Journal of Computers Vol. 31 No. 5, 2020, pp. 99-111
doi:10.3966/199115992020103105008

Video Region of Interest Extraction Algorithm Based on Improving Visual Background Extraction Model

Ren-Jie Song, Yuan-Dong Zhang*

School of Computer, Northeast Electric Power University, Jilin 132012, China
{1939811347, 1406632033}@qq.com

Received 13 April 2019; Revised 7 August 2019; Accepted 12 October 2019

Abstract. Aiming at the problems that the visual background extraction model is difficult to adapt to dynamic scenes, appears the target holes easily, spends much time to eliminate ghosts, and pixels at the junction of the foreground and the background are prone to retransmit wrong information. A video region of interest extraction algorithm based on visual background model improved is proposed. In the foreground segmentation phase, according to the spatio-temporal correlation, the global quantity threshold is adjusted by using the mean absolute deviation adaptively, and a dynamic quantity threshold is obtained. At the same time, according to the contribution rates of the scene regions to the human eyes, the variance is used to adjust the global distance threshold adaptively, and the dynamic distance threshold is obtained, and the propagation probability of misdetecting information is eliminated and the detection accuracy is improved, which adapts to the changes of dynamic scene, and the video region of interest is extracted. In the update phase of background model, the complexity of regional scene is used to adjust the update periods dynamically to eliminate the target holes and accelerate the elimination of ghosts effectively. The results show that the proposed algorithm can make up the shortcomings of the VIBE algorithm, and extract the region of interest of video with higher precision, which is suitable for the extraction of the region of interest in the dynamic video scenes.

Keywords: adaptive threshold, dynamic update period, region of interest extraction, visual background extraction

1 Introduction

Moving target detection algorithm is an important application in the field of video surveillance [1], the purpose is extracting the motion regions from the video sequences. At present, the moving target detection algorithms used commonly include the inter-frame difference algorithm [2], the optical flow algorithm [3], and the background modeling algorithm [4]. The inter-frame difference algorithm detects the moving targets by the differences between the adjacent frames, but the influences of the target speed and the time interval are easy to generate target holes and fake targets. The optical flow algorithm detects moving targets by calculating optical flow fields, but the algorithm have a weak ability to suppress noise, and the computational complexity is also higher, which is difficult to meet real-time requirements. The background modeling algorithms compare the video sequences with the background model to detect the moving targets, and the real-time performance is higher than the inter-frame difference algorithm and the optical flow algorithm, and the detection targets are more accurate. The existing background modeling algorithms can be divided into two categories roughly: (1) parametric background models, the type of models build a parametric model for each pixel to estimate the background, such as the mixed Gaussian model (GMM), which is based on statistical information of pixel samples, each pixel model in the model is obeying multiple Gaussian distributions in the space-time domain, they can be modeled in complex background, but the algorithm has high complexity, low real-time performance and high false alarm rate in the transformation scenarios. (2) Non-parametric background models, the type of models rely on

* Corresponding Author

historical pixel values to build models to estimate the background, such as the kernel density estimation algorithm, which relies on historical pixels to construct a background model, which avoids the introduction of prior knowledge, thus, the sample data can be approximated to the greatest extent, but the accuracy of the algorithm is low when the speed of targets detected is changing, and the algorithm is sensitive to noise. For example, the codebook model (CB) uses a quantitative sequences to construct a background model by training samples, the algorithm has an advantage of being simple in calculation aspect and strong in real-time aspect, but which is sensitive to illumination, and the accuracy is low in complex scenes. In addition, some scholars have also proposed some other background models, such as multi-frame averaging algorithm [5], the algorithm belongs to non-parametric background models, the model is simple and easy to be calculated, but the effect which processes dynamic scene video is poor; SIFT flow algorithm [6], the algorithm belongs to the non-parametric background models, and the model has a better effect on the dynamic scene, but the computational complexity is higher; histogram model under rough set framework (histon roughness Index, HRI) [7], which belongs to the non-parametric background models, and the model can overcome the changes of illumination well, but the real-time performance is poor; random sample consensus (RANSAC) [8], which belongs to the parametric background models, the model has high reliability and stability, however, along with the number of iterations increasing, the complexity of the algorithm will increase, which is not conducive to practical applications. These models improve the moving target detection effects compared to the traditional moving target detection algorithms, but the computational cost is greater, the real-time performance is also weaker, limiting the demand for practical applications. In 2009, Barnich et al. proposed a pixel-based non-parametric random sample model(visual background extractor, VIBE) [9], the algorithm is one of the most popular algorithms in the current target detection fields, which is characterized by simple background model and lower computational complexity. It has higher precision to detect the video sequences under a single background, and can detect moving targets quickly, which has advantages of strong robustness and real time. However, the VIBE algorithm uses fixed threshold and global update period, which leading to appear target holes in complex scenes of video easily, unable to adapt to dynamic scene changes, and the detection accuracy is low. In the update phase of background model, the model takes a long time to eliminate ghosts, and the pixels at the junction of foreground and background are prone to produce error retransmission of information, affecting the real time and accuracy of the detection results.

The human visual system (HVS) characteristics [10-11] show that the human eyes have different attention to different regions of images, and the texture-rich regions or the moving object in the image intensely are the region of interest (ROI), which are more likely to attract the attention of the human eyes, but less attention to the region-non-interest (RONI) that are the background regions. The region of interest [12] extraction is widely used in medical imaging, content-based image retrieval, video compression, machine vision and other fields. At present, the algorithm of extracting region of interest can be divided into two categories [13]. One is that the observers need to observe each image which obtains the region of interest by a human-computer interaction manner, such as the eye tracker acquires the region of interest in the video according to the human eyes; the other is using the algorithm to obtain the region of interest. At present, the algorithms of extracting the region of interest include visual attention model, image feature extraction algorithm, neural network algorithm and so on. However, most of the traditional algorithms focus on the extraction of region of interest in the field of images, there is relatively little research on the extraction of region of interest in the field of videos.

This paper is based on the idea of combining the extraction of region of interest in the field of videos with the VIBE algorithm, which has improved the original VIBE algorithm, and proposes a video region of interest extraction algorithm based on the improved VIBE algorithm. The proposed algorithm improves the original VIBE model by using the proposed adaptive threshold formulas and the designed dynamic update period, which overcomes the problems with regard to the VIBE algorithm that is prone to emerge target holes in complex video scenes, takes a long time to eliminate ghosts, and pixels at the junction of foreground and background are prone to produce error retransmission of information. And the region of interest of the video are extracted. The accuracy and robustness of the proposed algorithm are higher in dynamic scene.

The paper is organized as follows: in “introduction” section, the related definition and basic results used throughout the paper, are introduced; in “the VIBE algorithm analysis” section, the steps and problems of the VIBE algorithm are introduced; in “the improved VIBE algorithm” section, the adaptive

threshold formulas and dynamic update designed are proposed, and the experimental demonstrations are carried out in turn; in "experiment" section, the simulation results under the test sequences are provided, to corroborate our theoretical results and evaluate the accuracy and robustness of our proposed scheme; in "Conclusion" section, the main conclusions drawn from the present work are outlined and our ideas in the next work are proposed.

2 VIBE Algorithm Analysis

2.1 VIBE Algorithm Principle

The VIBE algorithm is a background modeling algorithm based on sample consistency. The algorithm applies the random selection mechanism and neighborhood propagation mechanism to the process of establishment and update of the background model for the first time [14]. The basic idea of the algorithm is that establishes a background model for each pixel according to the random selection mechanism. In the period of foreground detection, the pixels detected are compared with the pixels where are in the background model to segment the foreground; The spatial consistency is fully utilized in the period of background update, and the pixels which have been judged to be the background points are updated into the background model. At the same time, the pixels which belong to the neighborhood are updated to the neighborhood according to the neighborhood propagation mechanism.

2.2 VIBE Algorithm Steps

The VIBE algorithm is mainly divided into three parts: Initialization of the background model, Foreground detection and Background model update. The steps of the VIBE algorithm in detailed are as follows:

(1) Initialization of the background model

The initialization of the background model in the VIBE algorithm is done in the first frame of the image. The algorithm assumes that the pixel and its' neighboring pixels obey the same time domain distribution, and in the m neighborhoods of each pixel $v(x)$, where extracts N points v_i randomly to form the background model $M(x)$ of the current point in the equal probability, $M(x)$ is defined as:

$$M(x) = \{v_i \mid i = 1, 2, \cdots N\} . \tag{1}$$

$$v_i = (v_i^R, v_i^G, v_i^B) . \tag{2}$$

Where equation (1) represents the background model sets of the current pixel point, equation (2) represents that the pixel points of the VIBE algorithm are based on the RGB color space, and the VIBE algorithm extracts 20 points (N=20) from the eight neighborhood areas (m=8) to initialize the background model in the equal probability randomly. If the first frame of video has moving target areas in the initialization process of the VIBE algorithm, the phase of foreground segmentation is prone to produce ghosting.

(2) Foreground detection

The foreground detection phase in the VIBE algorithm is actually a classification process to pixel points. To measure the similarity between the current pixel point and the pixel points in the corresponding background model, defining $v(x)$ as the center of the sphere $S_R(v(x))$ and R as the radius of the sphere $S_R(v(x))$. The processes of comparing the current pixel point with the background model are as shown in Fig. 1. When the number of background pixels falling in the sphere satisfies a certain numbers I, it is judged that the pixel is a background (BG) pixel point, otherwise, it is judged as a foreground (FG) pixel point, and K indicates the detection result, then the foreground detection conditions are equation (3):

$$K = \begin{cases} *\{D(v(x), M(x)) \le R\} \ge \beta, BG \\ *\{D(v(x), M(x)) \le R\} < \beta, FG \end{cases} . \tag{3}$$

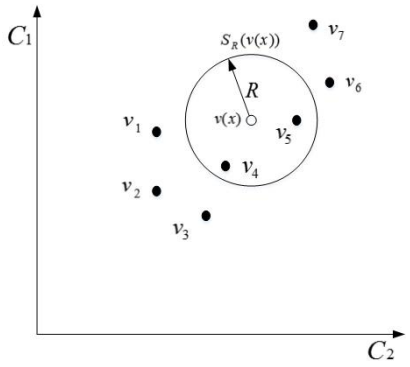


Fig. 1. VIBE background model

Where D is the Euclidean distance between the current pixel and the background pixels, *{*} indicates the number which meets the conditions, β indicates the global quantity threshold, R indicates the global distance threshold, and the global distance threshold is 20 and the global quantity distance threshold is 2 in the VIBE algorithm. If the distance between the current pixel and the pixel in the background model is less than or equal to R, the number of which is more than or equal to β, then the current pixel is judged to be the background pixel point, otherwise, the current pixel is the foreground pixel point.

The VIBE algorithm uses a fixed threshold to segment foreground in the dynamic scenes with more false detection results, resulting in inaccurate detection results.

(3) The update of the background model

In the update phase of background model, the VIBE algorithm uses a conservative update strategy and a spatial information neighborhood diffusion mechanism. When the pixel $v(x)$ is judged as a background pixel, a pixel of the corresponding background is replaced by $1/\phi$ (called time sampling factor) probability in the model set $M(x)$, when the pixels detected are as the background, which are updated into the model. Meanwhile, in order to maintain the spatial consistency of the pixels, selecting a pixel from the m neighborhood (m=8) of the current pixel randomly to replace the pixel where is in the corresponding background model according to the current update method. The update probability of the sample in the VIBE algorithm is independent of time. The probability of the samples reserved in the background model with the changes of time is as follows:

$$P(t,t+dt)=e^{-\ln(\frac{\gamma-1}{\gamma})dt}. \tag{4}$$

Where t is the entry time of the samples and dt is the reserved time of the samples. We can see that the probability of sample reserved $P(t,t+dt)$ decays exponentially with time from equation (4). The method can reserve the useful samples as long as possible. The update period of the background model is determined by the random number extracted. Selecting a value μ from $(0,K)$ in the equal probability randomly, when μ is equal to the predetermined values, updated; otherwise, not to update, the predetermined values is set 0 usually.

The update method produces error retransmission of information on the pixels, which are at the junctions of foreground and background, and causing false detection results on follow-up work. Although the update method suppresses ghosting to a certain extent, it takes more time to eliminate ghosting, which reduces the efficiency to extract moving targets.

3 Improved VIBE Algorithm

In view of the existing problems in the VIBE algorithm, this paper improves the foreground segmentation phase and the background update phase, and proposes a video region of interest extraction algorithm based on improving the visual background model.

3.1 The Determination of Adaptive Quantity Threshold Based on Space-time Domain

In the foreground segmentation phase of the VIBE algorithm, the global quantity threshold is used for foreground segmentation. This method has better effect on video sequences with little changes of scenes, but it is not ideal for video sequences segmentation in dynamic scenes, and it is prone to miss targets, which accompanied by noise points. Chen et al. [15] introduces the $OTSU$ algorithm to obtain the adaptive segmentation threshold to perform segment foreground. However, the introduction of the $OTSU$ algorithm which is used to calculate the threshold increases the complexity of the background model and the segmentation effect is not accurate enough. Studies have shown that [16], the texture complexity in the video can represent the contribution rates of the image regions to the human eyes. Generally, the higher the texture complexity, the higher the degree of state changes. For the high-texture regions or moving regions of the video sequence, which are more likely to attract the attention of the human eyes. The mean absolute deviation (MAD) is introduced to represent the texture complexity. At the same time, as a result of the certain correlation between adjacent frames in the video sequences, intra-frame pixels have similar distributions with their neighborhoods [17], and the adaptive quantity threshold is used to calculated texture complexity based on space-time domain.

Since the improved algorithm is based on the space-time domain to calculate the adaptive quantity threshold, the first frame uses the original VIBE algorithm to initialize the background model, and the improved algorithm is used to calculate the adaptive threshold from the second frame. In this paper, the video frame is divided into 8×8 blocks for processing. The implementation steps are as follows:

Calculating the MAD of the current pixel and its neighboring pixels, and calculating the MAD of the current pixel and the pixels where are at the corresponding position of the background model frame, and make the mean absolute deviation of the former is $MAD2$, the mean absolute deviation of the latter is $MAD1$, and the calculation formula is as follows :

$$MAD = \frac{1}{M \times N}\sum_{j=0}^{N-1}\sum_{i=0}^{M-1} | p(i,j) - avg | . \tag{5}$$

In equation (5), M, N represent the width and height of the currently video frame block processed respectively, $p(i,j)$ represents the pixel values of the current pixel point, and avg represents the average pixel values of the current frame block.

(1) After obtaining $MAD1$ and $MAD2$, judging the size of the two values to perform the first step of segmentation, and the judgment rules are as equation (6):

$$MAD1 < MAD2 = \begin{cases} true & FG \\ false & BG \end{cases}. \tag{6}$$

Since the first frame of the video uses the VIBE algorithm to initialize the background model, the VIBE algorithm does not distinguish the background regions and the regions with high texture complexity, and the background model contains regions with high texture complexity. At this time, when the foreground segmentation is performed according to the equation (6) in the second frame, when the judgment result is $true$, it is a motion region; otherwise, it is a region with background regions and high texture complexity. When the background model is initialized in the second frame by using the proposed algorithm of this paper, the original background model will only contain the background regions, which does not contain the regions with high texture complexity. When the foreground segmentation is performed according to the equation (6) in the subsequent frames, when the judgment result is $true$, it is a region with a high motion regions and texture complexity; otherwise, it is a background regions.

(2) After obtaining $MAD1$ and $MAD2$, the adaptive quantity threshold η_N is determined by the two values and the second segmentation step is performed. The expressions of η_N are as follows:

$$\eta_N = \frac{MAD1}{MAD1 + MAD2} \times 2 + \frac{MAD2}{MAD1 + MAD2} \times \beta . \tag{7}$$

$$\eta_N = \frac{MAD1}{MAD1 + MAD2} \times 2 + \frac{MAD2}{MAD1 + MAD2} \times 1 \times 50\% + \frac{MAD2}{MAD1 + MAD2} \times 0 \times 50\% . \tag{8}$$

As a result of the global quantity threshold is 2, when the foreground segmentation is performed by the VIBE algorithm, if the current pixel and the number of pixels in the background model meet the global distance threshold, which is greater than or equal to 2, it is determined as the background regions or the regions with high texture complexity, otherwise, it is determined as the motion regions. It can be understood that when the number of satisfying conditions equals 1 or 0, it is a motion region. The probability of occurrence of 1 or 0 is random, and the probability of occurrence of 1 or 0 is equal in this paper. When foreground segmentation is performed in the second frame, when the judgment result is *false*, the adaptive quantity threshold is determined by equation(7), $\beta \in (3,20)$ is a random number here; when foreground segmentation is performed in subsequent frames, when the judgment result is *true*, the adaptive quantity threshold is determined by the equation (8).

3.2 The Determination of Adaptive Distance Threshold Based on Space-time Domain

The VIBE algorithm uses the global fixed distance threshold to classify the pixels. It has better detection effect on the single scene of the video frames. However, the video frames will be detected wrongly under the dynamic scene with complex background, which is difficult to adapt to the changes of the dynamic scenes. In the field of image processing [18], the variance can measure the degree of difference between pixels, and the distances between pixels are also related to the degree of difference between pixel gray values closely. In view of the idea, this paper uses variance as the balance factors to adjust the distance threshold based on the space-time domain. The specific calculation steps are as follows:

(1) Calculating the variance of the current pixel and its neighboring pixel points, and make it σ_2; and calculating the variance of the current pixel point and the pixel points in the corresponding background model block, and make it σ_1.

(2) The VIBE algorithm uses the global fixed threshold for foreground segmentation, the global distance threshold $D_1 = 20$. The local distance threshold D_2 is defined here, the calculation method of D_2 is as follows :

$$D_2 = \frac{1}{N}\sum_{i=1}^{N} | v(x,y), v_i | . \tag{9}$$

Where N represents the number of pixels in the eight neighborhood of the current pixel point, $v(x,y)$ represents the current pixel point, v_i represents any pixel point in the eight neighborhood, and $|*|$ represents Manhattan distance between the current pixel point and its neighbor, the local distance threshold is obtained from the mean of the distance between the current point and all the pixels in its eight neighborhoods.

(3) The adaptive threshold T_N is obtained from the global distance threshold and the local distance threshold, and the calculation method of T_N is as follows:

$$T_N = \frac{\sigma_2}{\sigma_1 + \sigma_2} \times D_2 + \frac{\sigma_1}{\sigma_1 + \sigma_2} \times D_1 . \tag{10}$$

Where the adaptive distance threshold T_N is obtained when the variance is used as a balance factor between the global distance threshold and the local distance threshold.

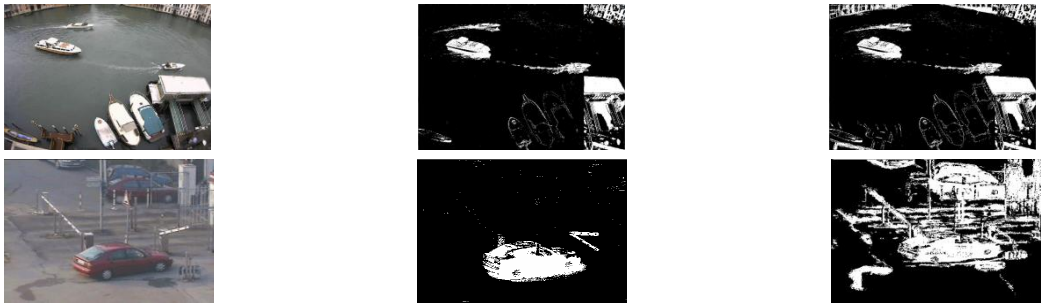
3.3 Extracting the Region of Interest of the Video Based on the Adaptive Threshold of Space-time Domain

After obtaining the adaptive quantity threshold and the adaptive distance threshold, which are brought into foreground segmentation discriminant (3) of the VIBE algorithm. The VIBE algorithm uses the Euclidean distance based on RGB space to measure the difference between the current pixel point and the pixel point in the background model. To reduce the amount of calculations, the Chebyshev distance is used to measure the difference between the current pixel point and the pixel point in the background model. The foreground segmentation equation of adaptive threshold based on the space-time domain is as follows:

$$K = \begin{cases} *\{CD(v(x), M(x)) \le T_N\} \ge \eta_N, BG \\ *\{CD(v(x), M(x)) \le T_N\} < \eta_N, FG \end{cases}. \tag{11}$$

In equation (11), $CD(x, y)$ represents the Chebyshev distance of the current pixel point and its background pixel points, and the background regions and the region of interest are segmented by the adaptive threshold, wherein the region of interest include the regions of high texture complexity and the moving regions. The adaptive threshold obtained by the above steps is robustness to dynamic scene. The experimental comparison between the improved VIBE algorithm based on adaptive threshold and the VIBE algorithm based on global threshold are as follows:

We can clearly see from Fig. 2 that the region of interest detected in the segmentation results based on the adaptive threshold are more accurate, and the regions with high texture complexity and darkness in the scene are detected as the region of interest. Using the global threshold based on the VIBE algorithm will detect the background regions as the foreground regions wrongly, resulting in inaccurate detection results. At the same time, the regions with high texture complexity in the region of interest can eliminate the false detection problems, which caused by producing error retransmission of information at the junction between the background and the foreground in the background update phase of the VIBE algorithm.



Real scene VIBE algorithm Improved VIBE algorithm

Fig. 2. The results of the two algorithms in the 68th frame

3.4 Dynamic Update Period Based on Regional Scene Complexity

In the update phase of the VIBE algorithm, besides adopting the neighborhood propagation update mechanism, the global random update period is also used to adapt to the changes of the scene. However, the global random update period causes that the real targets are quickly updated to the background model to emerge the foreground holes or the phenomenon that the fake targets are updated slowly, which will affect the subsequent detection results. Many scholars have proposed algorithms to distinguish the fake targets and the real targets, such as the algorithms based on statistical properties, the algorithms based on state feature extraction, and the algorithm based on tracking block contours [19]. However, these algorithms are more complex and have poor feedback to dynamic scenes. This paper does not distinguish the real targets and fake targets, but adjusts the update period based on the statistical information in space-time domain dynamically, which speeds up the elimination of fake targets and adapts to the dynamic changes of the scenes effectively.

In this paper, the video frames are divided into 8×8 areas for processing. The scene complexity is measured by the ratio P_j of the background pixels in each area. The expression of P_j is as follows:

$$P_j = \frac{*\{BG\}}{*\{BG\} + *\{FG\}}. \tag{12}$$

In the above formula, $*\{BG\}$ represents the number of pixels in background area, and $*\{FG\}$ represents the number of pixels of region of interest. With the changes of scene, the update period U_j also changes in the regions where P_j changes, and the dynamic update period U_j of each region is as

follows:

$$U_j = \begin{cases} \dfrac{U}{1+P_j} & P_j \geq \rho \\ 1 & P_j < \rho \end{cases}. \tag{13}$$

In the above formula, U represents the global update period, and ρ represents the threshold of regional scene complexity, which is obtained by the covariance in the corresponding background model regions. In statistical knowledge, the covariance represents the correlation between variables, which is used to measure the correlation between the pixel gray values or the degree of brightness between the pixels. The specific calculation method of this paper is that extracting ten sets of pixels in the corresponding background model area randomly, and two pixels in each sets are used to calculate the covariance. When P_j is greater than or equal to the threshold ρ, the dynamic update period is proportional to the global update period. Otherwise, the background model needs to be updated immediately. The detection results of the dynamic update period combined with the regional scene complexity and the global update period are as shown in Fig. 3.

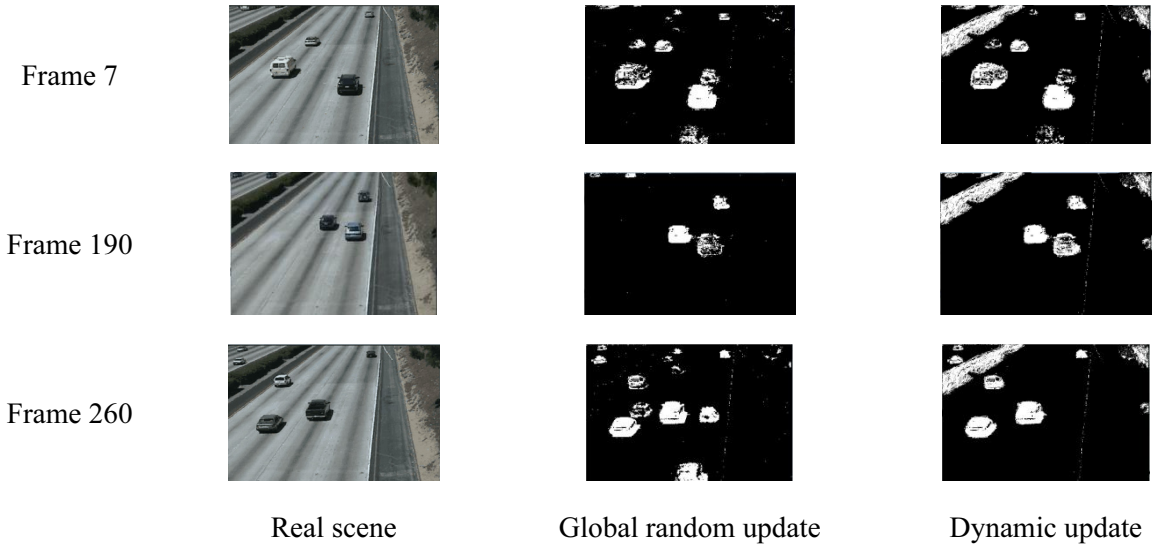


Fig. 3. The detection results of global random update mode and dynamic update mode

We can see from Fig. 3 that the global random update period is not ideal for the elimination of ghosts, and foreground holes are appeared in the detection results. However, the dynamic update method combined with the complexity of the regional scenes can quickly eliminate shadow interference and the ghost phenomenon, and which can reduce the problem of foreground holes effectively. And the detection results are more accurate.

4 Experiment

4.1 Experimental Parameters and Evaluation Indicators

Experimental parameters. The algorithm proposed in this paper is simulated under the Win10 system of Intel core i7 and 8G memory. The experiment is based on the environment of OPENCV 2.4.10 and Microsoft Visual Studio 2010. The experimental data sets which come from http://cvrr.ucsd.edu/aton/shadow/index.html and http://labrococo.dis.uniroma1.it/MAR/ provide the seven video sequences of Campus, Laboratory, Highway I, Highway II, View_001, refections-3 and 20071030_1355_c10-C. All the video sequences in the experiment are tested with the same parameters. The parameters are set as follows: the number of sample points $N = 20$; the background model uses 8 $(m = 8)$ neighborhoods; the global update period $U = 18$; the global quantity threshold $\beta = 2$; the global distance threshold $R = 20$.

Evaluation index. In order to analyze the performance between the proposed algorithm and several other algorithms quantitatively, the accuracy (Precision, PR), the recall rate (Recall, RE), the false positive rate (FPR), the false negative rate (FNR) and the F-Measure (FM) [20-21] are used as quantitative indicators to evaluate the algorithms:

$$PR = \frac{I_{TP}}{I_{TP} + I_{FP}}. \tag{14}$$

$$RE = \frac{I_{TP}}{I_{TP} + I_{FN}}. \tag{15}$$

$$FPR = \frac{I_{FP}}{I_{FP} + I_{TN}}. \tag{16}$$

$$FNR = \frac{I_{FN}}{I_{TP} + I_{FN}}. \tag{17}$$

$$FM = 2 \times \frac{PR \times RE}{PR + RE}. \tag{18}$$

In the above formulas, the parameter I_{TP} (True Positive) indicates the number of pixels detected as foreground correctly; the parameter I_{FP} (False Positive) indicates the number of pixels detected as the foreground pixels wrongly, but which actually belongs to the background pixels; the parameter I_{FN} (False Negative) indicates the number of pixels detected as background pixel wrongly, but which actually belongs to the foreground pixels; the parameter I_{TN} (True Negative) indicates the number of pixels detected as the background pixels correctly. PR indicates the proportion of the foreground pixels which are detected correctly to the total number of foreground pixels detected. Generally, the larger the value, the better the detection effect; RE indicates the proportion of the foreground pixels which are detected correctly to all the foreground pixels. Generally, the larger the value, the better the detection effect; FPR indicates the proportion detected as the foreground pixels wrongly, but which actually belongs to the background pixels to the actual background pixels. Generally, the smaller the value, the better the detection result; FNR indicates the proportion detected as the background pixels wrongly, but which actually belongs to the foreground pixels to the actual foreground pixels. Usually the smaller the value, the better the detection result. FM is a comprehensive indicator that reflects the performance of the algorithm more comprehensively.

4.2 Analysis of Experimental Results

Fig. 4 shows the results of a frame in four different scenarios under the three-frame difference algorithm, the CB algorithm, the GMM algorithm, the VIBE algorithm and the proposed algorithm. We can see that the ghosting phenomenon of the VIBE algorithm is serious, and there is almost no ghosting in the result images of the improved algorithm, which speeds up the elimination of ghosting and eliminates the probability of error retransmission of information at the junction of foreground and background, and the detection accuracy is improved. At the same time, the three-frame difference algorithm, the CB algorithm, the GMM algorithm and the VIBE algorithm emerge target hole phenomenon, and unable to adapt to the dynamic scene changes, Compared with the four algorithms, the proposed algorithm's extraction results about the moving targets are relatively complete, which overcome the target hole phenomenon effectively, and extract the ROI with high texture complexity in the videos from the detection results under test sequences with the complex scenes. In summary, the proposed algorithm in this paper has strong robustness in the different video scenes, which extracts the region of interest in the videos more accurately.

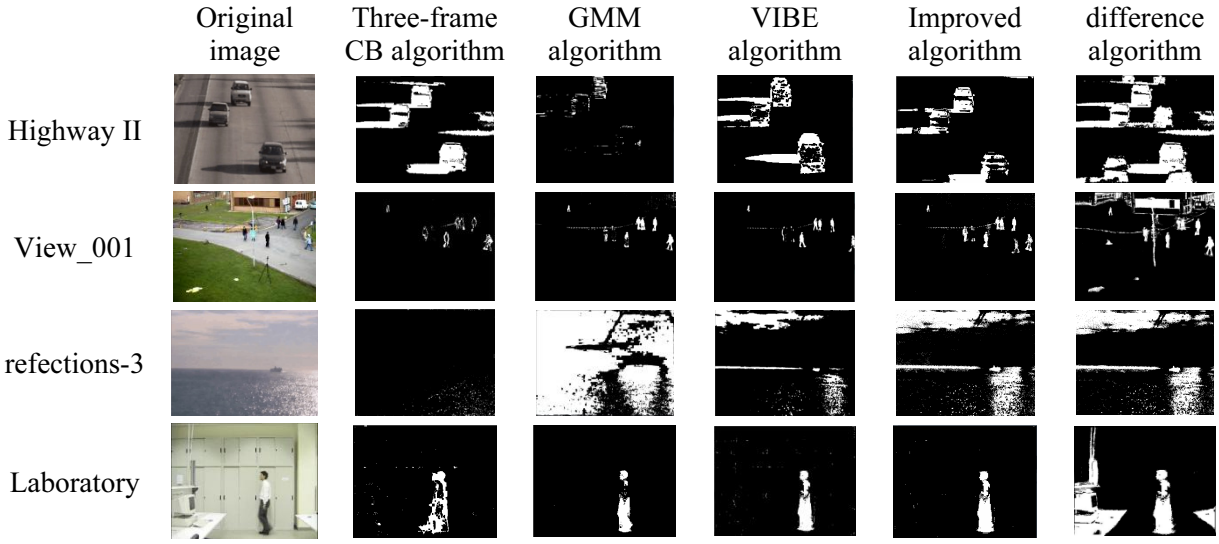
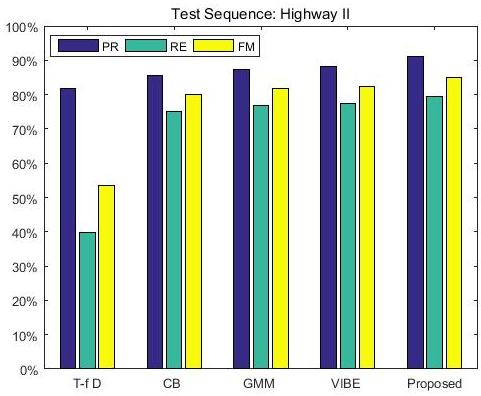
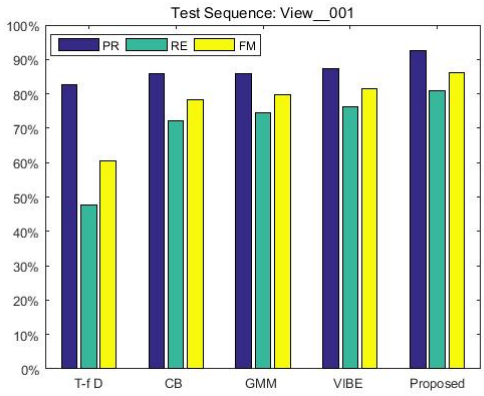


Fig. 4. The results graphs of five algorithms

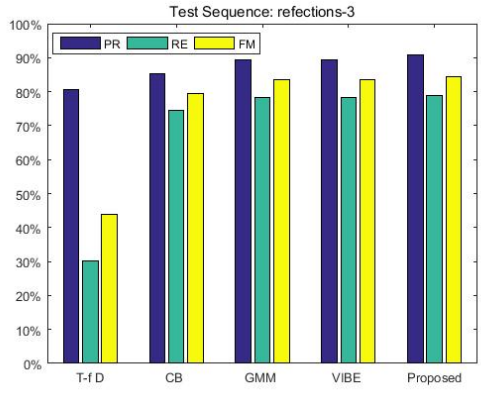
To analyze the advantage of the proposed algorithm specifically, the three-frame difference algorithm, the CB algorithm, the GMM algorithm, the VIBE algorithm and the proposed algorithm are experimented separately under four test sequences, which use the average values of PR, RE and FM to analyze the experimental results. Fig. 5 shows the comparison of experimental results of the three indicators under the four groups of scenarios.



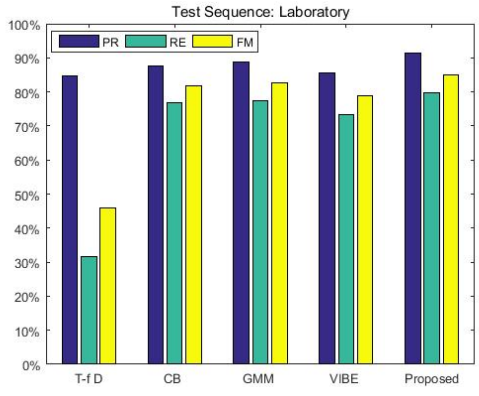
(a) the comparison of different indicators of the five algorithms under Highway II



(b) the comparison of different indicators of the five algorithms under View1_001



(c) the comparison of different indicators of the five algorithms under Refections-3



(d) the comparison of different indicators of the five algorithms under Laboratory

Fig. 5. The comparison of different indicators of the five algorithms under the four video scenarios

For convenience, the three-frame difference algorithm in Fig. 5 is shorted as T-f D. From the comparison results of the indicators in Fig. 5, we can see that the GMM algorithm and the VIBE algorithm are higher than the CB algorithm and the three-frame difference algorithm in terms of precision, but the proposed algorithm in this paper has the highest accuracy. The proposed algorithm which is under the Highway II sequences with complex scenes in Fig. 5(a) and the Refections-3 sequences with changes in light intensity in Fig. 5(c) have clear advantages; the GMM algorithm and VIBE algorithm are similar in terms of recall, but both of them are lower than the proposed algorithm in this paper. Under the laboratory sequences with the indoor scenes in Fig. 5(d) and the View1_001 sequences with the outdoor scenes in Fig. 5(b), the proposed algorithm has the highest RE values. Compared with the other four algorithms, the FM value is the highest of the proposed algorithm in this paper, therefore the result is optimal. In summary, the proposed algorithm is more optimal in terms of precision, recall and F-measure, and improves the overall performance compared to the original VIBE algorithm.

Table 1 shows the comparison results of a total of 6471 frames of five algorithms under 4 sets of scenes (Highway II, View_001, refections-3, and Laboratory). Compared with the other four algorithms, the proposed algorithm of this paper has the highest FM value, which indicates that the proposed algorithm has better detection effects from Table 1. As a result of the proposed algorithm which has high accuracy in detecting the region of motion and texture with high complexity, which has higher value between PR and RE. The PR value of the three-frame difference algorithm is significantly higher than the proposed algorithm in this paper, however, the RE value of the three-frame difference algorithm is lower than other algorithms, it is that the number of foregrounds pixels detected by the three-frame difference algorithm are small, which can suppress the influence of dynamic background factors. Therefore, the false detection rates are low, resulting in a high PR value. And the number of foregrounds pixels detected is small, causing that the false detection rates are large, resulting in a low RE values. Thus, the index values of the three-frame difference algorithm does not have a reference value. The FNR and FPR values of the VIBE algorithm are obviously higher than other algorithms except the three-frame difference algorithm, because the VIBE algorithm has a long ghost stay and is affected by the ghost regions, the maximum FNR values of the three-frame difference algorithm indicates that the false detection rates are very high. The FPR values and FNR values of the proposed algorithm in this paper are lower than other algorithms, which indicate that the proposed algorithm can guarantee the integrity of the targets while eliminating ghosts quickly.

Table 1. The Performance comparison between proposed algorithm and other algorithms

	PR	RE	FPR	FNR	FM
Three-frame difference algorithm	0.9324	0.3148	0.0591	0.2671	0.4707
CB algorithm	0.8527	0.6943	0.0732	0.0664	0.7654
GMM algorithm	0.8873	0.7275	0.0714	0.0742	0.7995
VIBE algorithm	0.8716	0.7164	0.0793	0.1501	0.8016
Improved algorithm	0.9138	0.7850	0.0584	0.0441	0.8445

In the process of the experiment, we can also find that using the global conservative update mechanism will cause much false detection to mutant scenes such as the scene with sudden changes in illumination, and cannot guarantee the correct detection of moving targets. Therefore, this paper improves the original VIBE algorithm to propose dynamic update methods, which can adapt to the changes of the scenes and accelerate the elimination of ghosts and target holes phenomenon, which improving the detection accuracy. As a result of the VIBE algorithm uses the spatial neighborhood diffusion mechanism, which causes the wrong propagation of information at the foreground and background junctions in the background model update phase, thus, the VIBE algorithm will detect the region with high texture complexity as the foreground region easily when it is used to detect the video sequences, which causes the false detection results. The proposed algorithm detects the regions with high texture complexity as region of interest in foreground regions, which avoids the problem that the error retransmission of information at the junctions of the foreground and background in the background model update phase.

5 Conclusion

Aiming at the problem that the fixed threshold can not adapt to the dynamic scene changes, and the pixels at the junction between the foreground and the background are prone to cause the error retransmission of information, the global conservative update mechanism adopted eliminates the ghost slowly and appears the target hole phenomenons easily, resulting in low detection accuracy and other issues, the background modeling algorithm is improved and an improved video region of interest extraction algorithm is proposed based on the VIBE algorithm. In order to adapt to the dynamic scene changes and eliminate the error retransmission of information in the background update phase, the adaptive quantity threshold is obtained by using the mean absolute deviation under the space-time domain combined with visual attention theory, then defining the local distance threshold, the adaptive distance threshold is obtained by the local distance threshold and the global distance threshold based on the pixels' variance, which are as the balance factor under the space-time domain. Finally, the adaptive quantity threshold and the adaptive distance threshold are brought into the VIBE foreground segmentation formula to extract the region of interest in the video, which enhance the ability to adapt to the changes of dynamic scenes and eliminate error information of retransmission in the background update phase. At the same time, an extraction algorithm for the region of interest of the video is proposed. In order to reduce the probability of occurrence of target holes and speed up the elimination of ghosts, the adaptive update period which is calculated by the regional scene complexity is combined with the proposed judgment formula to adapt to dynamic scene changes, which speeds up the elimination of ghosts and reduces the occurrence of target holes. The experiment compares the three-frame difference algorithm, the CB algorithm, the GMM algorithm, the original VIBE algorithm and the proposed algorithm under the same parameters, and which uses five kinds of evaluation indicators to analyze the experiment results. The results show that the proposed algorithm has better adaptability to dynamic scene than the other algorithms, and the false detection rates of the proposed algorithm is significantly reduced, which enhances the robustness and adaptability, and extracts the region of interest of the video accurately. Next, we will conduct a thorough study on the extraction of video region of interest combined with video coding.

References

[1] H. Xu, Z. Yang, M. Tian, Y. Sun, G. Liao, An extended moving target detection approach for high-resolution multichannel SAR-GMTI systems based on enhanced shadow-aided decision, IEEE Transactions on Geoscience and Remote Sensing 56(2)(2018) 715-729.

[2] H. Lu, K. Gu, W. Lin, W. Zhang, Object tracking based on stable feature mining using intraframe clustering and interframe association, IEEE Access 5(2017) 4690-4703.

[3] J.H. Lee, Y.N. Hwang, S.Y. Park, J.S. Jeong, S.M. Kim, An analysis of contrast agent flow patterns from sequential ultrasound images using a motion estimation algorithm based on optical flow patterns, IEEE Transactions on Biomedical Engineering 62(1)(2015) 49-59.

[4] J.D. Romero, M.J. Lado, A.J. Mendez, A background modeling and foreground detection algorithm using scaling coefficients defined with a color model called lightness-red-green-blue, IEEE Transactions on Image Processing 27(3)(2018) 1243-1258.

[5] E. Rebeiz, P Urriza, D. Cabric, Optimizing wideband cyclostationary spectrum sensing under receiver impairments, IEEE Transactions on Signal Processing 61(15)(2013) 3931-3943.

[6] S. Kang, C.Y. Lee, M. Goncalves, A.D. Chisholm, P.C. Cosman, Tracking epithelial cell junctions in C. elegans embryogenesis with active contours guided by SIFT flow, IEEE Transactions on Biomedical Engineering 62(4)(2014).

[7] F. Steinmetz, A. Wollschlager, R. Weitschat, Razer—a HRI for visual task-level programming and intuitive skill parameterization, IEEE Robotics and Automation Letters 3(3)(2018) 1362-1369.

[8] X. Tong, Z. Ye, Y. Xu, S. Liu, L. Li, H. Xie, T. Li, A novel subpixel phase correlation method using singular value decomposition and unified random sample consensus, IEEE Transactions on Geoscience and Remote Sensing 53(8)(2015) 4143-4156.

[9] L.L. Zhao, Y. Chen, Y. Zou, Q. Ye, Shooting for smarter motion detection in cameras: improvements for the visual background extractor algorithm using optical flow, IEEE Consumer Electronics Magazine 6(4)(2017) 81-91.

[10] J. Han, Y. Ma, B. Zhou, F. Fan, K. Liang, Y. Fang, A robust infrared small target detection algorithm based on human visual system, IEEE Geoscience and Remote Sensing Letters 11(12)(2014) 2168-2172.

[11] S.C. Pei, L.H. Chen, Image quality assessment using human visual DOG model fused with random forest, IEEE Transactions on Image Processing 24(11)(2015) 3282-3292.

[12] L. Zhang, A. Li, Region-of-interest extraction based on saliency analysis of co-occurrence histogram in high spatial resolution remote sensing images, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 8(5)(2015) 2111-2124.

[13] L. Zhang, K. Yang, H. Li, Regions of interest detection in panchromatic remote sensing images based on multiscale feature fusion, IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing 7(12)(2015) 4704-4716.

[14] O. Barnich, M.V. Droogenbroeck, Vibe: a universal background subtraction algorithm for video sequences, IEEE Transactions on Image Processing 20(6)(2011) 1709-1724.

[15] J. Chen, W. Bing, A Otsu threshold segmentation method based on rebuilding and dimension reduction of the two-dimensional histogram, Journal of Graphics. <https://en.cnki.com.cn/Article_en/CJFDTotal-GCTX201504013.htm>, 2015.

[16] J. Jeong, K. Lee, Bilateral frame rate up-conversion algorithm based on the comparison of texture complexity, Electronics Letters 52(5)(2016) 354-355.

[17] X. Wan, L. Gao, J. Song, H. Shen, Beyond frame-level CNN: saliency-aware 3D CNN with LSTM for video action recognition, IEEE Signal Processing Letters 24(4)(2017) 1-1.

[18] N.Q. Nguyen, R.W. Prager, Minimum variance approaches to ultrasound pixel-based beamforming, IEEE Transactions on Medical Imaging 36(2)(2017) 374-384.

[19] S. Smith, I. Williams, A statistical method for improved 3D surface detection, IEEE Signal Processing Letters 22(8)(2015) 1045-1049.

[20] G. Wang, B. Li, Y. Zhang, J. Yang, Background modeling and referencing for moving cameras-captured surveillance video coding in HEVC, IEEE Transactions on Multimedia 20(11)(2018) 1-1.

[21] L. Li, P. Wang, Q. Hu, S. Cai, Efficient background modeling based on sparse representation and outlier iterative removal, IEEE Transactions on Circuits and Systems for Video Technology 26(2)(2014) 1-1.