Moving target detection (MTD) is one of the emphases and difficulties in the field of computer vision and image processing. It is the basis of moving target tracking and behavior recognition. We propose two methods are improved and fused, respectively, and the fusion algorithm is applied to the complex scene for MTD, so as to improve the accuracy of MTD in complex and hybrid scenes. Using the main idea of the three-frame difference image method, the background difference method and the interframe difference method are combined to make their advantages complementary to overcome each other’s weaknesses. The experimental results show that the method can be well adapted to the situation of periodic motion interference in the background, and it can adapt to the situation of sudden background changes.

1. Introduction

In recent years, although the methods of video moving target detection (MTD) have been deeply studied and some problems existing in algorithms have been improved, most of these algorithms are for specific applications [1]. It is the follow-up tracking and recognition of moving objects and the basis of semantic understanding of moving object sequences [2–4]. For applications such as encoding and decoding of key information in the video, the moving target of interest in the video sequence needs to be extracted from the background, so as to achieve compressed storage of the video image [5–7]. Intelligent video surveillance [8] is one of the most typical applications of video MTD at home and abroad in recent years, and MTD has been applied to video surveillance, UAV target detection, and other fields [9–11].

In foreign intelligent video surveillance system developed earlier, as early as the 1960s, many universities and research institutions in the United States engaged in MTD, tracking, and abnormal behavior recognition and other fields to carry out research. One of the most representative is the developed at Carnegie Mellon University, and other well-known colleges and universities in the United States dedicated to the detection, identification, and analysis of human behavior and can realize the prediction of human action; according to the forecast results, we judge the abnormal behavior of pedestrian is going to happen and give warning and prompt the use of such as public surveillance system. At the same time, VSAM (visual surveillance and monitoring) is mainly used for unmanned operations in hostile environments, such as severe disasters, long-term battlefield monitoring, and effective detection and tracking of enemy moving targets in military situations [12]. Subsequently, the University of Maryland in the United States independently developed a real-time visual monitoring system W4 (Who? When? Where? What?) [13] to achieve the positioning, segmentation, and multitarget tracking of the moving human body and to simply judge the next human behavior by detecting whether the human body is carrying other items. The Defense Advanced Research Agency’s Airborne Video Surveillance (AVS) program uses target detection technology to capture static or dynamic moving target information from video in real time and then combines moving target tracking and recognition technology to obtain moving target speed, position, and trajectory information. Using such moving target information can improve the real-time detection function of UAV for specific environment [14].
Domestic universities and key research institutes have gradually studied the intelligent video surveillance system since the 1990s. There are many foreign experimental projects that are only under simulated conditions. Now many of them have been transformed into products, and they are widely used. The domestic market is not to be outdone. Now some fields can be independently innovated, and many products are gradually put into use. From automation, the Chinese Academy of Sciences, institute of independent research, develops a set of traffic monitoring prototype system [15] to detect moving vehicles, based on the license plate number, and automatically identifies and analyzes if there is any cause to track the vehicle, and then, tracking the identified vehicle, even automatically identifies and analyzes if there is any cause to detect moving vehicles, based on the license plate number, and develops a set of traffic monitoring prototype system [15] to obtain, such as the vehicle monitoring system of highway.

2. Concept Review

2.1. Extraction of Background. There are usually three methods to obtain background images: manual rendering, statistical method, and Surendra background update algorithm.

(1) The manual background method requires people to start the camera to obtain the background image when no foreground object is observed. This method of background extraction increases the demand of manpower and material resources, and the background image without the foreground is hard to obtain, such as the vehicle monitoring system of highway.

(2) Statistical methods: background images can be summed up as in a specific time period of pixel grayscale average and use this as a background image corresponding to the pixel point of the average gray; when read in a video of a pixel point observation, we will find that there is no prospect; through a moving target point gray value which remained stable, the change is very small; only when it prospects through a moving target, the grayscale of the point will change dramatically.

(3) Surendra background update algorithm: we can adaptively obtain the background image; the algorithm to extract the background idea is to find the movement of objects through the current poor Zhen image area; movement area on the background remains the same, rather than the sports area with the background of the current frame to replace updates. The algorithm steps are shown below:

(1) Image 1 of frame 1 is taken as background B.
(2) Select I value $T$, the number of generations $m = 1$, MAX STEPS.
(3) Find the frame difference image of the current frame:

$$D = \begin{cases} 
1, & |I_i - I_{i-1}| \geq T, \\
0, & |I_i - I_{i-1}| < T,
\end{cases}$$

(1)

where we denote $I_i$ to be the $i^{th}$ frame.

(4) Update background image B from binary image $D$:

$$B(m, n) = \begin{cases} 
B_{i-1} (m, n), & D(m, n) = 1, \\
\alpha I_i + (1 - \alpha)B_{i-1} (m, n), & D(m, n) = 0,
\end{cases}$$

(2)

where $B_i$ denotes the current background which is set to be the weighted average of the instantaneous background and previous background and $\alpha$ is a background update coefficient.
2.2. Detection of Moving Targets. Under normal circumstances, there is a great difference between the foreground and the gray value of the background, while the gray value of moving objects itself will not be very different. Thus, the current frame image is $I$ and background image is $B$; the background differential binary image is

$$DB(m, n) = \begin{cases} 255, & I(m, n) - B(m, n) \geq T, \\ 0, & I(m, n) - B(m, n) < T, \end{cases}$$

where we denote by $T$ to be the threshold which is given by the corresponding pixel value of searching toward increasing pixel intensity. In this way, pixels with the gray value of 255 in the differential binary image can be regarded as moving target points in the foreground.

2.3. Image Postprocessing. Due to the influence of noise, some points belonging to the background will be wrongly detected as moving targets in the foreground. Meanwhile, due to the slight disturbance of objects in the background, these background points will also be wrongly identified as moving target points. In order to eliminate these influences, it is necessary to process the difference images of foreground and background. For example, the mathematical morphology method is used to process the difference images. Because the shadow pixels and the pixels of the moving target is the same visual feature, the background of the direct reduction method cannot distinguish between moving objects and shadows; normally, the shadow will be wrongly judged to be moving targets; in RGB space, people’s perception of the difference and the consistency of the calculated difference are very poor. Calculate the color similarity between images. Therefore, shadow detection is suitable for HSV color space. When a pixel is covered by shadow, its brightness value changes greatly, while chromaticity signal changes little. In this way, the following method can be used for discrimination:

$$S(x, y) = \begin{cases} 1, & \alpha \leq \frac{I(x, y)}{B(x, y)} \leq \beta \cap (I(x, y) - B(x, y)), \\ 0, & \text{other}, \end{cases}$$

where $I(x, y)$ is the current frame, $B(x, y)$ is the background image, $S(x, y)$ is the shaded pixel point, and $A$ is set as the influence coefficient of the light source. The stronger the light source is, the smaller the value of “$A$” is. $B$ is set to remove the influence of noise.

3. Frame Difference MTD and Improved Algorithm

For video image sequence, there is continuity between adjacent frames, so it can be considered that the illumination of the two frames is basically unchanged in the case of a very short time interval. By calculating the difference of adjacent frames, the amount of the gray value change between corresponding pixels of adjacent frames can be calculated. Because of the moving object in video before and after the two frames relative to the movement of background image pixel values of the same as a prime spot for two frames of pixel values before and after the change amount is larger, we exercise background pixel values of a point. In the case where the pixel values of the two frames before and after the volume change are small, we can separate the moving target and the background according to the frame difference method. The frame difference method is the simplest MTD method among all the current MTD methods.

3.1. Two-F Difference Method. First of all, two consecutive frames of images obtained after filtering are represented by $I_{k-1}(x, y)$ and $I_k(x, y)$, respectively, from the extracted video sequence.

Secondly, the obtained two adjacent video sequence images are used to carry out difference operation to obtain the difference images $D(x, y)$:

$$D(x, y) = |I_{k-1}(x, y) - I_k(x, y)|.$$  \hspace{1cm} (5)

If the threshold value $T$ is small, the background information will be regarded as a part of the moving target, reducing the accuracy of MTD. In this paper, according to the dynamic adaptive generation of multiple segmentation thresholds for each frame of the video sequence, segmentation thresholds are obtained by adapting to the changes of foreground and background contrast. The binarization process can be expressed by

$$T_k(x, y) = \begin{cases} 0, & D(x, y) < T, \\ 1, & D(x, y) \geq T. \end{cases}$$  \hspace{1cm} (6)

The principle block diagram of two-FDM is shown in Figure 1:

Below is the experiment of two-FDM using laboratory video dataset, as shown in Figure 2:

In Figure 2, (a) is frame 97 of the video sequence, (b) is frame 98 of the video sequence, and (c) is the difference between frame 97 and frame 98 of the picture. Through the experimental results, we can see the simple algorithm and fast detection speed and other characteristics. Disadvantages: when the time interval is long or the object moves too fast, the phenomenon of “double shadow” appears in (c) in Figure 2. At the same time, if the moving target moves too slowly or the time interval is too short, the pixel values of the two frames before and after cannot be changed, resulting in the difference of pixel values basically unchanged.

3.2. Three-Frame Difference Method (TFDM). The moving target area obtained by the two-FDM is larger than the actual target, resulting in the phenomenon of “double shadow”. The binarization method of two-FDM can be used, as shown in formula (6) above, to obtain the final three-frame difference in the middle frame, as shown in formula (7).

Principle block diagram of TFDM is shown in Figure 3. The specific process of MTD by TFDM is as follows.
First of all, Figure 3 shows that two difference images are obtained by difference operation on three consecutive frames of images.

Secondly, the two difference images obtained are thresholding by using formula (6), respectively, through the selected threshold value.

Finally, the binary foreground image $T(m, n)$ and $T(m, n)$ are calculated.

Laboratory video dataset was used for the experiment of TFDM, and the result is shown in Figure 4:

3.3. Canny Edge Detection. The common principle of edge detection is to use the discrete gradient approximation function to find the points with great changes in the gray value in the image, and we connect these points with great changes to form the edge information of the image.

The common steps of edge detection are as follows:

1. The applied edge detection of the filter is realized by the derivative of image intensity, and the derivative is greatly affected by voice. Therefore, it is necessary to reduce the influence of voice on the derivative of image intensity through filtering, so as to improve the performance of edge detection and reduce the amount of calculation.

2. Image enhancement is to realize the effect of image enhancement by enhancing the intensity value of pixel points that have obvious changes in the neighborhood of image gray points (the gradient
amplitude of the image can be calculated to determine which points have obvious changes and need to be enhanced) and determine the pixel points in the candidate region of the image edge.

(3) The image edge detection, due to image enhancement, will make the neighborhood within many of the gradient value bigger, so we get to the edge pixels than the actual real; we are looking for more edge pixels, we can use the threshold method to select the points, while only keeping the real pixels at the edge of the image.

3.3.1. Gaussian Filtering. Since the result of edge detection is easily affected by image noise. In order to make the image smooth, a Gaussian filter is used to process the image to achieve the purpose of noise reduction.

3.3.2. Calculate Gradient Amplitude and Direction. The gray gradient value of the image can be approximately replaced by the first-order finite difference score. The gradient and direction of the edge can be obtained.
The partial derivatives of $x$ and $y$ directions used by Canny algorithm to calculate the convolution template are as follows:

$$A_m = \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix}, B_n = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}. \quad (7)$$

The first-order partial derivative matrix of the image gray gradient value in the $x$ direction can be obtained as

$$\alpha[m, n] = \frac{(f[m, n + 1] - f[m, n] + f[m + 1, n + 1] - f[m + 1, n])}{2}. \quad (8)$$

The first-order partial derivative matrix of the image gray gradient value in the $y$ direction can be obtained as

$$\beta[m, n] = \frac{(f[m, n] - f[m + 1, n] + f[m, n + 1] - f[m + 1, n + 1])}{2}. \quad (9)$$

Formulas (8) and (9) can obtain the gradient amplitude of the edge:

$$X[m, n] = \sqrt{\alpha[m, n]^2 + \beta[m, n]^2}. \quad (10)$$

With formulas (8) and (9), the gradient direction of the edge can be obtained:

$$Y[m, n] = \arctan\left(\frac{\alpha[m, n]}{\beta[m, n]}\right). \quad (11)$$

$M[i, j]$ and $A[i, j]$ can be obtained by calculating the first-order partial derivative matrix of $x$ direction and $y$ direction. If the angle obtained is not within these several ranges, the formula can be converted to between $0^\circ$ and $360^\circ$.

3.4. Background Modeling Method. As the scene environment changes slightly over time, the background model needs to be updated in real time for the background subtracting method so that a good background model can accurately detect the moving target from the image.

At present, most background modeling methods are based on the improvement and optimization of the original algorithm. Common modeling methods are as follows:

(1) Single Gaussian background modeling: Gaussian background modeling is mainly used in complex scene environments, such as leaves shaking or camera shaking. The principle of the method is, respectively, for each pixel point in the original image to a single modeling, assumes that each pixel in the image for a period of time has nothing to do with other pixels, and is independent, and the characteristic value of the background pixels’ fluctuation in a period of time satisfies the Gaussian distribution, according to the reasonable hypothesis for background modeling.

(2) Mean method background modeling: if the video image sequence scene is not too complex, the mean method can be used for background modeling. In essence, the mean method is an idea of statistical filtering, which is realized by summing up the multiple frames captured by the camera within a period of time, dividing the cumulative value by the number of captured frames, and finally obtaining the average value, which is used as the background reference model.

4. Experimental Results and Analysis

The software environment of the experiment is windows7, combined with VS2010 and Opencv. Video sequences from frames 81 to 115 of laboratory dataset are used to select representative moving objects of frames 97, 98, and 99 from 35 frames to conduct the experiment of improving the TFDM, as shown in Figure 5.

In Figure 5, figures (a), (b), and (c) represent sequential video sequences from the Laboratory dataset. Figure (a) is frame 97 of the video sequence. Figure (b) is frame 98 of the video sequence. Figure (c) is frame 99 of the video sequence. Figure (d) is frame 98 and frame 97 difference diagram. Figure (e) is frame 99 and frame 98 difference diagram. Figure (f) is frame three difference diagram.

Analysis of experimental results is as follows:

(1) Figures (d) and (e) represent that the two-frame difference method shows that moving targets detected by the two-FDM have “double shadow” phenomenon. Figure (f) is the traditional TFDM. It performs the logical “and” operation on the two-frame difference images by making the difference between the front and back frames of the three adjacent frames and takes the common part of the moving target of the two-FDM. Figure (f) is the MTD result of the traditional TFDM.
It can be seen from Figure (g) that a relatively complete contour of the moving target can be obtained by using Canny edge detection algorithm. This method can not only obtain more complete contour information but also obtain part of the internal information of the moving target.

Table 1 compares the accuracy of MTD between the traditional TFDM and the improved frame difference method. Through experimental results and comparison experiments, we can see that the improved algorithm can get more accurate moving targets compared with the traditional TFDM.

Table 2: Comparison of processing time between the traditional TFDM and the improved TFDM.

| Current frame picture | Traditional TFDM (ms) | Improved TFDM (ms) |
|----------------------|-----------------------|-------------------|
| Laboratory           |                       |                   |
| The 97th frame       | 3.15043               | 4.93475           |
| The 98th frame       | 3.30144               | 4.95418           |
| The 99th frame       | 3.22263               | 5.10116           |

(2) It can be seen from Figure (g) that a relatively complete contour of the moving target can be obtained by using Canny edge detection algorithm. This method can not only obtain more complete contour information but also obtain part of the internal information of the moving target.
The laboratory dataset of three consecutive frames at frame 97, 98, and 99 were improved by the frame difference method before and after running time statistics, and the experimental results are shown in Figure 6:

In Figure 6, the right and the left images represent the running time of the traditional frame difference method and the improved frame difference method, respectively. Since the improved frame difference method performs Canny edge detection based on the moving target image obtained by the two-frame difference method and then performs “or” operation with the traditional TFDM, the increased running time is the Canny edge detection time and “or” operation time based on the two-frame difference method. Therefore, after the improvement, the time increase is relatively small, but the experimental accuracy has achieved a high improvement.

Table 2 shows the comparison of processing time between the traditional TFDM and the improved TFDM. When the processing time is increased little, the detection accuracy improved the frame difference.

Both intelligent room and laboratory datasets are due to the similarity between moving target and moving background in complex scenes, so the improved frame difference method detects more complete contour of the moving target than the traditional frame difference method, and the frame difference method is less affected by illumination changes.

5. Conclusion

Through the study of Canny edge algorithm, we can know that Canny edge detection can detect more complete edge information according to the small difference between the color value of the moving target and complex scene when the color of the moving target and the background is similar. Background model-based moving target detection is a very effective detection technology, which subtracts the current frame image from the existing background model image. If a pixel value in the difference image is greater than the threshold, it belongs to the moving target; otherwise, it belongs to the background region. From the above introduction and simulation test, it can be seen that background subtraction method is simple in operation, accurate in detection position, and fast in speed, and the results obtained after threshold operation directly give the location, size, shape, and other information of the target, which has a good application prospect. In this study, simulation experiments verify that the improved frame difference method is better than the traditional TFDM when the moving target color is similar to the background color in complex background.

Data Availability

The data used to support the findings of this study are available upon request to the author.

Conflicts of Interest

The author declares that he has no conflicts of interest.

Acknowledgments

This paper was supported by Outstanding talented person cultivation projects 2019 colleges and universities (Item no. gxyq201XXX2), the 2017 Anhui finance & trade.

References

[1] L. Najman, “Using mathematical morphology for document skew estimation,” in Proceedings of the Electronic Imaging 2004, International Society for Optics and Photonics, pp. 182–191, San Jose, CA, USA, January 2004.
[2] E. Jayabalan and A. Krishnan, “Detection and tracking of moving object in compressed videos,” Computer Networks and Information Technologies, vol. 142, pp. 39–43, 2011.
[3] X. Li, Q. Chen, and H. Chen, “Detection and tracking of moving object based on PTZ camera,” in Proceedings of the 2012 IEEE Fifth International Conference on Advanced Computational Intelligence (ICACI), pp. 493–497, Nanjing, China, October 2012.
[4] D. Parikh and C. L. Zitnick, “The role of image understanding in contour detection,” in Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition, pp. 622–629, Washington, NJ, USA, June 2012.
[5] X. Liu, H. Yao, R. Ji, P. Xu, and X. Sun, “Bidirectional-isomorphic manifold learning at image semantic understanding & representation,” Multimedia Tools and Applications, vol. 64, no. 1, pp. 53–76, 2012.
[6] T. Y. Lee, S. H. Nam, and J. S. Baek, “A motion estimation VLSI architecture for video image compression[J],” Aesthetic Surgery Journal, vol. 33, no. 3, pp. 456–457, 2013.
[7] H. Ke, H. Sun, and L. Gao, “A video image compression method based on visually salient features,” Journal of Digital Information Management, vol. 12, no. 5, pp. 333–339, 2014.
[8] C. Held, J. Krumm, P. Markel, and R. P. Schenke, “Intelligent video surveillance,” Computer, vol. 45, no. 3, pp. 83–84, 2012.
[9] J. Kim, D. Yeom, and Y. Joo, “Fast and robust algorithm of tracking multiple moving objects for intelligent video surveillance systems,” IEEE Transactions on Consumer Electronics, vol. 57, no. 3, pp. 1165–1170, 2011.
[10] D. Conte, P. Foggia, and M. Petretta, “Meeting the application requirements of intelligent video surveillance systems in MTD,” Pattern Recognition and Image Analysis, vol. 46, no. 4, pp. 180–187, 2011.
[11] N. Chen and F. Ding, “Flame object segmentation by an improved frame difference method,” in Proceedings of the 2012 Third International Conference on Digital Manufacturing and Automation (ICDMA), pp. 422–425, IEEE, Guilin, China, August 2012.
[12] L. L. Xia, X. Y. Pan, and X. D. Yang, “MTD method based on three-frame difference and improved by markowitz portfolio model,” Journal of Chinese Inertial Technology, vol. 22, no. 2, pp. 200–204, 2014.
[13] L. I. Qiu lin and H. E. Jia feng, “Vehicles detection based on three-frame-difference method and cross-entropy threshold method[J],” Computer Engineering, vol. 37, no. 4, pp. 172–174, 2011.
[14] B. Karasulu and S. Korukoglu, “Moving object detection and tracking by using annealed background subtraction method in videos: performance optimization,” Expert Systems with Applications, vol. 39, no. 1, pp. 33–43, 2012.
[15] H. Sang and C. Xu, “MTD based on background subtraction of block updates,” in Proceedings of the 2013 6th International
Conference on Intelligent Networks and Intelligent Systems (ICINIS), pp. 51–54, IEEE, Shenyang, China, November 2013.

[16] Y. Xin, J. Hou, L. Dong, and L. Ding, “A self-adaptive optical flow method for the moving object detection in the video sequences,” Optik, vol. 125, no. 19, pp. 5690–5694, 2014.

[17] E. Reinhard and T. Pouli, “Colour spaces for colour transfer,” Computational Color Imaging, Springer, Berlin, Germany, 2011.

[18] Y. H. Kim and H. Y. Jeong, “RGB color model based the fire detection algorithm in video sequences on wireless sensor network,” International Journal of Distributed Sensor Networks, vol. 4, no. 1, pp. 1–10, 2014.

[19] M. Song, D. Tao, C. Chen, J. Bu, and Y. Yang, “Color-to-gray based on chance of happening preservation,” Neurocomputing, vol. 119, no. 16, pp. 222–231, 2013.

[20] Y. Zhang, X. Wang, and B. Qu, “Three-frame difference algorithm research based on mathematical morphology,” Procedia Engineering, vol. 29, no. 4, pp. 2705–2709, 2012.

[21] J. Canny, “A computational approach to edge detection,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 8, no. 6, pp. 679–698, 1986.

[22] S. Q. Sun, K. Y. Liu, and S. Y. Liu, “MTD in complex environment of railway station,” Journal of Traffic and Transportation Engineering, vol. 13, no. 3, pp. 113–120, 2013.

[23] C. Chinrunrueng and C. H. Sequin, “Optimal adaptive K-means algorithm with dynamic adjustment of learning rate,” IEEE Transactions on Neural Networks, vol. 6, no. 1, pp. 157–169, 1995.