

Research Article

A Multiframes Integration Object Detection Algorithm Based on Time-Domain and Space-Domain

Yifan Liu , Zhenjiang Cai , and Xuesong Suo 

College of Mechanical and Electrical Engineering, Hebei Agricultural University, Baoding 071000, China

Correspondence should be addressed to Zhenjiang Cai; czj65@163.com

Received 26 September 2017; Revised 16 January 2018; Accepted 21 January 2018; Published 20 February 2018

Academic Editor: Simone Bianco

Copyright © 2018 Yifan Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to overcome the disadvantages of the commonly used object detection algorithm, this paper proposed a multiframes integration object detection algorithm based on time-domain and space-domain (MFITS). At first, the consecutive multiframes were observed in time-domain. Then the horizontal and vertical four-direction extension neighborhood of each target pixel were selected in space-domain. Transverse and longitudinal sections were formed by fusing of the time-domain and space-domain. The mean and standard deviation of the pixels in transverse and longitudinal section were calculated. We also added an improved median filter to generate a new pixel in each target pixel position, eventually to generate a new image. This method is not only to overcome the RPAC method affected by lights, shadows, and noise, but also to reserve the object information to the maximum compared with the interframe difference method and overcome the difficulty in dealing with the high frequency noise compared with the adaptive background modeling algorithm. The experiment results showed that the proposed algorithm reserved the motion object information well and removed the background to the maximum.

1. Introduction

The research and application of the motion object detecting and tracking are on the rise in various fields; object detection is one of the essential technical issues. Motion object detection technology is to extract moving objects (also known as the background) from the scene (also known as the foreground) which in the video of every frame image. The better the results of more reserved objects information and less background information, the better the ability of object detection.

The common motion objects detection algorithms include the background difference method, the interframe difference (IFD) method, RPCA method, and the adaptive background modeling method. The advantages of background difference method are being simple, being able to maintain the integrity of motion objects to the maximum, and having more accurate positioning; this method is suitable for moving object detection with fixed cameras and no changed background. Its shortcomings are obvious; in most practical application, the background information

will be influenced by illumination change, target shadow, and some outside influence of impurities and noise [1–3]. Interframe difference method has the advantage of simple operation, being suitable for unfixed cameras, and also having strong robust character and acclimatization. The disadvantage is being unable to extract the complete object information, in the case of object too slow or too fast motion, which may cause object missing or error detection in two objects [4, 5].

Robust principal component analysis (RPCA) method transformed video frames into vectors. The background matrix was decomposed into low rank matrix after determinant transformation; the foreground matrix was decomposed into sparse matrix after determinant transformation. Rodriguez and Wohlberg proposed a simple alternating minimization algorithm for solving a minor variation on the Fast Principal Component Pursuit (FPCP) [6, 7]. This method decomposed the sparse matrix even after the first outer loop. But detection results showed difficulty in dealing with the high frequency noise. He et al. proposed Grassmannian Robust Adaptive Subspace Tracking Algorithm (GRASTA)

[8]; this method improved the processing speed and was effective for removing the background, but the effect of extracting motion object information was not good.

In recent years, adaptive background modeling method has received the widespread attention and research. It mainly includes the following several kinds of algorithms: (1) Gaussian mixture model (GMM) algorithm [9–12]: the algorithm is to establish initial background frame by single or multiple Gaussian filters and adaptively updated as the background frame changes. This algorithm can extract the moving object information well and remove most of the background noise. (2) Codebook algorithm [13, 14]: this algorithm is a compressed sample background extraction algorithm based on the background codebook, and in the meanwhile codebook is updated correspondingly. (3) Visual background extraction (Vibe) algorithm [15–17]: this algorithm updates background model adaptively and detects the motion objects based on consecutive pixels having similar characteristics in space-domain. And it possesses better real-time performance and robustness. Essentially speaking, these three algorithms are all in the background difference method, so they have some common faults. For example, they easily lead to “ghosting” in background model, difficulty in dealing with the high frequency noise (flicker leaves, fluctuating water surface), and hysteresis of background noise removal.

In this paper, a multiframes integration object detection algorithm based on time-domain and space-domain (MFITS) was presented aiming at explaining the advantages and disadvantages of these common object detection algorithms. At first, the consecutive multiframes were observed in time-domain. Then the horizontal and vertical four-direction extension neighborhood of each target pixel were selected in space-domain. Transverse and longitudinal section were formed by fusing of the time-domain and space-domain. The mean and standard deviation of the pixels in transverse and longitudinal section were calculated. We also added an improved median filter to generate a new pixel in each target pixel position, eventually to generate a new image. The proposed algorithm is fundamentally different from the common algorithms in image processing within the image frames calculation, but in a transverse and longitudinal section of the multiframes.

The MFITS algorithm is not only to overcome the PRAC method affected by light, shadows, and noise, but also to reserve the object information to the maximum compared with the interframe difference method and overcome the difficulty in dealing with the high frequency noise compared with the adaptive background modeling algorithm.

2. Common Object Detection Algorithms Principle

Common object detection algorithms often use difference between frames to detect the motion object, in other words, time-domain operation. For example, the background difference method and the interframe difference method both compute grey-scale changes of each pixel between two frames. Optical flow method is to compute changed trend

of the pixel values for multiple frames in the time-domain. Adaptive background modeling algorithm is to generate an adaptive background model and then difference with other frames. The principle is described as follows generally:

$$d(i, j) = |m(i, j, n) - m(i, j, k)|. \quad (1)$$

$m(i, j, n)$ is pixel value in n frame (i, j) position. $m(i, j, k)$ is pixel value in k frame (i, j) position.

Then the different values between two frames are compared with a threshold value T . For less than the threshold grey value as the background, directly set pixel grey value of 0, grey value greater than the threshold value is as the foreground set to 255 pixels directly.

$$\begin{aligned} b(i, j) &= 0, & d(i, j) < T, \\ b(i, j) &= 1, & d(i, j) \geq T. \end{aligned} \quad (2)$$

$b(i, j)$ is pixel value in (i, j) position of the difference image.

3. MFITS Algorithm Principle

At first, the consecutive multiframes were observed in time-domain. Then the horizontal and vertical four-direction extension neighborhood of each target pixel were selected in space-domain. Transverse and longitudinal section were formed by fusing of the time-domain and space-domain. The mean and standard deviation of the pixels in transverse and longitudinal section were calculated. We also added an improved median filter to generate a new pixel in each target pixel position, eventually to generate a new image.

In this paper, the experiment of the five consecutive frames is used in the time-domain; 5 frames can reflect the enough displacement change of moving target, instead of the traditional frame difference method moving targets between two near frames, a motion too small for the displacement to detect. And moving targets can ensure not having much displacement within 5 frames, located in a reasonable range. Equation is expressed as follows:

$$\begin{aligned} M(i, j) &= [m(i, j, n), m(i, j, n+1), m(i, j, n+2), \\ &\quad m(i, j, n+3), m(i, j, n+4)]. \end{aligned} \quad (3)$$

$M(i, j)$ is n pixels sequence in (i, j) position of n consecutive frames. Figure 1 shows the schematic of five consecutive frames.

In practical applications, the target could be affected by noise and even a static background; the disturbances occur in the time-domain. The target can be considered as stationary signal when the pixel value did not change much (background). On the other hand, the target can be considered as impact signal when the pixel value changed a lot (foreground). So different from the common difference moving target detection algorithm in (1), we adopt the mean

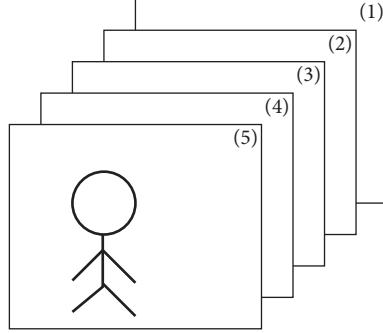


FIGURE 1: The five consecutive video images.

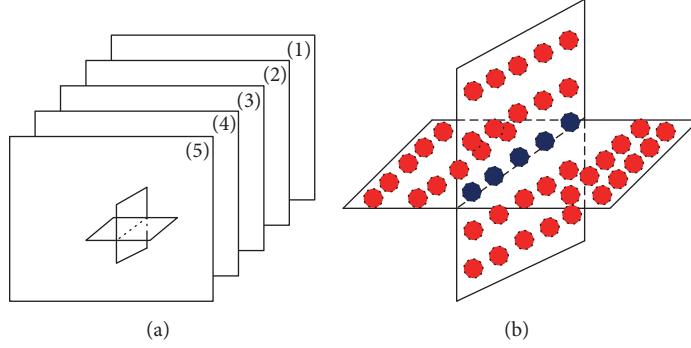


FIGURE 2: Transverse and longitudinal section of the five consecutive video images.

and standard deviation of the pixels value in the time-domain to distinguish signal characteristics.

$$\bar{m}(i, j) = \frac{\sum_{n=n}^{n+4} m(i, j, n)}{5}, \quad (4)$$

$$S(i, j) = \sqrt{\frac{\sum_{n=n}^{n+4} (m(i, j, n) - \bar{m}(i, j, n))^2}{4}}. \quad (5)$$

$\bar{m}(i, j)$ is the mean of the target $m(i, j)$ in five consecutive frames. $S(i, j)$ is the standard deviation of the target $m(i, j)$ in five consecutive frames.

Combined with (4), (5) analysis shows that target in foreground standard deviation is greater than that in background, so moving targets are segmented by setting threshold T , finally generating a new binary image. But the motion targets are mistaken for background noise with the moving targets' neighborhood pixels with no significant difference leading to small standard deviation. Or the background pixels are affected by high frequency noise which leads to big standard deviation; the background is mistaken for motion target. Aiming at this problem, the pixels information of target extension neighborhood in each frame (space-domain) is combined with five consecutive frames (time-domain), which improved the detection accuracy. In order to get enough pixel information and minimize the data computation, two pixels are selected in the four horizontal and vertical

directions of each target's extension neighborhood. Equations are expressed as follows:

$$\begin{aligned} f_n &= [m(i-2, j, n), m(i-1, j, n), m(i, j, n), \\ &\quad m(i+1, j, n), m(i+2, j, n)], \\ g_n &= [m(i, j-2, n), m(i, j-1, n), m(i, j, n), \\ &\quad m(i, j+1, n), m(i, j+2, n)]. \end{aligned} \quad (6)$$

f_n is the horizontal direction extension neighborhood of target; g_n is the vertical direction extension neighborhood of target.

Transverse and longitudinal section were formed by fusing of the time-domain and space-domain. Combined with (3), (6) analysis shows that the transverse and longitudinal section of each have $5 * 5 = 25$ pixels. Equations are expressed as follows:

$$\begin{aligned} F &= [f_n, f_{n+1}, f_{n+2}, f_{n+3}, f_{n+4}], \\ G &= [g_n, g_{n+1}, g_{n+2}, g_{n+3}, g_{n+4}]. \end{aligned} \quad (7)$$

F is the transverse section of target $m(i, j)$ generated in time-domain and space-domain; G is the longitudinal section of target $m(i, j)$ generated in time-domain and space-domain. The details are shown in Figure 2.

Figure 2(a) is as the position sketch map of the transverse and longitudinal section in consecutive frames. Figure 2(b) is

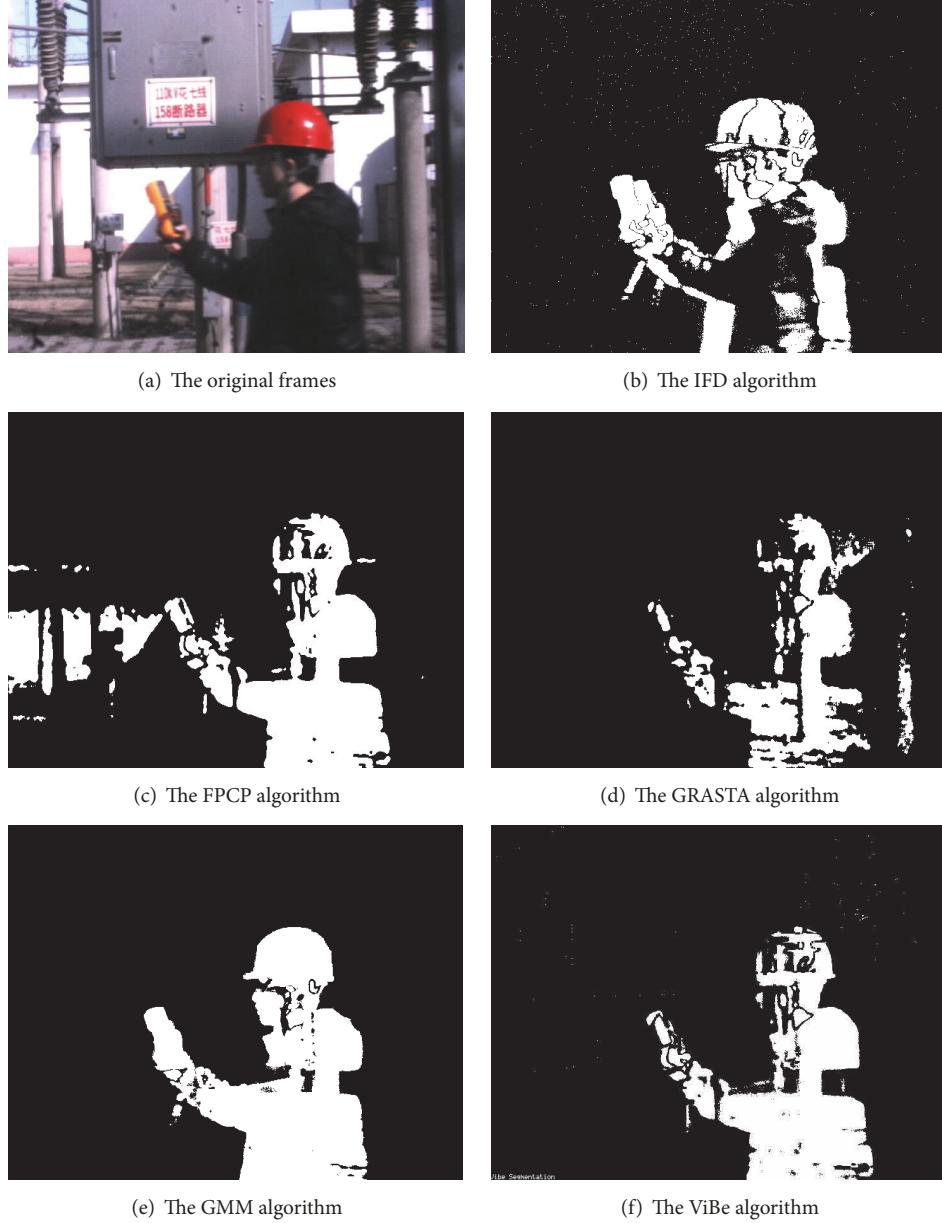


FIGURE 3: The common used detection algorithm results of outdoor video.

as the magnification view of the transverse and longitudinal section. Each point in the figure represents a pixel. Longitudinal section consists of 25 pixels; transverse section is the same. Dark blue points represent the position of the target $m(i, j)$ in the 5 frames.

Combined with (4) and (5), the mean and standard deviation of target pixels on five consecutive frames (time-domain) expand to the transverse and longitudinal section (time-domain and space-domain). As shown in Figure 2, each transverse section contains five sets of data, each group of data with 5 samples; each sample is the grey level of each pixel. Longitudinal section is the same. There are 10 groups in total. The standard deviation for each set of data was resolved firstly, and then mean of the standard deviation of 10 sets of data can

be solved. If the mean is less than the threshold T (2 as the threshold in experiments), then the target is regarded as the background. The mean of the target $m(i, j)$ in five consecutive frames in (4) is as a new grey value in the corresponding location of the new image.

$$\bar{S} = \frac{\sum_{i=i-2}^{i+2} S(i, j) + \sum_{j=j-2}^{j+2} S(i, j)}{10}, \quad (8)$$

$$m_{\text{new}} = \bar{m}(i, j), \quad \text{if } \bar{S} < T. \quad (9)$$

\bar{S} is the mean of the standard deviation of 10 sets of data. m_{new} is the new grey value into the corresponding location of the new image.

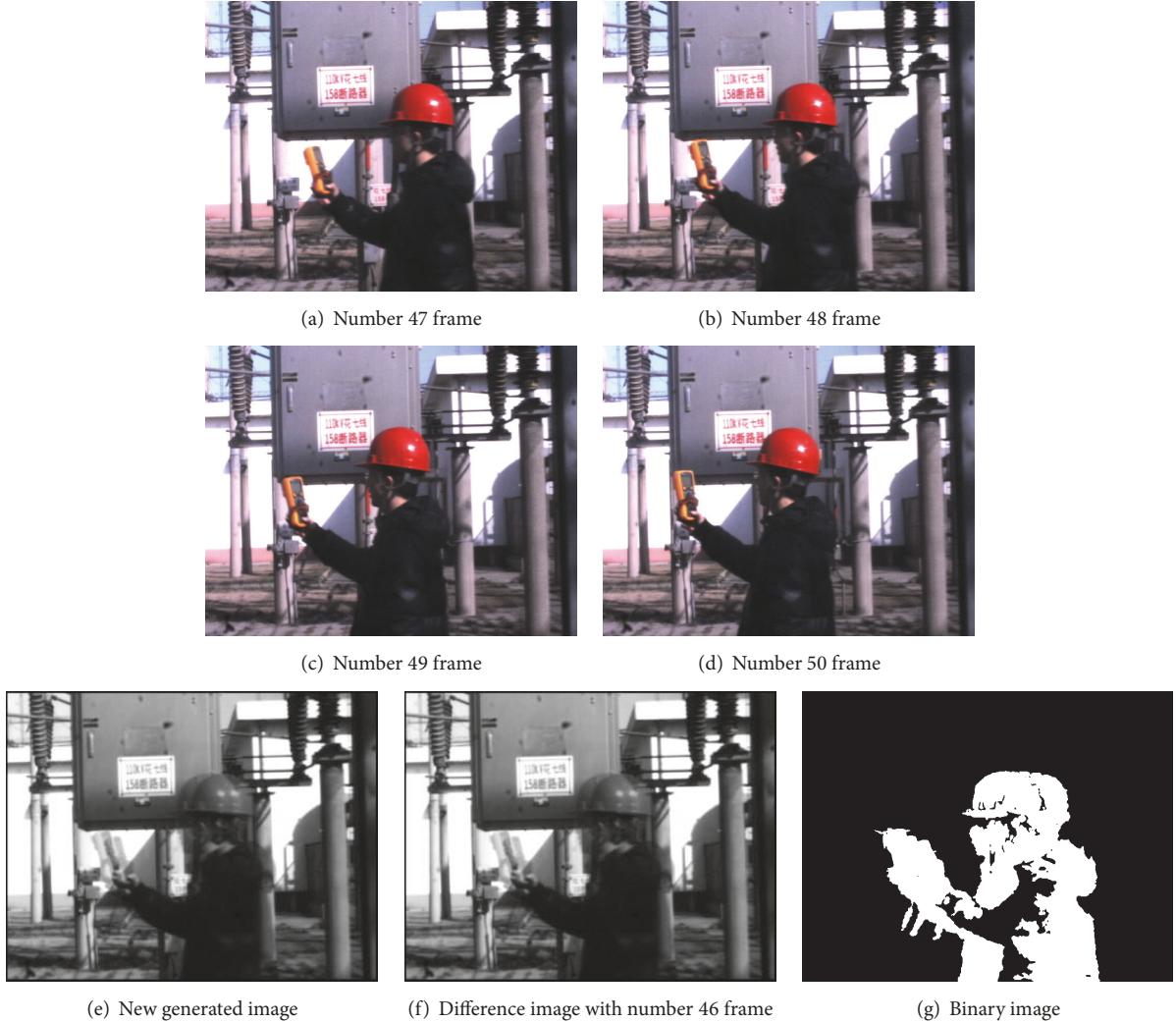


FIGURE 4: The MFITS algorithm results of outdoor video.

If the mean is greater than the threshold T , the target is regarded as the motion object. The median of the data of the transverse and longitudinal section is calculated separately by median filtering. The median of the transverse and longitudinal section is as a new grey value in the corresponding location of the new image. In the traditional median filter, we usually calculate the median value by a $3 * 3$ or $5 * 5$ square template, where

$$W_1 = \begin{Bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{Bmatrix},$$

$$W_2 = \begin{Bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{Bmatrix}. \quad (10)$$

In this paper, the pixels in transverse and longitudinal section need to median filter, respectively; then the mean of the two medians is as the new grey value. The weight of the traditional median filter all is 1 (shown in (10)). Because the pixels in transverse and longitudinal section are different from the traditional image plane, the same weight is not suitable for this algorithm. So we proposed an improved median filtering algorithm [18], to find out the most suitable pixel as the new pixel into the corresponding location of the new image to best represent the motion object. $5 * 5$ square template in (5) is still adopted, but each pixel in the template is combined with the median filter and mean filter given different weights; finally the new grey value is calculated. Traditional median filter is given the weight of 0.3; the mean filter is given the weight of 0.7. The new median value calculated by improved median filter both reflects the changed pixels of motion object and suppresses the background noise to a great extent, where

$$C_1 = N_{\text{med}}(F) = 0.3 * \text{med}(F) + 0.7 * \text{mean}(F),$$



FIGURE 5: The commonly used detection algorithm results of indoor video.

$$\begin{aligned}
 C_2 &= N\text{med}(G) \\
 &= 0.3 * \text{med}(G) + 0.7 * \text{mean}(G), \\
 m_{\text{new}} &= \frac{C_1 + C_2}{2}, \\
 &\quad \text{if } \bar{S} > T.
 \end{aligned} \tag{11}$$

C_1 is as the median value of transverse section, C_2 is as the median value of longitudinal section, med is as the traditional median filter, $N\text{med}$ is as the improved median value, mean is as the traditional mean filter, and m_{new} is as the new grey value in the corresponding location of the new image.

The new grey value of pixel fills the corresponding location to complete a new image. Then we calculate binary images of difference images for the new image and the five consecutive frames, respectively. Information of motion objects can be extracted.

4. Experiment

In this section, several typical object detection algorithms were presented, and a comparison of their features was made with the MFITS algorithm. Shooting video equipment is the CCD industrial camera in Microvision Co., Ltd., model MV-VS220, resolution of $720 * 480$; the computer used in experiment is Intel I5 processor, 2 GB of memory, Windows

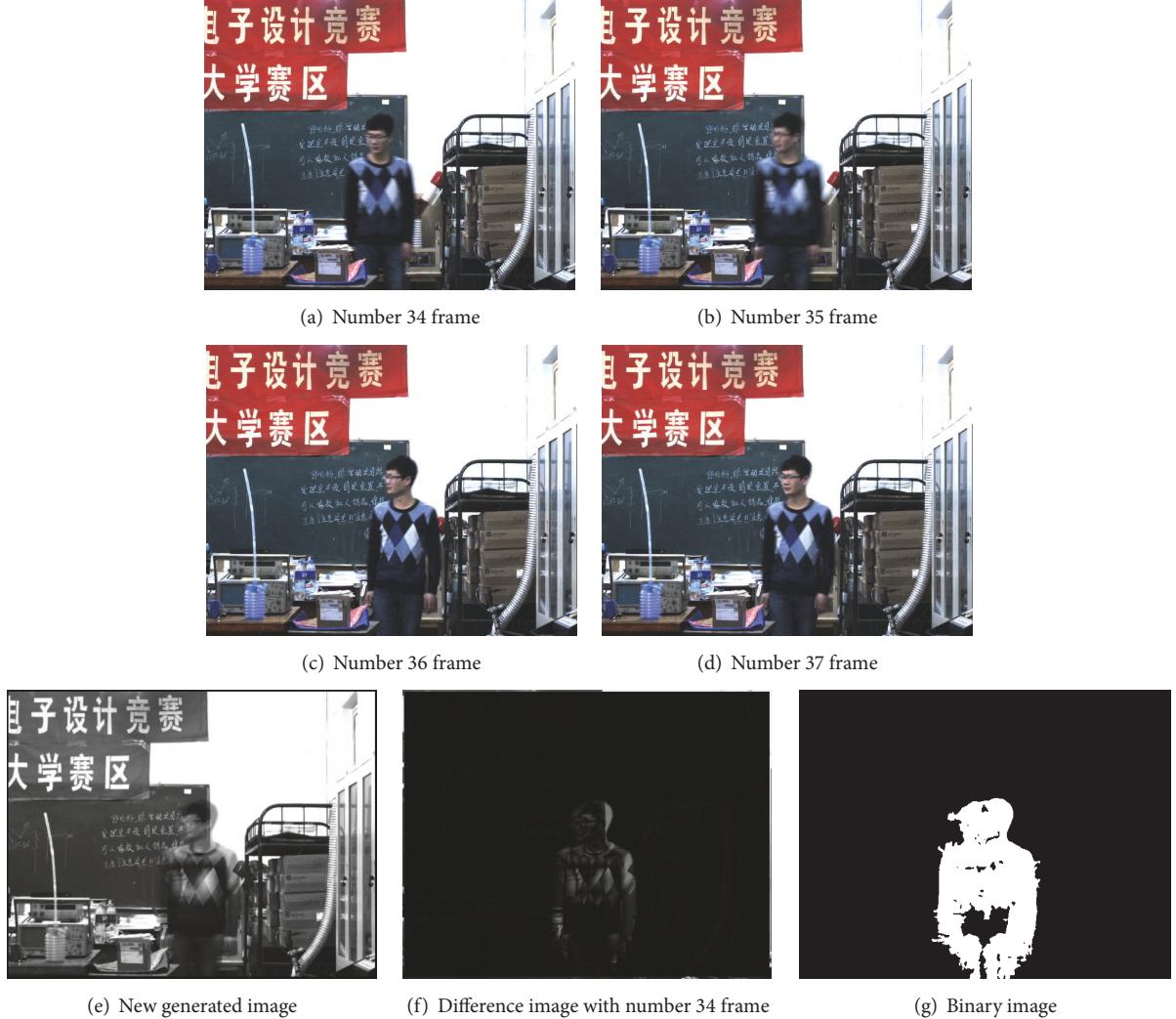


FIGURE 6: The MFITS algorithm results of indoor video.

7 operating system. The programs are compiled with Matlab 2014.

Experiments selected number 46 frame as the key frame in the outdoor video with small background disturbance. For comparison, experiments also selected number 34 frame as the key frame in the indoor video with large background disturbance, number 41 frame in shop video, and number 17 frame in escalator video. The shop and escalator video were both sourced from *LRS-library*. The experiments compared kinds of commonly used algorithms with the key frames. In the multiframes integration object detection algorithm based on time-domain and space-domain, we need to deal with multiple consecutive frames of video; meanwhile to facilitate comparing with other algorithms, number 46–50 frames of the outdoor video, number 33–37 frames of the indoor video, number 39–43 frames of the shop video, and number 16–20 frames of the escalator video were chosen separately as the key frames. The binary images of different algorithms were listed to get the simpler, intuitive comparing result. Binary

image threshold is grey value 10. In Gaussian mixture model (GMM) algorithm, the number of Gaussian filters is 3; the initial background modeling frames are 10; learning rate is 0.7. In ViBe algorithm, the number of sample adjacent pixels is 20; the threshold of matching points number is 20 and $\# \min = 2$; update rate is 16. Details are shown in Figures 3–6.

To compare and analyze these kinds of algorithm more efficiently and accurately, this paper adopted the accuracy rate and recall rate as the quantitative indicators. The equations were as follows:

$$P = \frac{N_J}{N_Z}, \quad (12)$$

$$R = \frac{N_J}{N_G}.$$

P is as the accuracy rate, R is as the recall rate, N_J is as the correctly detected moving object pixel, N_Z is as the actually detected moving object pixel, and N_G is as the

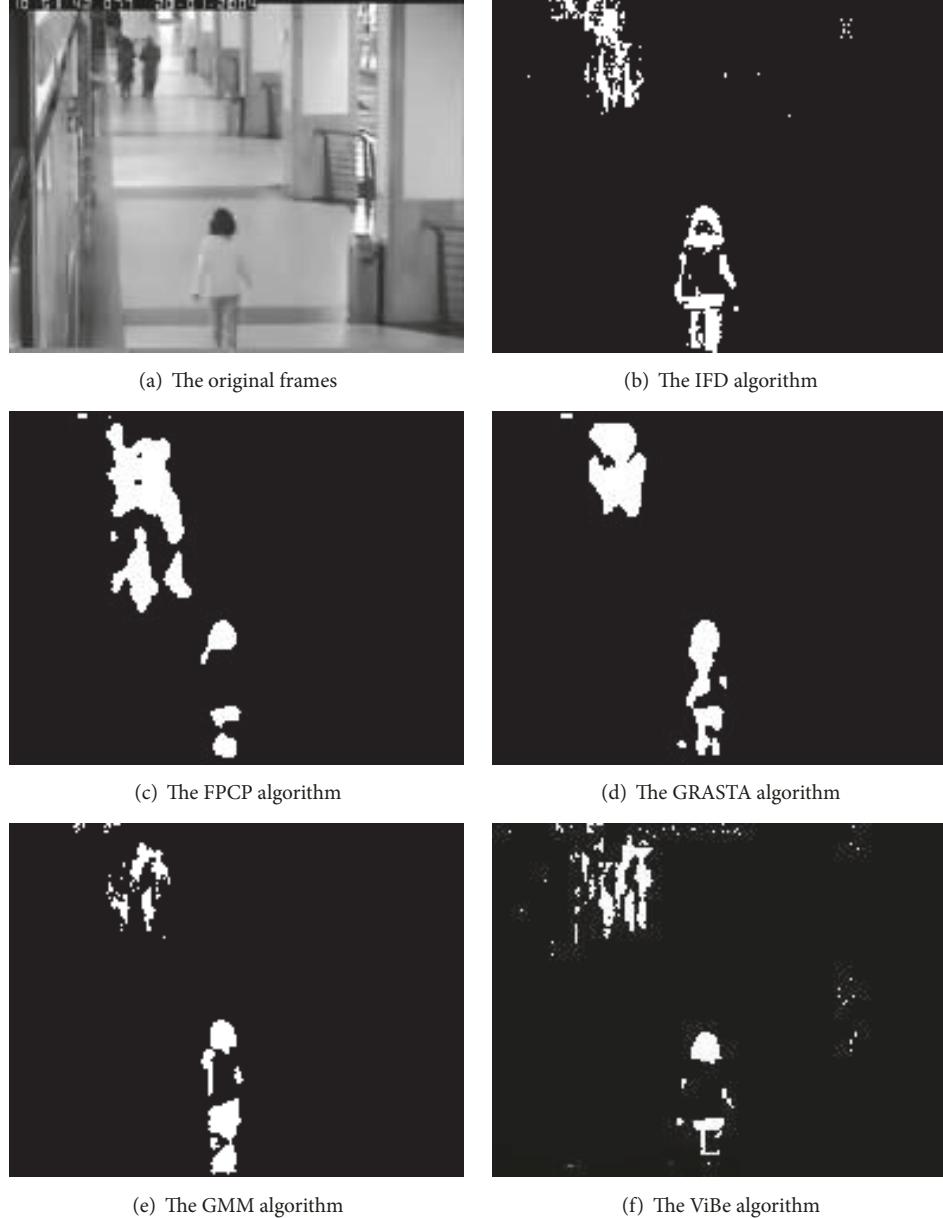


FIGURE 7: The common used detection algorithm results of shop video.

ground truth moving object pixel. Accuracy rate reflected the denoising performance of the algorithms. Recall rate reflected the retention of the moving object information. Details were shown in Table 1.

Comparing with these kinds of algorithm combined with Figures 3–10 and Table 1, interframe difference (IFD) method can remove noise effectively (mean of accuracy rate was 91.63%), but the object information retention performs poorly (mean of recall rate was 61.37%). The FPCP algorithm performed generally in all aspects. The GRASTA algorithm showed that either background denoising effect (mean of accuracy rate was 78.19%) or object information retention (mean of recall rate was 70.52%) performed well compared

with FPCP algorithm. However, the algorithm performed worse in denoising when the background noise disturbance was large (recall rate of the escalator video was 42.61%). The GMM algorithm showed that either background denoising effect or object information retention performed well when the background noise disturbance is small (outdoor video). However, the algorithms for object information reserving have their shortcomings to a certain extent when the background noise disturbance is large (the other three videos). The ViBe algorithm had the lowest recall rate (57.11%) compared with these algorithms. The MFITS algorithm showed stable performance with different conditions and higher robustness. The MFITS algorithm had the optimal recall rate (82.98%),

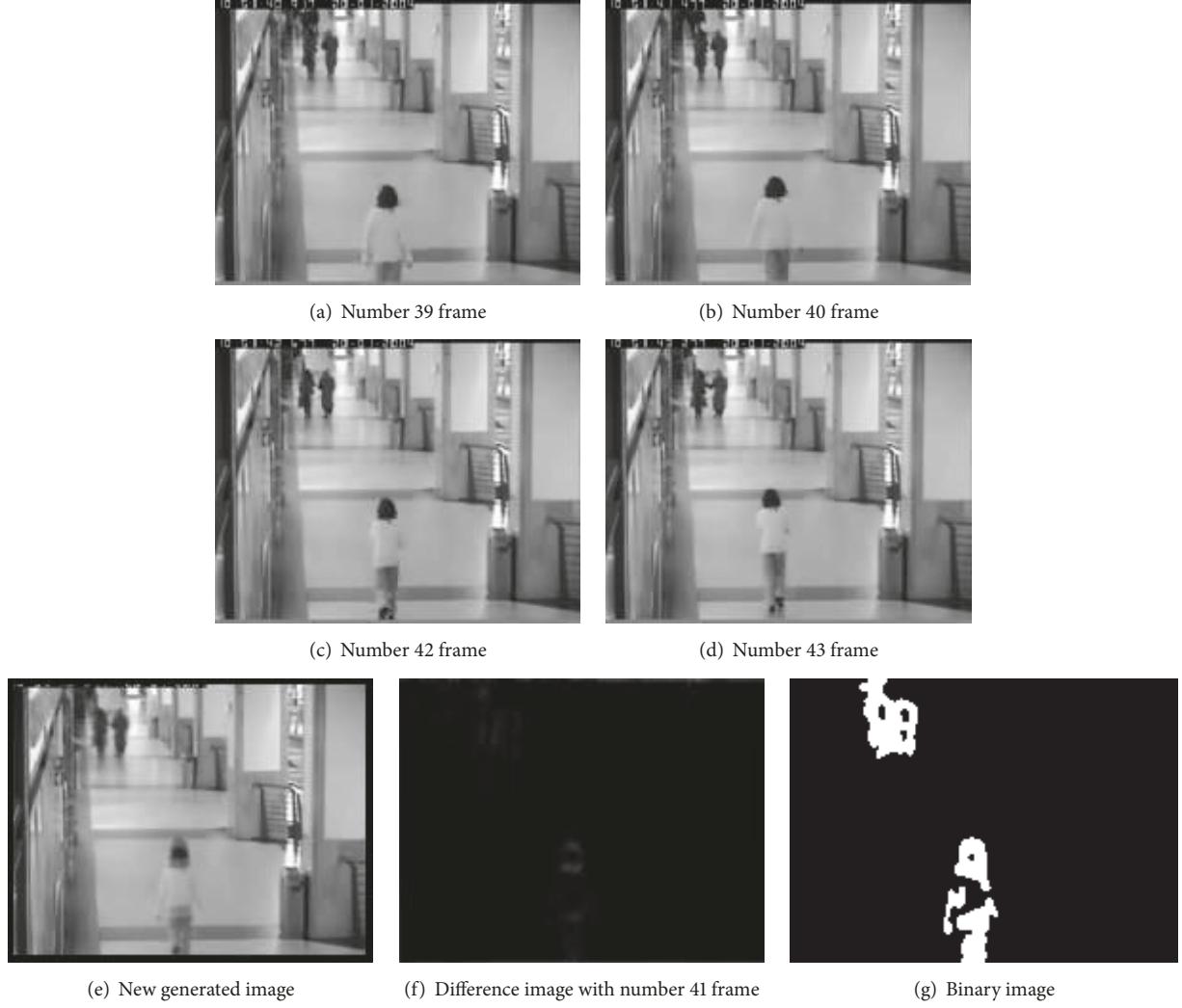


FIGURE 8: The MFITS algorithm results of shop video.

TABLE 1: The performance of different algorithms through precision and recall rate.

Algorithm	IFD	FPCP	GRASTA	GMM	ViBe	MFITS
Outdoor video						
Accuracy rate	94.92%	67.53%	83.62%	99.22%	96.19%	94.99%
Recall rate	64.22%	73.69%	67.12%	91.11%	71.83%	86.27%
Indoor video						
Accuracy rate	94.45%	61.47%	90.46%	95.98%	92.06%	95.25%
Recall rate	60.47%	55.83%	61.73%	71.66%	56.29%	75.98%
Shop video						
Accuracy rate	91.36%	64.25%	95.86%	98.77%	91.34%	93.54%
Recall rate	57.41%	68.01%	71.91%	70.33%	55.16%	79.81%
Escalator video						
Accuracy rate	85.77%	56.54%	42.61%	93.48%	77.32%	97.16%
Recall rate	63.38%	68.46%	81.33%	40.17%	45.62%	89.84%
Mean						
Accuracy rate	91.63%	62.44%	78.19%	96.86%	89.22%	95.24%
Recall rate	61.37%	66.49%	70.52%	68.32%	57.11%	82.98%

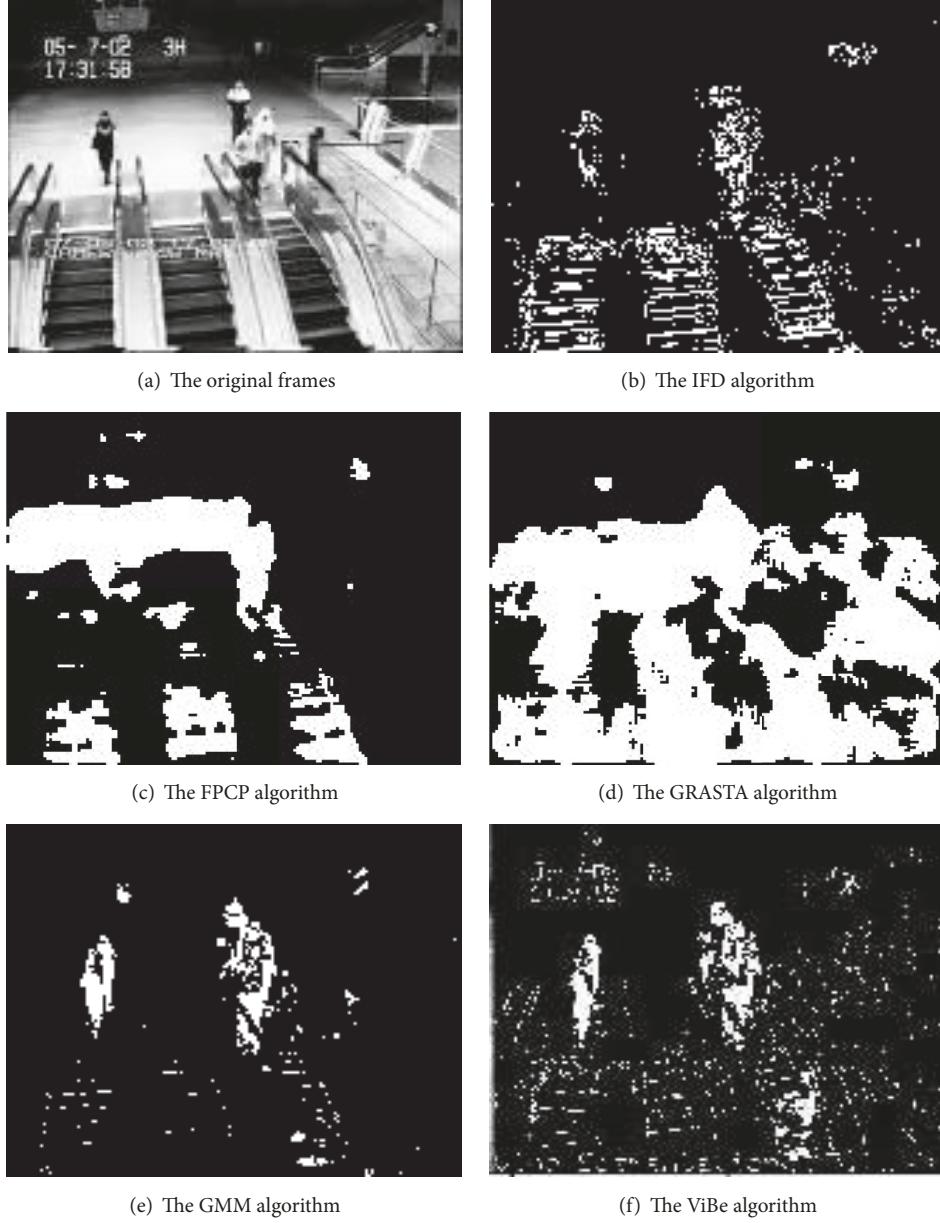


FIGURE 9: The common used detection algorithm results of escalator video.

which is promoted a lot compared with these kinds of algorithms. Denoising effect was almost the same as the GMM algorithm (mean of accuracy rate was 95.24%).

According to (9), (11) in new image generation, threshold selection is very important. The data was compared with the three different thresholds (0.5, 2, and 4). In order to facilitate observation, new images generated by three different thresholds are made by the binarization processing. Details are shown in Figure 11.

Figure 11 shows the image processing effect in the different threshold. When the threshold T value is 0.5 in Figure 11(a), the generated image contains a large amount of background noise. When the threshold T value is 4 in Figure 11(c),

background suppression effect is good, but part of the moving object information is lost compared with the two previous images (center area of the moving object). When the threshold T value is 2 in Figure 11(b), the generated image not only retained the effective information to the utmost extent, but also suppressed background noise effectively. In conclusion, this paper selected the threshold T to 2.

Figure 12 shows the transverse and longitudinal section (center coordinates 350,246) in 46–50 frames of the outdoor video.

Figure 12(a) is the shown magnification of transverse section (actual size is 5 * 5 pixel size). Figure 12(b) is the shown magnification of longitudinal section. In Figure 12, the

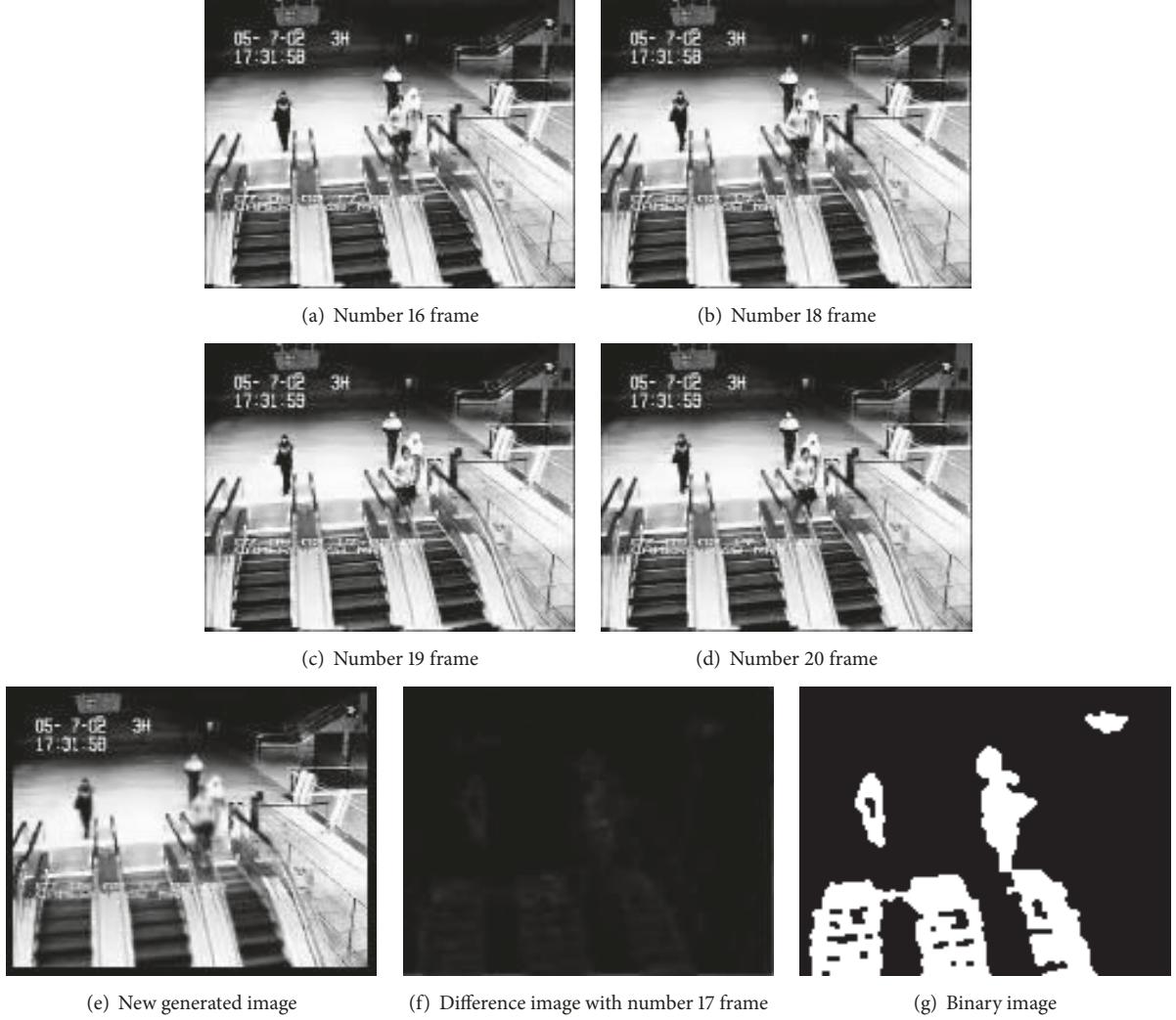


FIGURE 10: The MFITS algorithm results of escalator video.

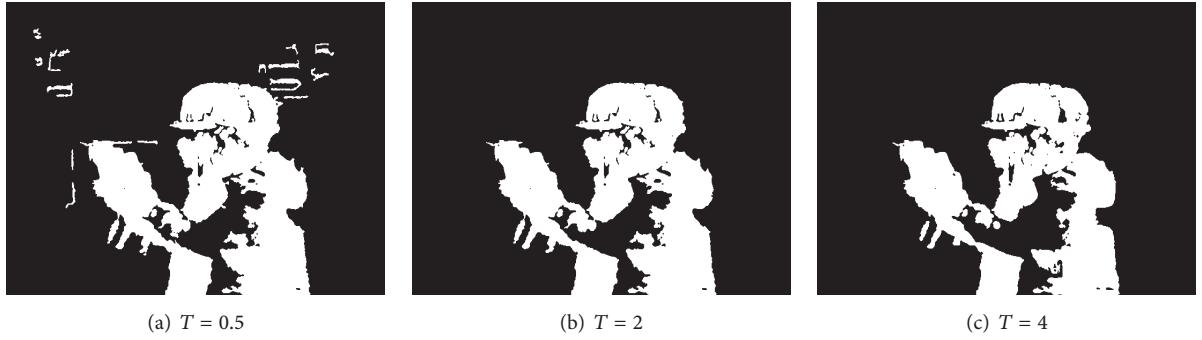


FIGURE 11: Image processing effect in the different threshold.

boundary line of transverse section is vertical; the boundary line of longitudinal section is horizontal. The transverse and longitudinal section do not exist with any motion object.

Aiming at the improved median filter we proposed in this paper the effect of generating new images compared with the

average filter, median filter, and improved median filter from outdoor video frames. Details are shown in Figure 13.

Figure 13(a) is the new image generated by mean filter; this filter retained the relatively complete foreground, but a lot of backgrounds are mistaken for the moving object, which

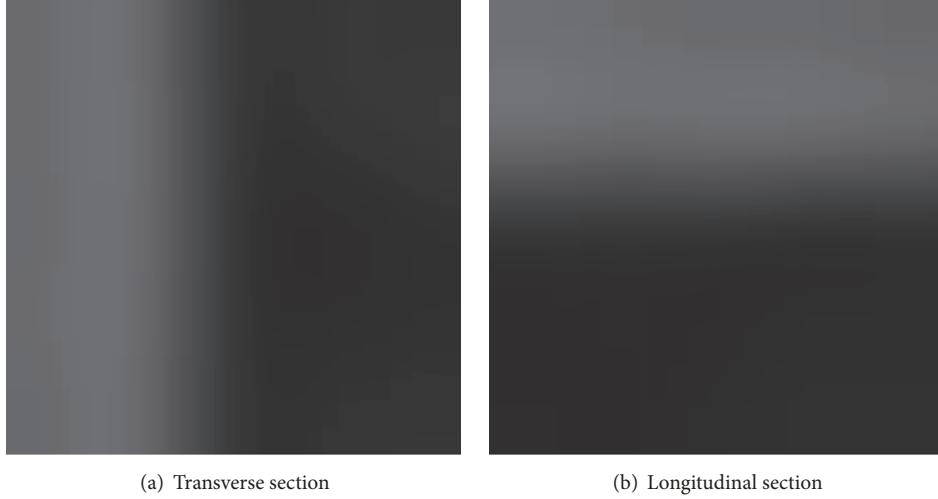


FIGURE 12: The plan sketch of transverse and longitudinal section.

TABLE 2: Performance comparison between improved median filter and traditional median filter.

Algorithm	MSE	PSNR
Traditional median filter in Lena	$3.0771e + 03$	13.2494
Improved median filter in Lena	$2.8540e + 03$	13.5763
Traditional median filter in video	$7.8654e + 03$	9.1736
Improved median filter in video	$7.5073e + 03$	9.3760

led to blurring the edge of the motion object. Figure 13(b) is the new image generated by traditional median filter; a lot of foregrounds are mistaken for the background, which led to loss of information of the motion object. Figure 13(c) is the new image generated by improved median filter; algorithm has the optimal effect of removing background and maximum reserving of the motion object information in these three kinds of filters.

In order to verify the denoising performance of the improved median filtering algorithm, we compared with the different median filter denoising effect in Lena image with salt-pepper noise and outdoor video frames. Experimental images adopt the mean square error (MSE) and peak signal to noise ratio (PSNR) as the evaluating criterion for image denoising effect and object information reserving effect. The mean square error method gauges the distortion degree of the image by calculating the mean square error between the original frame and denoised frame. The smaller value indicates that algorithm for noise suppression effect is better. Peak signal to noise ratio method is the value of maximum signal and intensity of the noise. The bigger value indicates that algorithm for noise suppression effect is better. Details are shown in Figure 14 and Table 2.

Comparing with the different median filter denoising effect in Figure 14 and Table 2, the improved median filter is superior to traditional median filter in denoising effect and object information retention effect.

Because the transverse and longitudinal section of multiframe need to be traversed in MFITS algorithm, the computational time and the complexity have increased. In order to analyze the complexity of the MFITS algorithm, the computational time of different algorithm processing in different videos was compared separately. Details were shown in Table 3.

Comparing with the computational time of different algorithms in Table 3, the MFITS algorithm improved motion foreground detection accuracy and completeness but spends much time on computation. The computational time and the complexity of MFITS algorithm will be improved in the following study.

5. Conclusion

In this paper, a multiframe integration object detection algorithm based on time-domain and space-domain (MFITS) was presented aiming at the advantages and disadvantages of commonly used object detection algorithms. The MFITS algorithm is different from commonly used algorithms using difference between frames. Instead of extracting sequential N frames of video, it forms a pair of new images constituted with N frames transverse and longitudinal section. Then the new image is formed by traversing through the multiframe. We also added an improved median filter combined with the traditional median filter and mean filter. The experiment results showed that the MFITS algorithm reserves the object information well and removed the background to the maximum. The MFITS algorithm also has strong robustness in dealing with different situations of video frames.

Conflicts of Interest

The authors Yifan Liu, Zhenjiang Cai, and Xuesong Suo declared that there are no conflicts of interest regarding the publication of this paper.



FIGURE 13: The image progress of different filter.

TABLE 3: The computational time of different algorithms.

Algorithm	IFD	FPCP	GRASTA	GMM	ViBe	MFITS
Outdoor video	3.29	53.52	83.78	7.23	10.35	143.75
Indoor video	2.17	40.29	68.25	6.62	8.77	126.26
Shop video	0.95	3.78	4.38	3.48	3.62	23.41
Escalator video	1.13	3.51	1.77	4.45	3.09	22.03

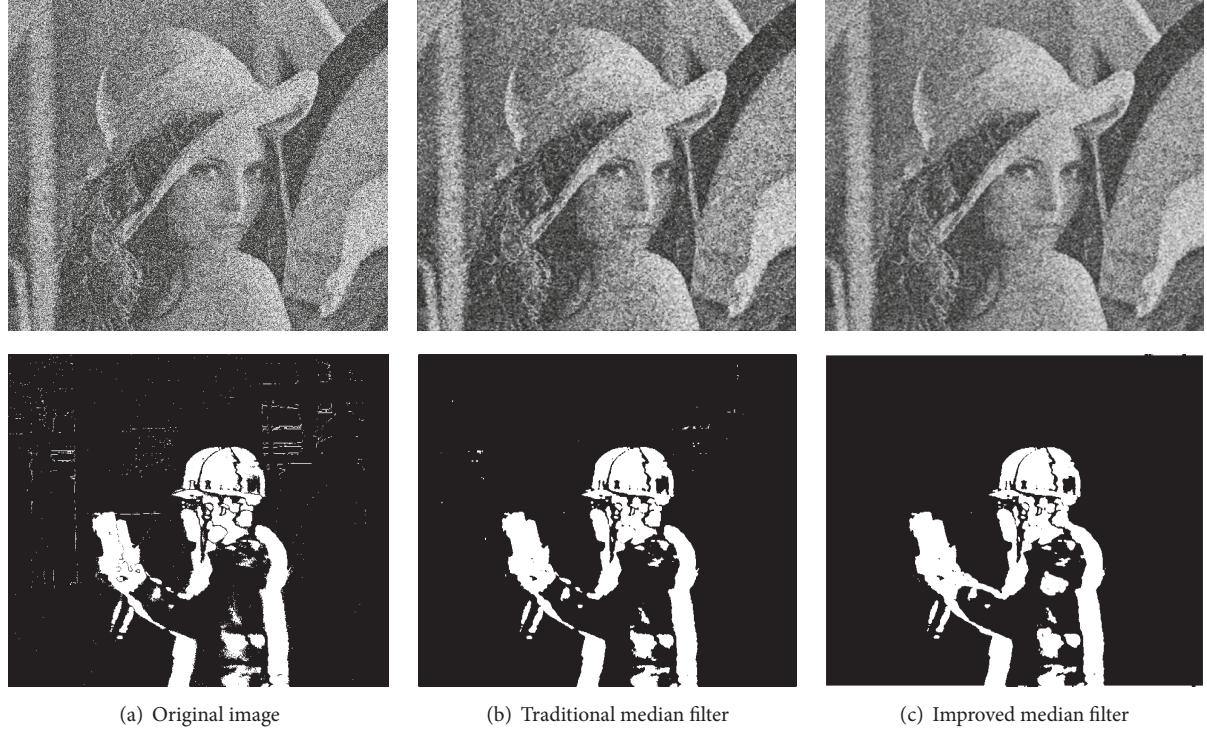


FIGURE 14: The image progress of different median filter.

Acknowledgments

This paper is supported by the Hebei Province Key Research and Development Project (17227206D).

References

- [1] Y. Benerezeth, P. M. Jodoin, B. Emile, H. Laurent, and C. Rosenberger, "Review and evaluation of commonly-implemented background subtraction algorithms," in *Proceedings of the 19th International Conference on Pattern Recognition (ICPR '08)*, pp. 1–4, Tampa, Fla, USA, December 2008.
- [2] Z. Yusi, *The Illumination Robustness Based on Background Subtraction Motion Target Detection and Tracking Technology Research*, Southwest Jiaotong University, Sichuan, China, 2011.
- [3] L. Chen and Z. Jibei, "Symmetric difference and background difference method based on dynamic threshold of motion object detection algorithm," *Application Research of Computers*, vol. 25, no. 2, pp. 488–494, 2008.
- [4] E. Stringa and C. S. Regazzoni, "Real-time video-shot detection for scene surveillance applications," *IEEE Transactions on Image Processing*, vol. 9, no. 1, pp. 69–79, 2000.
- [5] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, "Wallflower: principles and practice of background maintenance," in *Proceedings of the 7th International Conference on Computer Vision (ICCV '99)*, vol. 1, pp. 255–261, Kerkyra, Greece, September 1999.
- [6] P. Rodriguez and B. Wohlberg, "Fast principal component pursuit via alternating minimization," in *Proceedings of the 2013 20th IEEE International Conference on Image Processing, ICIP 2013*, pp. 69–73, Australia, September 2013.
- [7] P. Rodriguez and B. Wohlberg, "Incremental principal component pursuit for video background modeling," *Journal of Mathematical Imaging and Vision*, vol. 55, no. 1, pp. 1–18, 2016.
- [8] J. He, L. Balzano, and A. Szlam, "Incremental gradient on the Grassmannian for online foreground and background separation in subsampled video," in *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2012*, pp. 1568–1575, USA, June 2012.
- [9] Y. Chen, K. Ren, G. Gu, W. Qian, and F. Xu, "Moving object detection based on improved single Gaussian background model," *Chinese Journal of Lasers*, vol. 41, no. 11, Article ID 1109002, pp. 245–253, 2014.
- [10] W. Tong and W. Ling, "Based on the frame difference of mixed Gaussian background model," *Computer Engineering and Application*, vol. 50, no. 23, pp. 176–180, 2014.
- [11] H. Yuanlei and L. Wanjun, "Improvement of the gaussian mixture model moving target detection algorithm," *Computer Application*, vol. 34, no. 2, pp. 580–584, 2014.
- [12] L. Yanan, Zhouyong, and T. Ruijuan, "Background modeling based on gaussian mixture model and three frame difference," *Ordnance Industry Automation*, vol. 34, no. 4, pp. 33–35, 2015.
- [13] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, "Real-time foreground-background segmentation using codebook model," *Special Issue on Video Object Processing*, vol. 11, no. 3, pp. 172–185, 2005.
- [14] T. Sun, S.-C. Hsu, and C.-L. Huang, "Hybrid codebook model for foreground object segmentation and shadow/highlight removal," *Journal of Information Science and Engineering*, vol. 30, no. 6, pp. 1965–1984, 2014.
- [15] O. Barnich and M. Van Droogenbroeck, "ViBe: a universal background subtraction algorithm for video sequences," *IEEE*

Transactions on Image Processing, vol. 20, no. 6, pp. 1709–1724, 2011.

- [16] Y. Qin, S. Sun, X. Ma, S. Hu, and B. Lei, “A background extraction and shadow removal algorithm based on clustering for ViBe,” in *Proceedings of the 13th International Conference on Machine Learning and Cybernetics, ICMLC 2014*, pp. 52–57, China, July 2014.
- [17] A. Sanin, C. Sanderson, and B. Lovell, “Shadow detection: a survey and comparative evaluation of recent methods,” *Pattern Recognition*, vol. 45, no. 4, pp. 1684–1695, 2012.
- [18] Z. Gaochang, Z. Lei, and W. Fengbo, “The improved median filtering algorithm in the application of image denoising,” *Applied Optics*, vol. 32, no. 4, pp. 678–682, 2011.

