An Improved Visual Background Extraction Algorithm Combining Depth Information

Feng Yang1, * , Shijun Jin1* and Wenwu Ye2
1. School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China
2. School of Physics and Technology, Nanjing Normal University, Nanjing 210023, China
*Email:jinsj@seu.edu.cn

Abstract. ViBe is one of the most commonly used background subtraction method, which conducts foreground detection on each frame pixel-by-pixel. When the algorithm is applied to video sequence with large depth differences such as forest fire surveillance, the problem of traditional ViBe is appearing. If the algorithm parameters of different depth are the same, it will inevitably lead to two situations: the nearby shaking of trees is easy to cause false detection because of the large pixel area, and the distant smoke is easy to miss as it occupies small pixels of the image. To solve this problem, an improved visual background extraction algorithm combining depth information is proposed. The method calculates the dark channel information as depth information, and a conversion function is designed to adjust sensitivity to accommodate moving targets with huge depth difference. The experimental results demonstrate that the improved ViBe algorithm has better performance than traditional ViBe algorithm in forest fire surveillance. In addition, the improvement of the algorithm does not excessively increase the computational time-consuming.

1. Introduction
Detection of moving objects refers to extracting the change region from the video sequence image, and it is a fundamental task in many intelligent video monitoring application [1]. For this technology is a prerequisite in other subsequent processing such as behaviour recognition, real-time tracking and behaviour analysis. Moving objects detection has been widely used in the field of security, military, intelligent transportation, aerospace and medical.

Common motion detection methods include the optical flow, frame difference, and background subtraction methods. The frame difference is simple and has good robustness, however, the detection result is easily affected by noise and is easy to occur the problem of inaccurate target edge detection [2]. The optical flow method can accurately calculate the speed of the moving target and can handle the rotation of the target, but in most cases it is computationally complex and sensitive to noise, making it difficult to meet real-time requirements [3]. The background subtraction method is the most commonly used target detection method[4]. The algorithm is simple and the calculation amount is small [5]. It can respond to the real-time requirements of video detection.

ViBe is one of the most commonly used background subtraction method, which is based on the background extraction with spatiotemporal random selection by Barnich and Vanogenbroeck [6]. In the forest fire surveillance video image, the distant smoke accounts for fewer pixels in the image, so
when the smoke moves, the changing pixels is small and it is difficult to detect. Moreover, this small number of pixels is easily filtered out as noise. Meanwhile, the shaking of near trees in video sequence is likely to detect to large-area foreground pixels, which actually are a part of the background in forest fire prevention application.

To solve this problem, we propose an improved moving target detection algorithm by joining depth information in the process of detection. This paper is distinguished by the following main contributions:

- We propose a improved moving target detection algorithm for large-area surveillance video with huge depth difference, which combines depth information. And the cost of this method is expected to be very small.
- In this method, a simple mapping relationship is proposed, which can directly correspond the depth information to the detection sensitivity of the ViBe;
- We can estimate depth information from a single image without the need to deploy new sensors. This method does not require particularly accurate depth information. We use dark channels values as rough depth information and achieve excellent experimental results.

2. Visual background extraction

ViBe is a powerful samples-based background subtraction algorithm. By building a set of sample values as a pixel model, the moving target detection task is transformed into a classification task for each pixel [7-8]. The ViBe method mainly consists of three steps: establishing a background pixel model, pixel classification, and updating the background model.

2.1. Model initialization

Model initialization is to establish a background model. A random selection policy is used to select the pixel value in the neighborhood as the background sample set for the pixel, and the sample set represents a background model of the pixel. If the value of pixel \( x \) at time \( t \) is \( p_t(x) \) the set of pixel models can be expressed as \( \{ p_1, p_2, \cdots, p_n \} \). To achieve instantaneous initialization technique, ideally we would like to make our algorithm work from the second frame. So the first frame was being used to initialize the model. But the single frame image does not contain motion information, in that the moving target in the first image will generate a ghost object that has to fade over time.

2.2. Pixel Classification Process

After the background model is established, the pixels in the image need to be classified one by one. To classify \( p_t(x) \), first define a sphere \( S_R(p_t(x)) \) centred on \( p_t(x) \) with a radius of \( R \). Define the number of intersection elements of the neighborhood scope and the model as:

\[
I = S_R(p_t(x)) \cap \{ p_1, p_2, \cdots, p_n \}
\]

(1)

Define the minimum number of intersections as \( I_{min} \). If the number of intersections is greater than the minimum intersection number \( I_{min} \), then \( p_t(x) \) is classified as a background point, otherwise classified as a foreground pixel.

2.3 Model Update

The background update is performed when the current pixel is the background pixel. Assuming that a sample in the background model is added at time \( t_0 \), then the probability that it still exists in the sample set at time \( t_1 \) is \( \left( \frac{n-1}{n} \right)^{t_1-t_0} \) and the background update is divided into two parts: current sample set update and neighborhood update.

3. Improved ViBe combining depth information

In the forest fire video surveillance scene, the detection difficulty is totally different with the general monitoring scene. In the surveillance image, the farthest visible distance can reach several kilometers. As shown in Figure 1, the distance between the trees in the image and the monitoring platform may be
only a few tens of meters, and the location of the suspected fire is about one kilometer away from the monitoring platform.

The ViBe execution effect is usually as shown in Figure 2. The swaying trees caused a large number of foreground pixel detection results. As the distance increases, the movement of the distant smoke is difficult to trigger a large area of foreground pixels. Considering this problem, this paper presents an improved ViBe method that combines depth information.

From the dark channel prior theory, the pixel value of the dark channel of the near image in the actual scene is very low[9]. The dark channel value of the image refers to the minimum of the three color channels in a certain area of the image, expressed as follows, and its mathematical definition is as follows:

$$f_{dark}(x) = \min(\min(f^c(y)))_{c \in \{r,g,b\}, y \in \Omega(x)}$$

(2)

Where $\Omega(x)$ is a local patch centered at $x$, and $f^c$ is a color channel of image. As shown in Figure 3, the dark channel values of image pixels at different distances in the image vary widely.

Because this method only needs to calculate the approximate depth of field change trend, in the actual calculation, the calculation speed can be further increased by reducing the image, and then the original size is adjusted after the calculation is completed. Through a large number of tests, the value of the dark channel varies from 0 to 255, and the dark channel value of the partial image sky area is usually close to 255. In practice, even on sunny days, the atmosphere is not absolutely free of particles. In addition, the existence of particles is the basis of human perception of the depth of the scene[10-11].

Forest fires that are relatively far away from fire in nearby trees are more likely to trigger moving target detection. Therefore, it is desirable to have a lower motion sensitivity in a close range (i.e., a region where the dark channel value is smaller), and a higher motion sensitivity in a long distance region (a region having a larger dark channel value). However, the sensitivity of the motion detection in the long-distance area cannot be increased indefinitely, because there is usually a cloud block floating in the sky area, and there is no practical significance for the motion sensitivity of the sky to be too high.
It can be seen from the above that the larger the radius $R$ of the neighborhood range $S_R(p_t(x))$ in ViBe, the easier it is to satisfy the condition of the minimum matching threshold and the more easily classified as the background when the number of matched samples is subsequently calculated. Therefore, the detection accuracy of ViBe is negatively correlated with the radius $R$ of the neighborhood range, and the foreground detection sensitivity can be improved by reducing the neighborhood radius.

![Radius and Threshold Graph](image)

**Figure 4.** Mapping function

Based on the above discussion, this section designs a mapping function for converting the dark channel to the ViBe detection radius as shown in Figure 4. Max refers to the highest detection sensitivity, Min refers to the lowest detection sensitivity, and threshold indicates the dark channel threshold. Above this threshold, the sensitivity of the motion detection does not continue to increase in order to avoid unnecessary false detection. Through the mapping function shown in Figure 4, the ViBe moving target detection algorithm can adapt to moving targets with large difference in depth.

4. Experimental results

To evaluate the effectiveness of the improved algorithm, an experimental system is set up. The Experiments are implemented on a PC with Intel i5-6500 and 8G memory.

We compared the algorithm presented in this paper with traditional ViBe algorithm on four surveillance videos, which is 1280×720. Figure 5 shows the test results on the test videos. From the figure5(a), we can see that the rectangular area occupied by the plume flow area in video 1 is about 50×100 pixels, and its area is much smaller than the area of a single ordinary tree below the image. Therefore, it can be seen in Figure5(b) that there are a large number of erroneous foreground pixels in the image due to the shaking of the trees, and the moving smoke that needs to be completely detected at a distance has only a relatively sparse motion check result. Because in smoke detection applications, the swaying trees are the background. In Figure5(c), the improved ViBe results of four videos clearly highlight the area where the plume flutters, and because of the lower detection sensitivity used in the close-range part of the image, false foreground pixel results caused by tree shaking are significantly reduced.

Finally, we tested the speed of the improved algorithm in this paper. The average consumption time of the image sequence of 1280×720 is shown in Table 1. It can be seen from the table that although the improved ViBe has one more processing in the algorithm flow, it hardly increases the computational complexity. The dark channel calculation only needs to be calculated when the first frame is initialized, and the subsequent frame calculation does not need to be repeated. In the improved ViBe method, the calculation of the dark channel in the first frame takes about 5ms.
5. Conclusion
The algorithm proposed in this paper has the following improvements: (1) The method of calculating value of dark channels as depth information is creatively proposed. We can estimate depth information from a single image without the need to deploy new sensors. (2) This algorithm designs a mapping function to adjust the detection sensitivity. (3) The improvement of the algorithm hardly increase the computational time-consuming due to the actual optimization. (4) The algorithm proposed in this paper can be applied not only to the scene of forest surveillance video, but also to other scenes with large spatial spans. It extends the application range of the traditional ViBe algorithm. But the proposed algorithm are still some deficiencies that need to be improved, for example, the parameters of the algorithm cannot be adaptively adjusted. And regions where the depth changes drastically (such as the shaking treetop) are not well processed. This will be the focus of future research on the algorithm.

Reference
[1] Lifen T U , Sidong Z , Qi P , et al. Moving object detection based on Gaussian pyramid[J]. Journal of Central South University, 2013, 44(7):2778-2786.
[2] Xiao B, Cheng L, Hao C, et al. Moving object detection and recognition based on the frame difference algorithm and moment invariant features[C]// Control Conference. 2008.

[3] Baker S, Scharstein D, Lewis J P, et al. A database and evaluation methodology for optical flow[J]. International Journal of Computer Vision, 2011, 92(1): 1-31.

[4] Cheung S S, Kamath C. Robust techniques for background subtraction in urban traffic video[C]// Visual Communications & Image Processing. Visual Communications and Image Processing 2004.

[5] Suo P, Wang Y. An improved adaptive background modeling algorithm based on Gaussian Mixture Model[C]// Signal Processing, 2008. ICSP 2008. 9th International Conference on. IEEE Xplore, 2008.

[6] Barnich O, Droogenbroeck M V. VIBE: A POWERFUL RANDOM TECHNIQUE TO ESTIMATE THE BACKGROUND IN VIDEO SEQUENCES[C]// 2009 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2009.

[7] Barnich O, Droogenbroeck M V. ViBe: A Universal Background Subtraction Algorithm for Video Sequences[J]//. IEEE Transactions on Image Processing, 2011, 20(6):1709-1724.

[8] Bouwmans T, Porikli F, Hoferlin B, et al. Background modeling and foreground detection for video surveillance[M]// Chapman and Hall/CRC, 2014: 7.1-7.23.

[9] He K, Sun J, Tang X. Single Image Haze Removal Using Dark Channel Prior [J]. IEEE Trans Pattern Anal Mach Intell, 2011, 33(12):2341-2353.

[10] James M K, James K, Goldstein E B. Sensation and Perception [J]. Leonardo, 1983, 15(2):160.

[11] Preetham A J, Shirley P, Smits B. A practical analytic model for daylight[C]// Conference on Computer Graphics and Interactive Techniques. ACM Press/Addison-Wesley Publishing Co. 1999:91-100.