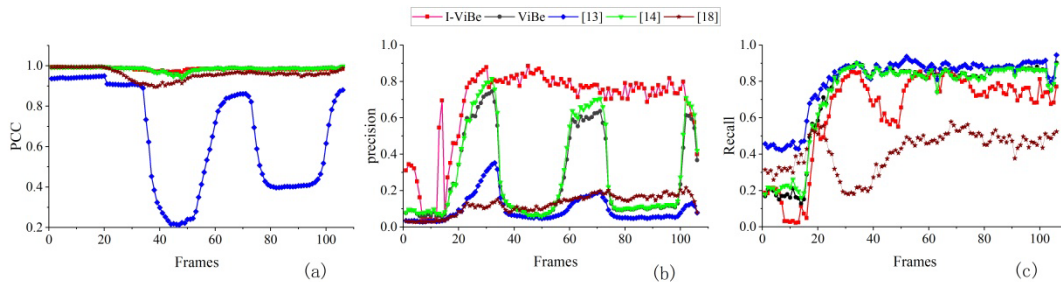


Fig. 7. Evaluate Metrics

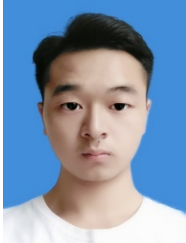
5. Conclusion

In this paper, given the characteristics of active infrared imaging cameras, at night, when the moving target approaches the camera, it will cause dramatic changes in illumination. We propose a moving target extraction algorithm (I-ViBe) based on infrared surveillance video. In this paper, the I-ViBe algorithm is applied to the moving target extraction of infrared video surveillance, and three different moving target extraction methods are adopted for the situation of light changes. When there is no illumination change, the ViBe is used to calculate, an adaptive segmentation threshold is introduced when the illumination change is small, and an improved algorithm based on region growth is proposed by introducing image entropy when the illumination changes sharply. Given the ghost defect in the ViBe algorithm, the visual saliency characteristic of image is used to suppress the ghost. Given the misjudgment of the ViB algorithm and the influence of noise, 24-domain filtering is used. The experimental analysis results show that, compared with the ViBe algorithm, the I-ViBe algorithm in this paper is more effective in extracting moving targets from night infrared surveillance video.

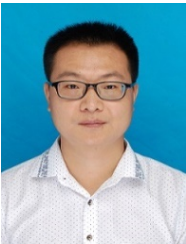
References

- [1] J. Lipton, H. Fujiyoshi, R. S. Patil, "Movingtarget classification and tracking from real-time video," in *Proc. of the 4th IEEE workshop on applications of computer vision WACV'98 (Cat. No.98EX201)*, pp. 8-14, 1998. [Article \(CrossRef Link\)](#)
- [2] D. Sun, S. Roth, M. J. Black, "Secrets of optical flow estimation and their principles," in *Proc. of the IEEE computer society conference on computer vision and pattern*, pp. 2432-2439, 2010. [Article \(CrossRef Link\)](#)
- [3] A. F. Bobick, J. W. Davis, "The recognition of human movement using temporal templates," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 23, no. 3, pp. 257 - 267, Mar. 2001. [Article \(CrossRef Link\)](#)
- [4] I. Haritaoglu, D. Harwood, L. S. Davis, "W/sup 4: real-time surveillance of people and their activities," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 8, pp. 809 - 830, Aug. 2000. [Article \(CrossRef Link\)](#)
- [5] O. Barnich, M. V. Droogenbroeck, "ViBe:A Universal Background Subtraction Algorithm for Video Sequences," *IEEE Transactions on Image Processing*, vol. 20, no. 6, pp. 1709-1724, Jun. 2011. [Article \(CrossRef Link\)](#)
- [6] W. Huang, L. Liu, C. Yue, et al., "The moving target detection algorithm based on the improved visual background extraction," *Infrared Physics & Technology*, vol. 71, pp. 518-525, Jul. 2015. [Article \(CrossRef Link\)](#)
- [7] M. Shaowen, D. Xinpu, W. Shuai, et al., "Moving object detection algorithm based on improved visual background extractor," *Acta Optica Sinica*, vol. 36, no. 6, pp. 204-213, Jan. 2016. [Article \(CrossRef Link\)](#)
- [8] W. Yuanbin and R. Jieying, "An Improved Vibe Based on Gaussian Pyramid," in *Proc. of the 4th International Conference on Control and Robotics Engineering (ICCRE)*, pp. 105-109, 2019. [Article \(CrossRef Link\)](#)
- [9] Y. Yue, D. Xu, Z. Qian, H. Shi, et al., "Ant_ViBe: Improved ViBe Algorithm Based on Ant Colony Clustering under Dynamic Background," *Mathematical Problems in Engineering*, vol. 2020, pp.1-13, Aug. 2020. [Article \(CrossRef Link\)](#)
- [10] L. Liu, G. Chai, Z. Qu, "Moving target detection based on improved ghost suppression and adaptive visual background extraction," *Journal of Central South University*, vol. 28, no. 3, pp.747-759, Apr. 2021. [Article \(CrossRef Link\)](#)
- [11] X. Liu, T. Zhong, D. Fu, "Robust Compositional Method for Background Subtraction," in *Proc. of the 12th International Conference on Control Automation Robotics & Vision (ICARCV)*, pp. 1419-1424, 2012. [Article \(CrossRef Link\)](#)
- [12] H. Guang, J. Wang, C. Xi, "Improved visual background extractor using an adaptive distance threshold," *Journal of Electronic Imaging*, vol. 23, no. 3, pp. 036005, Nov. 2014. [Article \(CrossRef Link\)](#)
- [13] H. Wang, Q. Wang, Y. Li, et al., "An illumination-robust algorithm based on visual background extractor for moving object detection," in *Proc.of the 10th Asian Control Conference (ASCC)*, pp. 1-6, 2015. [Article \(CrossRef Link\)](#)
- [14] Z. Peng, F. Chang, W. Dong, "Vibe Motion Target Detection Algorithm Based on Lab Color Space," in *Proc.of the Chinese Conference on Image and Graphics Technologies*, pp 45-54, 2015. [Article \(CrossRef Link\)](#)
- [15] J. Cao, S. Zhao, X. Sun, et al., "Algorithm of moving object detection with illumination robustness based on confidence," in *Proc.of the IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, pp. 1058-1062, 2017. [Article \(CrossRef Link\)](#)
- [16] Q. Zhang, W. Lu, C. Huang, et al., "An adaptive vibe algorithm based on dispersion coefficient and spatial consistency factor," *Automatic Control and Computer Sciences*, vol. 54, no. 1, pp. 80-88, Mar, 2020. [Article \(CrossRef Link\)](#)

- [17] Y. Zhang, W. Zheng, K. Leng, et al., "Background Subtraction Using an Adaptive Local Median Texture Feature in Illumination Changes Urban Traffic Scenes," *IEEE Access*, vol. 8, pp. 130367-130378, 2020. [Article \(CrossRef Link\)](#)
- [18] X. Hou, L. Zhang, "Saliency detection: A spectral residual approach," in *Proc. of the 2007 IEEE Conference on computer vision and pattern recognition*, pp. 1-8, 2007. [Article \(CrossRef Link\)](#)
- [19] Y. Ren, J. Zhou, Z. Wang and Y. Yan, "An Improved Saliency Detection for Different Light Conditions," *KSIIT Transactions on Internet and Information Systems*, vol. 9, no. 3, pp. 1155-1172, 2015. [Article \(CrossRef Link\)](#)
- [20] F. Kallel, A. B. Hamida, "A new adaptive gamma correction based algorithm using DWT-SVD for non-contrast CT image enhancement," *IEEE transactions on nanobioscience*, vol. 16, no. 8, pp.666-675, Nov. 2017. [Article \(CrossRef Link\)](#)
- [21] J. Jeng, S. Hsu, Y. Chang, "Entropy improvement for fractal image coder," *Int Arab J Inf Technol*, vol. 9,no. 5, pp. 403-410, Sep. 2012. [Article \(CrossRef Link\)](#)
- [22] M. D. Albuquerque, I. Esquef, A. Mello, Mello, "Image thresholding using Tsallis entropy," *Pattern Recognition Letters*, vol. 25, no.9, pp. 1059-1065, July, 2004. [Article \(CrossRef Link\)](#).
- [23] R. Adams, L. Bischof, "Seeded region growing," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 16, no. 6, pp. 641-647, Jun. 1994. [Article \(CrossRef Link\)](#)



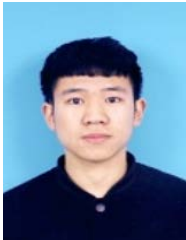
Zhiqiang Feng was born in Nanchong City, Sichuan Province, China in 1996. He is a M.S. candidate in the Key Laboratory of Artificial Intelligence of Sichuan Province, School of Automation and Information Engineering, Sichuan University of Science and Engineering. His current interests are in the area of computer vision and Artificial Intelligence.



Xiaogang Wang was born in Baoji City, Shanxi Province, China in 1984. He received his Ph. D degree from Chongqing University, Chongqing, China. He is currently an associate professor, with Artificial Intelligence Key Laboratory of Sichuan Province, School of Automation and Information Engineering, Sichuan University of Science and Engineering. His current interests are in the area of wireless sensor network and security, IoT, and Artificial Intelligence.



Zhongfan Yang was born in Dazhou City, Sichuan Province, China in 1997. He is a M.S. candidate in the Key Laboratory of Artificial Intelligence of Sichuan Province, School of Automation and Information Engineering, Sichuan University of Science and Engineering. His research interests lie in 3D computer vision and point cloud processing.



Shaojie Guo was born in 1997 in Sanmenxia City, Henan Province, China. He received his bachelor's degree from Henan Polytechnic University in Henan, China. He is a M.S. candidate in the Key Laboratory of Artificial Intelligence of Sichuan Province, School of Automation and Information Engineering, Sichuan University of Science and Engineering. His current interests are in the area of target defect detection, artificial intelligence.



Xinzhong Xiong was born in Guangan City, Sichuan Province, China in 1971. He received his Ph. D degree from the University of Electronic Science and technology, Sichuan, China. He is currently a professor, with Artificial Intelligence Key Laboratory of Sichuan Province, School of Automation and Information Engineering, Sichuan University of Science and Engineering. His current interests are in the area of wireless communication, IoT, Artificial Intelligence.