A Novel Motion Detection Approach Based on the Improved ViBe Algorithm

PAK Chun-Hyok[1,2], ZHAO Hai[1], ZHU Hongbo[1], PAN Yilin[3],

1. School of Computer Science and Engineering, Northeastern University, Shenyang 100819, China

2. Electronic Engineering Faculty, Kim Chaek University of Technology, Pyongyang, 999093, DPRK

3. School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China

(E-mail: Vlsi126@126.com)

Abstract: Ghost elimination is considered as a vital phase towards the moving object detection problem. Despite all the efforts of the existing methods have been made so far, finding an accurate and fast computational approach to solve ghost elimination is still a challenging problem for moving detection. In this paper, we propose a novel motion detection approach based on the ViBe algorithm. We employ the weighted window to select the dynamic adaptive optimal parameter for the ViBe model firstly. After that, the model integrates the inter-frame difference algorithm to accelerate the process of the ghost elimination. Experiment results show that our approach can detect the moving objects in real-time and get the higher accuracy result in the comparisons with the existing approaches.

Key Words: Video surveillance, Moving object detection, Background subtraction algorithm, Ghost elimination, Adaptive threshold

1 INTRODUCTION

Automatic video analysis system is more and importance recently with the rapidly development of economy and the uncontrolled urbanization. It started from the public security demand is widely used in field of the public security and auto-navigation of car or airplane, and so on. During the real-time video surveillance realization, to separate the foreground and background from video frame is became first stage for the real-time system realization. The motion detection algorithm can be divided into six categories; the optical flow [1-2], frame difference[3,4], background subtraction[5-10], cluster analysis[11], image segmentation[12-13] and point detection[14-15] method, respectively. The frame difference and background subtraction algorithms among them are widely used algorithm at present, because to do not the prior knowledge about the surveillance place. GMM algorithm that models each background pixel by using of a few of the Gaussian distribution is background subtraction algorithm [5-6]. And then, for new pixel, GMM is starting to search that new pixel value is matched to one of a few Gaussian distribution, isn't. By the search result, new pixel is estimated as foreground or background pixel. Its advantage is that extracted result is very exact. The dis-advantage is, because to demand the probability calculation for each pixel, the operation speed is very slow.

The codebook algorithm is also the moving detection algorithm based on background subtraction [7,8]. According to situation of the color and luminance, the training patterns of pixels are clustered on by this algorithm. Sample background pixel values at each pixel are quantized into codebooks which represent a compressed form of background model for a long image sequence. The codebook representation is efficient in memory, and because to do not calculate the probability, its' speed is very fast. Generally, the background image accuracy by this algorithm is a little lower than that by GMM. Finally, let's introduce the frame difference algorithm. This uses the difference between the current frame and previous frame on the image sequence, and each pixel on the current frame is estimated as background pixel or foreground pixel with gained difference. This algorithm advantage is very simple and has the low computational complexity. But, this algorithm can only extract the contour of motion object. In this paper, to realize the real-time motion detection, we selected the ViBe algorithm that has better shake-resistance and real-time than another, and improved a few its drawback. The remainder of paper is organized as follows. Section 2 describes the initial ViBe algorithm. In section 3, we present the principles and methods of our improved ViBe. Experiments are provided in Section 4, and Section 5 concludes this paper.

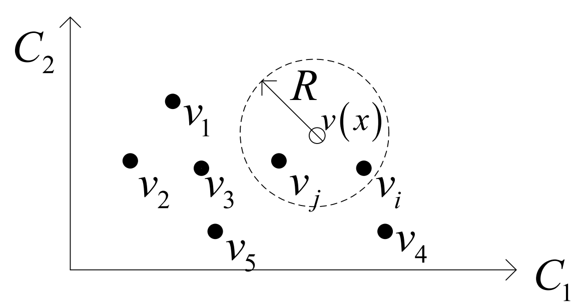


Fig. 1 the comparison between a pixel value and sample value

2 Related Work

The initial ViBe(Visual Background Extractor) is motion detection algorithm based on non-parameter random sample

This work is supported by National Nature Science Foundation under Grant ×××××××

model[9,10]. In this algorithm, the background models for each pixel are initialize on the basis of first frame, and then, by using of matching performance between each background model and corresponding pixel on the next frame, each pixel on next frame estimated as background or foreground pixel. This algorithm are divided into 3 stages; initial stage, matching stage and updating stage for background model[10].

2.1 The Initialization of Background Model for Each Pixel

Each background model are made of 20 background sample values for each pixel, background samples are random- repeatedly selected from 9 pixels around each pixel.

2.2 The Background Pixel Estimation for Each Pixel on Next Frame

In this algorithm, the background estimation is started from next frame. In the (C_1, C_2) color space, we denote $v(x)$ as the value of pixel x, and $S_R\left(v(x)\right)$ as the sphere of radius R around $v(x)$. Then the current pixel belongs to foreground or background pixel according to the number of samples contained in sphere.(See Fig. 1)

2.3 The Background Model Updating

The updating method of each background model is as following. Once a pixel belongs to background, a random sample in model has to be substituted as that pixel value. And, meanwhile have to insert that pixel value into model of random one among neighboring pixels around that pixel. It is called subsampling.

This algorithm representation has fast operation speed, accurate detection result, and robustness against the noise and illumination variation[16]. But, when we use this algorithm, the ghost is often appeared by its own characteristic.

3 The Improved ViBe Algorithm

3.1 Ghost Elimination

ViBe algorithm often generates the ghost by its own characteristic. Fig. 2 shows the ghost appearance in the result of the ViBe algorithm.

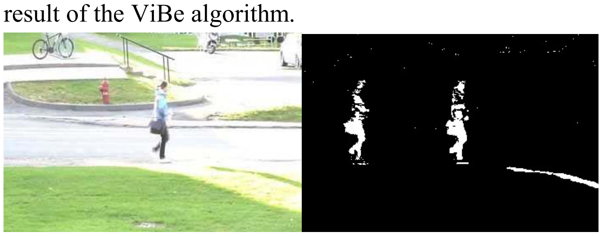


Fig. 2 The ghost appearance in the motion detection result by using of ViBe algorithm.

In ViBe algorithm, once the model contains at least two right background pixels value among own samples, the a miss-estimated pixel as foreground (i.e. ghost pixel) can be extended to background by two samples, but elimination speed is very slow. As you can see in Fig. 3, when the ghost began to generate, the result image by initial ViBe and the result image by the frame difference method are approximately equal. This is different from Fig. 2 where the actual foreground is completely separated from the ghost. We brought inspirations from this picture, and decided to remove the ghost pixel by using the frame difference algorithm.

For the fast ghost elimination, we use own characteristic of this algorithm. Once we find ghost pixel, and its pixel value is inserted into its own background model, the probability that current pixel belongs to background is very high. To find the ghost pixel, we used the frame difference method. The detail method is as following. If the difference between a pixel value on current frame and the corresponding pixel value on previous frame exists within the certain limit, this point can be regarded as the candidate ghost pixel. To consider the difference between the non-ghost pixel generated by slow moving and the exact ghost pixel, after certain number of sequent frame, that pixel is still remaining candidate ghost status, it is finally estimated as the exact ghost pixel. It can be written as following equation (1).

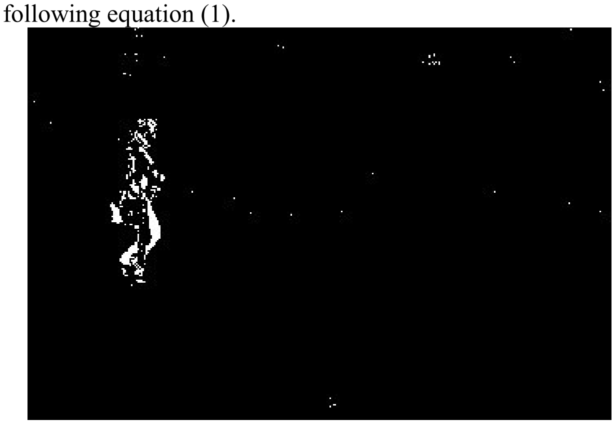


Fig. 3 The generation of initial ghost.

For the fast ghost elimination, we use own characteristic of this algorithm. Once we find ghost pixel, and its pixel value is inserted into its own background model, the probability that current pixel belongs to background is very high. To find the ghost pixel, we used the frame difference method. The detail method is as following. If the difference between a pixel value on current frame and the corresponding pixel value on previous frame exists within the certain limit, this point can be regarded as the candidate ghost pixel. To consider the difference between the non-ghost pixel generated by slow moving and the exact ghost pixel, after certain number of sequent frame, that pixel is still remaining candidate ghost status, it is finally estimated as the exact ghost pixel. It can be written as following equation (1).

$$TOS(x) = \begin{cases} TOS(x)+1, & if\ |d(x)| \le T_d \\ 0, & else \end{cases} \tag{1}$$

where, $TOS(x)$ is counter for the candidate ghost pixel estimation of pixel x, T_d is the preset threshold value,

$d(x)$ is difference between pixel values as mentioned above.

3.2 Self-adaption Model Parameter

We find that the use of radius leads to excellent result in every situation [10]. In ViBe algorithm, the greater the size of , the more pixels on the frame belong to background pixel. Instead, the more little the size of , the greater pixels on frame belong to foreground pixel. By using of standard deviation of background model for each pixel, according to sample change, Droogenbroeck et al. have been calculated the dynamic size of , which is limited from 20 to 40[18]. This dedicates to enhance the accuracy of foreground image, but it leads to large amount of debasement in processing rate. In our many experiments, in processing rate, it drops down to even 1/4 of initial algorithm.

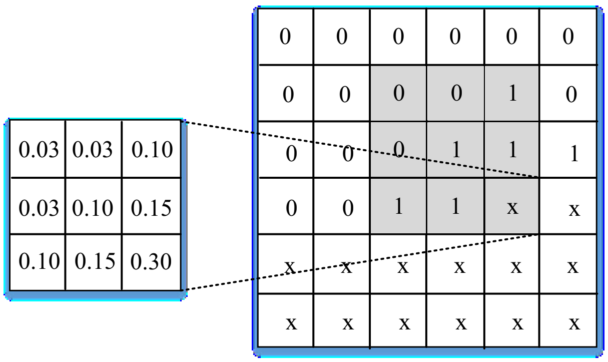


Fig. 4 The weighted window for the adapting parameter

For improving the initial ViBe performance, we assume as following. In case of perfect algorithm, it has to do not affect the interest foreground pixel, meanwhile, belong to background pixel for non-interest foreground pixel. And, in the viewpoint of region size, the region formed by the interest foreground pixels is generally big, but region by non-interest foreground pixel is little and its region is often formed by blinking pixel. To guaranty the demand in processing rate, the our dynamic model parameter is designed as following. The status of treated pixels on frame certainly influences on the estimation result of a current pixel. Therefore, we decided this characteristic of non-interest foreground pixel. To realize these, according to the bit-wise distance between treating pixel and treated pixel, we set the influence on treating pixel of treated pixel. The influence degree is as following Fig. 4, that is, far treated pixel from treating pixel, more small its influence . In view of processing rate, the influence scale of treated pixels is designed as 3×3, the sum of cells in this weight window is about 1, the position of treating pixel is right-most lower corner. Unlike in literature[18], the size of R is limited from 2 to 20. Thus, if the convolution between this weighted window and binary state of corresponding treated pixel around treating pixel is equal about 1, $R=2$, otherwise, if the convolution is equal 0.3, $R=20$. The following equations shows this processing.

$$T = W\square \tag{2}$$

where, W is weighted window, B is matrix by corresponding treated pixels(for exact detection about motion object, the treating pixel status is assumed as 1), □ is convolution operator. And, by using the equation (3), the size of R is calculated for corresponding treating pixel by using linear relation between R and T, as in literature[18].

$$R = -25.7 \cdot T + 27.7 \tag{3}$$

4 Experiment

The experiments is divided into three parts: we show our result of ghost elimination to verify our proposed the approach at first. After that, the adaptive ViBe model is utilized to select the optimal parameter R. Finally, we compares our final results with that of the existing methods. The accuracy for gained binary result image is estimated by using of 3 factors, *recall*, *precision* and their harmonic mean, the *F-Measure*, respectively[17]. These expressions is as following equation (4) - (6).

$$recall = \frac{\#\, correctly\ classified\ foreground\ pixels}{\#\, foreground\ pixels\ in\ GT} \tag{4}$$

$$precision = \frac{\#\, correctly\ classified\ foreground\ pixels}{\#\, pixels\ classfied\ as\ foreground} \tag{5}$$

$$F = 2\frac{recall \cdot precision}{recall + precision} \tag{6}$$

To improve the effect of our method, the two result image produced by initial ViBe and our improved ViBe were compared by 3 factors above mentioned. In our experiments, under the many of environment and for the variable size of frame (i.e. 320×240, 360×240, 640×480), the experiments were compared. For the initial ViBe, we used the parameter values mentioned in literature[10], i.e. $N=20$, $R=20$, $\#_{\min}=2$ and $\phi=16$.

4.1 Part I: the Result of Ghost Elimination

Fig. 5 shows the ghost elimination effect in 360×240 result images produced by initial ViBe and our method respectively. In the picture, the image from first to fourth column show 607th, 667th, 717th and 767th images among images before and after process, respectively.

The images on second row and third row show the result images produced by initial and our algorithm, respectively. As you can see in Fig. 6, the ghost elimination speed by our algorithm is faster than initial algorithm. In view of processing rate, for the 360×240 sequent frame, the initial and our algorithm have 43 and 41 in frame per second (fps), respectively. And, Fig. 6 graphically shows the number of ghost pixel on each result frame produced by initial and our algorithm, respectively.

4.2 Part II: the Result of the Adaptive ViBe Model

Fig. 7 shows the effect by using of adapting threshold model parameter R. This frame is under the complex environment, i.e. many of the blinking pixels by fountain and leafs, these blinking pixels are non-interest pixels. The (a), (b), (c) and (d) in Fig. 7 are the original frame, grand truth image, and result image by initial and our improved

ViBe, respectively. As you can see in Fig. 7, for all of region on the original frame, because the initial ViBe algorithm used fixed model parameter R , many of non-interest pixels are detected on result image. In the case of our improved algorithm, many of non-interest pixels are removed on the result image by using of adapting threshold parameter R .

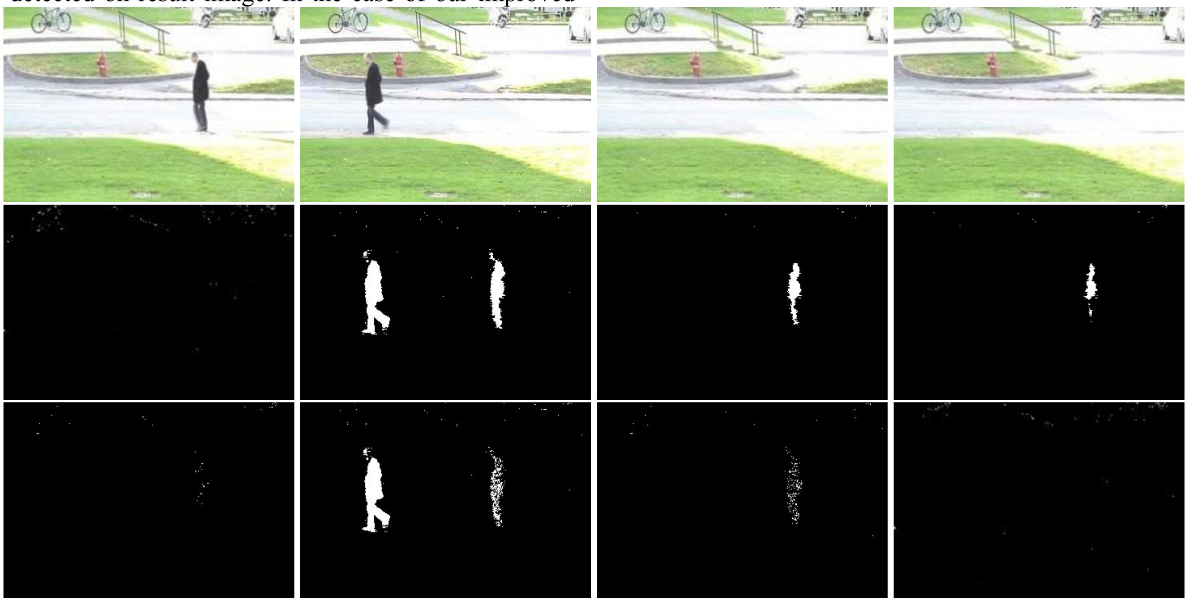


Fig. 5 The compare of ghost elimination effect by initial and improved ViBe algorithm
1th row – sequence image   2th row – initial ViBe   3th row – improved ViBe

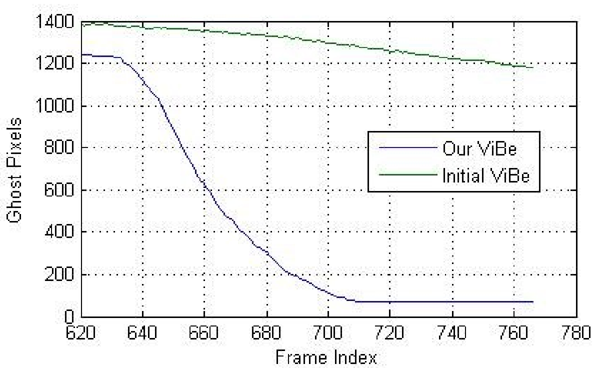


Fig. 6 The number of ghost pixels on the each result frames produced by two method respectively

Table 1 and 2 show a part of evaluation result by equation (4) and (5), and Fig. 8 shows a part of evaluation result by equation (6). As you can see in table 1 and 2, the *precision* and *recall* by our algorithm are little higher than initial algorithm. For *F-measure*, our algorithm is little higher than the initial ViBe.

Table 1 Precision of initial and improved ViBe algorithm

Frame Number	Fixed parameter	Adapting parameter
733	0.862	0.878
735	0.870	0.884
737	0.867	0.885
739	0.886	0.871
741	0.881	0.889
743	0.907	0.955

Table 2 Recall of initial and improved ViBe algorithm

Frame Number	Fixed parameter	Adapting parameter
733	0.598	0.634
735	0.608	0.639
737	0.618	0.647
739	0.604	0.645
741	0.624	0.648
743	0.622	0.644

(a) (b) (c) (d)

Fig. 7 Result images by initial and improved ViBe algorithm

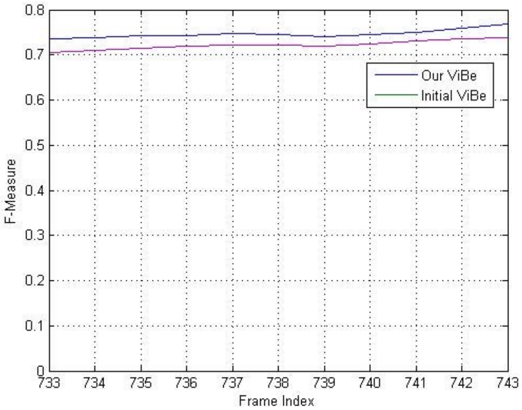
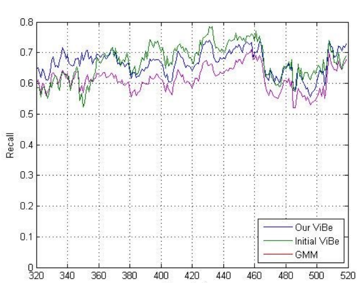


Fig. 8 Comparison chart of *F-measure* of initial and improved ViBe algorithm

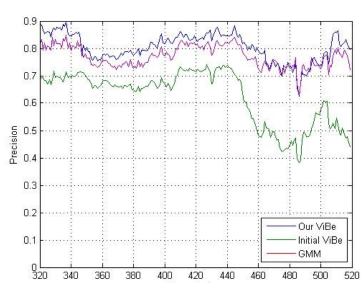
4.1 Comparison with the Existing Methods

We applied our improved ViBe method to the natural video. The following Fig. 9 shows the extracted background result frames by our improved ViBe, the initial ViBe and GMM algorithm, respectively. Fig.10 shows the accuracy analysis results on result frames by three of the factors.
As shown in Fig. 10, the difference between tree of factors is not great at recall value, but performance of our proposed method is very good at the precision and *F-measure*. Further, for the captured actual video data, our method is more faster than GMM at FPS. Certainly, the performance of GMM often depends on GMM's model parameter, its biggest weakness is the real-time. In Fig. 11, we illuminated the real-time of different method, as initial ViBe, ViBe by using of variance of the each pixel's background model[18], GMM and our method.

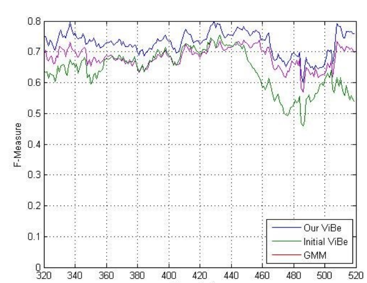
Fig. 9 Result images by initial and improved ViBe algorithm: 1th row - Original frame; 2th row - Initial ViBe; 3th row – Improved ViBe; 4th row – GMM



(a)



(b)



(c)

Fig. 10 The Result Analysis by the initial ViBe, the improved ViBe and GMM algorithm.
(a) Recall Value (b) Precision Value (c) *F-Measure*

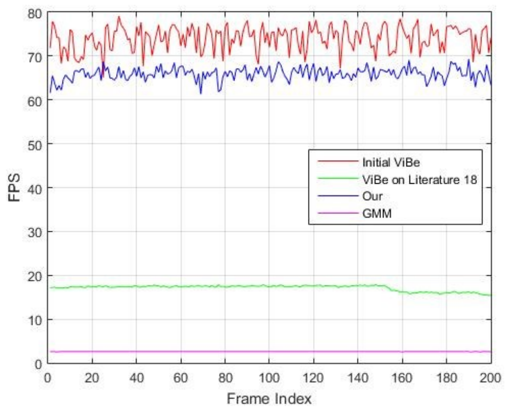


Fig. 11 The Real-time Analysis by the initial ViBe, the ViBe on the literature[18], the our improved ViBe and GMM algorithm.

5 Discussion and Conclusion

During the realization of the moving detection system, the most important factors are the real-time and the precision. Generally, the FPS of video data captured by camera is about 25~30fps. Under the circumstance of variable size and environment, many of the experiment results show that our method does not greatly drop down the real-time performance and the results achieved by this method are better than initial ViBe. Generally, the resolution of frame influences the real-time detection of system. In our experiments, for 640×480 frame, initial ViBe and our improved ViBe have about 20fps and 13fps, respectively. But, for 360×240 frame, these have about 70fps and 60fps, respectively. Thus, our method can be used to realize the real-time motion detection for about 360×240 frame. Especially, because of our method has both the ghost elimination and adapting model parameter function, it can be effectively contributed to enhancing the accuracy of the motion detection.

REFERENCES

[1] Horn B K P. Determining optical flow[J]. Artificial Intelligence, 1981(17):185-203.

[2] Brox T, Bruhn A, Papenberg N, et al. High accuracy optical flow estimation based on a theory for warping[M]//Computer Vision-ECCV 2004. Springer Berlin Heidelberg, 2004: 25-36.

[3] Chaohui Z, Xiaohui D, Shuoyu X, et al. An improved moving object detection algorithm based on frame difference and edge detection[C]//Image and Graphics, 2007. ICIG 2007. Fourth International Conference on. IEEE, 2007: 519-523.

[4] Collins R. A system for video surveillance and monitoring: VSAM final report[R]. Pittsburgh, Peen, America: Carnegie Melon University, 2000.

[5] Stauffer C, Grimson E. Adaptive background mixture models for real-time tracking[C]. IEEE Conference on Computer Vision and Pattern Recognition, 1999:246-252.

[6] Zivkovic Z. Improved adaptive Gaussian mixture model for background[A]. Proceedings of the 17th International Conference on Pattern Recognition[C], 2004, 2:28-31.

[7] Kim K, Chalidabhongse T, Harwood D, Davis L. Real-time foreground-background segmentation using codebook model[J]. Real-Time Imaging, 2005, 11(3):172-185.

[8] Kim K, Chalidabhongse T, Harwood D. Background modeling and subtraction by codebook construction[C]. International Conference on Image Processing, 2004:3061-3064.

[9] Barnich O, Droogenbroeck M V. ViBe: A powerful random technique to estimate the background in video sequences[C]. IEEE International Conference on Speech and Signal Processing, 2009:1709-1724.

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

[11] Papageorgiou T D, Curtis W A, McHenry M, et al. Neurofeedback of two motor functions using supervised learning-based real-time functional magnetic resonance imaging[C]//Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE. IEEE, 2009: 5377-5380.

[12] Comaniciu D, Meer P. Mean shift: A robust approach toward feature space analysis[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2002, 24(5): 603-619.

[13] Mirhosseini M, Malik M, Shi Z, et al. Efficient separation of the orbital angular momentum eigenstates of light[J]. Nature communications, 2013, 4.

[14] Forssen P E, Lowe D G. Shape descriptors for maximally stable extremal regions[C]//Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on. IEEE, 2007: 1-8.

[15] Mikolajczyk K, Schmid C. Scale & affine invariant interest point detectors[J]. International journal of computer vision, 2004, 60(1): 63-86.

[16] Bouwmans T, Porikli F, Hoferlin B, Vacavant A. Background Modeling and Foreground Detection for Video Surveillance[M]. Chapman and Hall/CRC, 2014.

[17] Brutzer S. Evaluation of background subtraction techniques for video surveillance[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2011:1937-1944.

[18] Droogenbroeck M V, Paquot O. Background Subtraction: Experiments and Improvements for ViBe[C]. Change Detection Workshop, Rhode Island, 2012: 32-37.