Detection of Abnormal Crowd Behavior Based on ViBE and Optical Flow Methods

Xinqian Zhao, Jian Chen[(✉)], Hongwei Lin, Yibo Zhao, and Changchuan Wei

School of Mechanical Engineering, Yangzhou University, Huayang Weststr. 196,
Yangzhou 225000, People's Republic of China
`chenjian.tud@hotmail.com`

Abstract. With the rapid development of computer vision and image processing technology, the detection of abnormal behavior in video has gradually attracted more and more scholars' attention. This paper proposes a fast method to detect abnormal behavior in surveillance videos. Firstly, the Visual Background Extraction (ViBE) method was used to extract the foreground object. Then, the image speed was separated according to the optical flow method. Finally, the Otsu method was used to binarize the image, the detection effect was evaluated by defining the density of moving pixels in a single frame of image relative to all pixels. In this paper, an experiment with the sudden scattering of concentrated crowds was conducted to simulate crowd abnormal behavior, the results show that the experimental effect is in line with expectations. Compared with other commonly used machine learning methods, the proposed method can be quickly used to detect crowd abnormal behavior in a surveillance video, the detection results can also be quickly evaluated, which has the advantages of low cost and convenience.

Keywords: Abnormal behavior detection · ViBE method · Foreground object extraction · Optical flow method · Otsu method

1 Introduction

Abnormal crowd behavior usually refers to the occurrence of illegal and criminal incidents in public places, the sudden dispersal of the crowd that had gathered together, or the sudden appearance of people who violate the rules of crowd movement in the regular movement of the crowd [1, 2]. With the continuous increase of urban population and the number of densely populated areas, there are more and more security risks, the occurrence probability of illegal and criminal behaviors also increases correspondingly, such as fighting and robbery, which greatly endangers public security. If these abnormal events can be detected in time, the harm of abnormal events to public security may be reduced.

The main work of abnormal behavior detection is to extract target features from a video and classify detection results. There are many ways to extract character features and classify the detection results. Xu et al. [3] used the Principle Component Analysis (PCA) for feature selection and the Support Vector Machine (SVM) for classification

of human behaviors. Nady et al. [4] employ space-time auto-correlation of gradients (STACOG) descriptor to extract spatio-temporal motion features from video sequence, the K-medoids clustering algorithm was used to partition the STACOG descriptors of training frames into a set of clusters. With considering the significance of the temporal comparison of motion states for detecting changes in crowds, a method based on the temporal context of motion was presented, to measure changes in the distribution of the physical characteristic descriptors of crowd motion [5]. Rodriguez et al. [6] proposed to leverage information on the global structure of the scene and to resolve all detection jointly, especially for some particularly challenging detection tasks, such as: heavy occlusions, high person densities and significant variation in people's appearance. In the paper of Wang et al. [7], pedestrian features were extracted by the multi-feature fusion method, the similar features in current frame of all candidate objects were matched with the characteristic information of pedestrians in the previous frame. This method has good robustness in complex traffic, but at the same time, the calculation is relatively large.

In order to improve the efficiency of video analysis, this paper proposes to extract the target foreground of the image firstly, then, carry out motion feature detection. The development of foreground object extraction technology has been relatively mature so far, most of them apply different processing algorithms, such as inter-frame difference method [8], the principle is simple, easy to be implemented, it also has good robustness, but is limited by the principle of inter-frame difference between adjacent frames, it has poor speed recognition effect on moving objects and is prone to produce cavitation phenomenon. Another example is the mixed Gaussian model method [9], although it has the adaptability of real-time update and better extraction effect compared with the former, it is difficult to obtain better results in the case of more noise in the image. In the year 2011, Barnich et al. [10] proposed Vibe method, experimental results show that compared with the mixed gaussian model method, the ViBe method has the advantages of higher speed and accuracy. Van Droogenbroeck [11] and Biao et al. [12] put forward their own improvement suggestions based on ViBe method respectively in 2012 and 2016, which made ViBe method more efficient and reduced the degree of "ghost phenomenon" in foreground images.

In addition, the optical flow method [13–17] can accurately identify the optical flow velocity of pixels. For the problem studied in this paper, the abnormal event detection of sudden crowd scattering or irregular behavior in public places, speed characteristic is the key to determine abnormal or not. Therefore, the concept of optical flow velocity separation was proposed in this paper, the optical flow method was used to filter the velocity, so as to filter out the low-speed information in the optical flow field and retain the high-speed information.

In the presented work, the abnormal behavior detection based on the improved ViBe algorithm as well as optical flow method was conducted. In order to detect the occurrence of abnormal events more easily, an abnormal event scoring mechanism was proposed, the mean score of high-speed part of optical flow field was calculated, the occurrence of abnormal events was detected according to the mean score.

2   Foreground Object Extraction

Background subtraction approach is one of the relatively common algorithms used in foreground detection. By which the foreground target was judged according to the changes of gray scale and other features. In other words, the current image and the background are differentiated. When the difference value is greater than a certain threshold, the target is considered to be detected [8]. The principle flow is shown in Fig. 1.
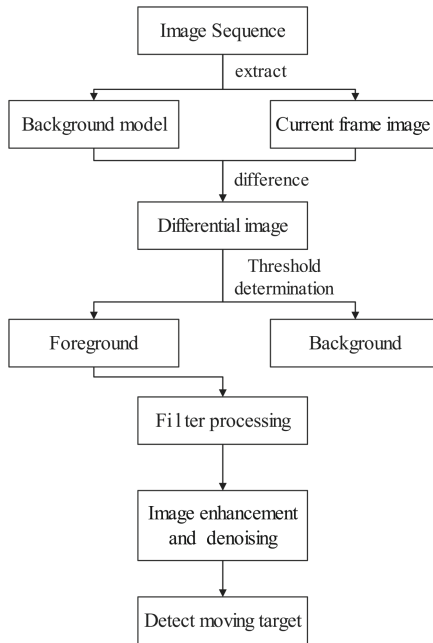

Fig. 1. Flow diagram of the background subtraction method

For the background subtraction method, the accuracy of the final foreground target detection was directly affected by the selected background model. Therefore, it is particularly important to select an appropriate algorithm to solve the background model. This paper mainly refers to the current advanced and efficient ViBe algorithm [10–12] for solving the problem.

2.1   ViBE Algorithm

ViBE algorithm uses neighborhood pixels to create a background model, and detects the foreground by comparing the background model with the current input pixel value, which can be subdivided into three steps:

Step1: Initialize the background model for each pixel in a single frame of image
It is assumed that the pixel values of each pixel and its neighbor pixel have similar distribution in the spatial domain. Based on this assumption, each pixel model can be

represented by its neighborhood pixels. In order to ensure that the background model conforms to statistical law, the range of neighborhood should be large enough. When the first frame of the image is entered, that is, when $T = 0$, the background model of pixels can be expressed as follows:

$$BK_M^0 = f^0\left(x^i, y^i\right) \big| \left(x^i, y^i\right) \in N_G(x, y) \tag{1}$$

in which, $N_G(x, y)$ represents the adjacent pixel values in the airspace, $f^0(x, y)$ represents the pixel value of the current point. In the process of N initialization, the possible times of pixel (x^i, y^i) in $N_G(x, y)$ being selected are $L = 1, 2, 3, …, N$.

Step2: Foreground object segmentation for the subsequent image sequence
When $t = k$, the background model of pixel point (x, y) is $BK_M^{k-1}(x, y)$, the pixel value is $f^k(x, y)$, then, the following formula is used to determine whether the pixel value is a foreground.

$$f^k(x, y) = \begin{cases} BK_M^{k-1}(x^r, y^r) > T & \text{foregound} \\ BK_M^{k-1}(x^r, y^r) \leq T & \text{background} \end{cases} \tag{2}$$

in this case, the superscript r is chosen at random, T is the preset threshold. If $f^k(x, y)$ meets the background N times, the pixels $f^k(x, y)$ can be regarded as the background, otherwise are the foreground.

Step3: Update method of background model
The update of Vibe algorithm is random in both time and space.

Randomness in time: Among N background models, one is randomly selected and set as image P_G, Table 1 represents the image of P_G position x and its eight pixels in the neighborhood. When a new frame of image P_t is obtained, if the pixel $P_t(x)$ corresponding to the x position in the image P_t is judged to be the background, P_G needs to be updated. This extraction process reflects the randomness of time.

Table 1. Selecting PG randomly from N frame background

$P_G(1)$	$P_G(2)$	$P_G(3)$
$P_G(4)$	$P_G(x)$	$P_G(5)$
$P_G(6)$	$P_G(7)$	$P_G(8)$

Randomness in space: A pixel $P_G(r)$ is randomly selected in the eight neighborhood of $P_G(x)$, and $P_G(r)$ is replaced by $P_t(x)$, which reflects the randomness of model update space.

2.2 Improved Vibe Algorithm

(1) Median filtering

In order to further improve the quality of the image obtained, the correlation method of morphological processing was adopted in this paper, to conduct secondary processing on the foreground target image extracted by Vibe algorithm, and filter out the relevant noise in the image.

As a signal processing method, median filtering [18] technology can effectively filter out noise. In this paper, the median filtering technology was used to replace the element points in the extracted foreground target image with the median of each element point in the specified field. The pixels in the foreground object in the image are closer to the real value, and the noise in the image was further eliminated. In two-dimensional images, the output formula of median filtering is as follows:

$$g(x, y) = \text{med}\{f(x - k, y - l)|k, l \in W\} \tag{3}$$

in which, $f(x, y)$ represents the original image, $g(x, y)$ represents the processed image, W is the used two-dimensional template.

(2) Corrosion expansion treatment

Supposing there are two images A and B, where A is the image to be processed and B is the corresponding template image of corroding or expanding. Usually, B is called the structural element. For the corrosion or expansion of A by B, it is manifested as the translation of B in the plane of A in any way. For corrosion, only when B is included in A, the black pixels in A are left, and the remaining pixels that do not meet the requirements are turned to white. For expansion, if the intersection of A and B is not empty, the corresponding white pixels in A are turned into black pixel, and the pixel that does not meet this condition are still white. This indicates that corrosion can reduce the boundary of the original image and eliminate the isolated noise points, while expansion can expand the boundary of the image and connect the isolated points.

By using the processing method in morphology [19], the target image is now corroded, the boundary points of the object are eliminated in a relevant way, so that the area of the processed image is reduced by one pixel along its periphery compared with that before corrosion. Then, the image is expanded, and all background pixels in contact with the object are integrated into the object, so that the object increases the area of the corresponding number of points.

The mathematical definition of corrosion is similar to expansion, A is corroded by B can be recorded as: $A \ominus B$, which is defined as:

$$A \ominus B = \left\{ z | (B)_z \cap A^c \neq \varnothing \right\} \tag{4}$$

In other words, A corroded by B is the set of origin positions of all structural elements, where the translation of B does not stack with the background of A, in which, $\varnothing$ is an empty set, and B is a structural element. Inflation is defined as a set operation. A is expanded by B, denoted as $A \oplus B$, and defined as:

$$A \oplus B = \left\{ z | (\hat{B})_z \cap A \neq \varnothing \right\} \tag{5}$$

(3) **Opening operation**

In this paper, the method of open operation was adopted to eliminate and separate the pixels representing noise in the original image without significantly changing the area of the original image. The morphological opening operation of A by B can be written as $A \circ B$, which is the result of expansion and corrosion of B after A is corroded by B.

Based on the original Vibe algorithm, the background difference method was used to obtain the extraction results of foreground objects, then, the median filter was used to remove the noise, and the corrosion and expansion algorithms in morphology were used to optimize the results. Finally, after a series of improved and optimized, the Vibe algorithm was obtained.

$$A \circ B = (A \ominus B) \oplus B \tag{6}$$

2.3 Foreground Extraction

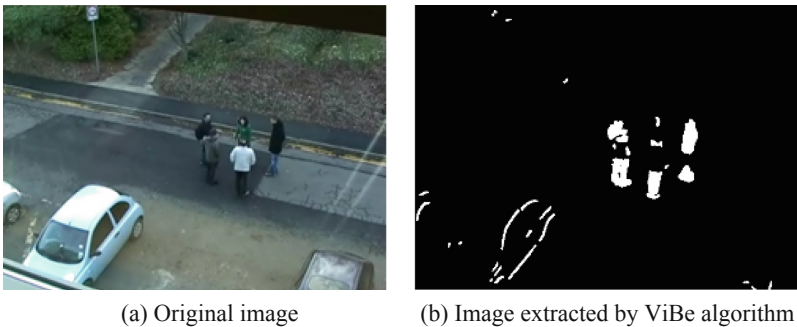The improved Vibe algorithm was used to extract the foreground target in the video, and the result is shown in Fig. 2.

(a) Original image   (b) Image extracted by ViBe algorithm

Fig. 2. Frame 412 in the video

3 Target Feature Extraction Based on Optical Flow Method

According to the movement information of the crowd in the video image, the occurrence of abnormal events can be judged. In order to highlight the influence of motion information on abnormal event detection, the speed separation of the processed image was carried out, the low-speed part of the image was filtered out while the high-speed part was retained. The high-speed part of the optical flow image reflects the high-speed moving region in the video, which is often the place where people are most interested, it is also the key region used to judge whether abnormal events occur or not. However, in the high-speed images extracted from optical flow images, high-speed regions are scattered, and some regions even have a few tiny points. These tiny high-speed regions do not contribute to abnormal behavior detection, and even affect the normal abnormal behavior detection results. In order to eliminate these effects, morphological processing of high-speed images can eliminate these small regions, while preserving large high-speed regions.

3.1   Optical Flow and Its Estimation Method

Optical flow is a motion mode, which is the movement of the target between the observer and the background. To be exact, optical flow is the projection of the target velocity to the wall in the three-dimensional space on the two-dimensional plane. Such projection on a two-dimensional plane can accurately reflect the movement information of the target.

Let $I\,(x,\,y,\,t)$ be the illumination intensity at the position of the image $(x,\,y)$ at the moment I, $w(x,\,y)$ and $v(x,\,y)$ are the optical flow components of the point in the horizontal and vertical directions respectively:

$$I(x, y, t) = I(x + u\Delta t, y + v\Delta t, t + \Delta t) \tag{7}$$

There are two unknowns u and v in the above equation, and the equation cannot be solved. To solve this equation, some new assumptions need to be added. Assuming that the velocity of the target sports field is continuous and smooth in the horizontal and vertical directions, we can perform first-order Taylor series expansion on it and obtain:

$$I(x, y, t) \approx I(x, y, t) + \frac{\partial I}{\partial x}\Delta x + \frac{\partial I}{\partial y}\Delta y + \frac{\partial I}{\partial t}\Delta t + \varepsilon \tag{8}$$

At this point, the second-order infinitesimal ε is ignored, the basic equation of optical flow method is as follows:

$$\frac{\partial I}{\partial x}\Delta x + \frac{\partial I}{\partial y}\Delta y + \frac{\partial I}{\partial t}\Delta t = 0 \tag{9}$$

The following set of equations can be obtained from pixels:

$$\begin{cases} I_{x_1}u + I_{y_1}v = -I_{t_1} \\ I_{x_2}u + I_{y_2}v = -I_{t_2} \\ \cdots \\ I_{x_n}u + I_{y_n}v = -I_{t_n} \end{cases} \tag{10}$$

The equation set has only three unknowns, but there are more than three equations, which means there is redundancy in the set. But it is clear that this equation set can be solved, and the system can also be expressed as:

$$\begin{bmatrix} I_{x_1} & I_{y_1} \\ I_{x_2} & I_{y_2} \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ I_{x_n} & I_{y_n} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} -I_{t_1} \\ -I_{t_2} \\ \cdot \\ \cdot \\ \cdot \\ -I_{t_n} \end{bmatrix} \tag{11}$$

3.2   Normalization of Optical Flow Images

The optical flow field calculated by the above optical flow method can be divided into horizontal direction and vertical direction. The optical flow field images represent the velocity components in horizontal and vertical directions respectively. However, in the study of abnormal behavior detection, we do not care about the direction of velocity, but only its magnitude. To facilitate the study, the two directions of the optical flow image can be fused, as shown in the formula:

$$I = \sqrt{u^2 + v^2} \tag{12}$$

In order to display the optical flow field intuitively, some additional display methods need to be defined. Since the gray level range of general images is [0,255], the size of optical flow is normalized to this range for display:

$$I' = \frac{I}{\text{Max}(I)} \cdot L \tag{13}$$

where, I' is the image after normalization, I is the image to be normalized, $Max(I)$ is the maximum number of pixels in the image. I' represents the gray level range, where the optical flow image of [0,255] can be normally displayed on the computer after it is normalized to the range of [0,1].

3.3   Separation of Optical Flow Velocity

Optical flow reflects the motion information of video images. In abnormal event detection, the occurrence of abnormal events can be easily judged by the motion information of images. According to the motion state of the crowd in the image, the human eye can easily distinguish the occurrence of abnormal events: when abnormal events occur, they are usually accompanied by large speed changes, which can be well reflected by the optical flow field image. By observing the optical flow field image of the abnormal event video, it can be found that when the abnormal event occurs, the gray image block in the image gradually becomes brighter due to the sharp increase of velocity, and the bright block spreads around. This paper determines the occurrence of abnormal events by analyzing the change characteristics of bright blocks in the optical flow field. In this paper, the concept of optical flow velocity separation was proposed: the optical flow image was divided into high-speed part and low-speed part, and the low-speed part of the optical flow field was filtered out and the high-speed part of the image was analyzed, which can reflect the occurrence of abnormal events more intuitively. The rules of optical flow image speed separation are shown in the formula:

$$\begin{cases} I_H = I_{x,y} \ I'_{x,y} \ > \ thresh \\ \ \ I_L = 0 \ \ \ I_{x,y} \ \leq \ thresh \end{cases} \tag{14}$$

here, the threshold is a constant value.

The high-speed part of optical flow image can directly reflect the occurrence state of abnormal events, and the selection of threshold value is very important, which is related

to the effect of separation. If the method of fixed threshold value is used, there is no way to apply to all scenarios, so this paper adopts dynamic threshold method. Generally, it is most appropriate to set the threshold as the median value of the optical flow field image. Considering the efficiency of the algorithm, a fast threshold calculation method is needed. Here, multiple of the maximum value of the image was taken as the threshold value, namely:

$$thresh = a\text{Max}(I) \tag{15}$$

in which, $Max(I)$ is the maximum number of pixels in the image. $thresh$ is a constant value, generally between 0.25 and 0.5. According to this strategy, the high-speed part and the low-speed part in the optical flow field can be clearly separated.

After separation, there will be many scattered gray areas in the high-speed part of the optical flow field image. These areas are not helpful for abnormal detection, but will affect the normal detection results. In order to eliminate the influence of these scattered areas, the corrosion operation in morphological filtering was used to filter out the separated high-speed images, many noise information affecting abnormal detection results were removed.
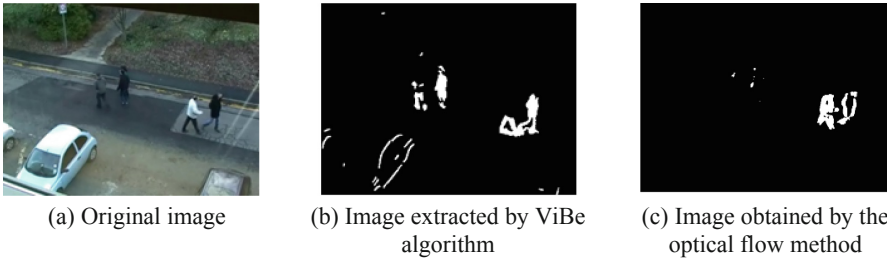

(a) Original image   (b) Image extracted by ViBe algorithm   (c) Image obtained by the optical flow method

Fig. 3. The processing result of frame 482 of the video, frame 482 was chosen because this is the middle segment where the abnormal behavior occurs

By comparing Fig. 3(b) and (c), it can be found that the image processed by optical flow method only contains the high-speed part, and the low-speed part of the foreground extraction part (such as the residual image of the car in the image) was completely filtered, which makes it easier to find the occurrence of abnormal events in the video.

4   Evaluation of Detection Effect

Before the occurrence of abnormal events, the distribution of abnormal event scores is gentle. According to the distribution characteristics of abnormal event score, this paper proposes to use the mean value of high-speed partial optical flow image to judge the changing process, so as to ensure the high accuracy of abnormal event detection.

As can be seen from the high-speed optical flow image, when the crowd is moving at a relatively low speed, the high-speed part of the image appears dark, and only a small part of the image has a non-zero pixel value. However, when the crowd is moving at high

speed, the high-speed part of the optical flow image appears bright, and the pixels in the large area are not zero anymore. The process from dimming to brightening of high-speed partial optical flow images can determine the occurrence of abnormal events. According to this characteristic, the mean value of high-speed partial optical flow image was used to judge the changing process.

$$S_M = \frac{\sum I_{x,y}}{N} \qquad (16)$$

In Eq. (16), S_M is the mean score of the image, also known as optical flow density, $\Sigma I_{x,y}$ represents the sum of all white pixels in the image, the Otsu method was used to binarize the image, N represents the number of pixels. This mean score can well reflect the overall motion speed of the video area, so as to distinguish whether abnormal events occur.

Otsu's method, named after its inventor Nobuyuki Otsu [20], is one of the most popular binarization algorithms. In computer vision and image processing, Otsu's method is used to automatically perform clustering-based image thresholding or reduction of a gray level image to a binary image.

In order to visualize the detection effect of optical flow method on abnormal behaviors, the changing trend of video score before and after optical flow method filtering was plotted in a certain video, as shown in Fig. 4. Black curve for the extraction of vision after this video grading, whole presents the score higher status, the score only said the white pixels of the change of situation, no practical significance, although in the beginning and end stage, had a tendency of lifting, but small change is not obvious, is impossible to judge whether there are abnormal event occurs. The blue curve is the score of the video processed by optical flow method. Compared with the black curve, the score of the blue curve decreases overall. Blue curve for the optical flow method after processing the segment ratings on video, compared to the black curve.
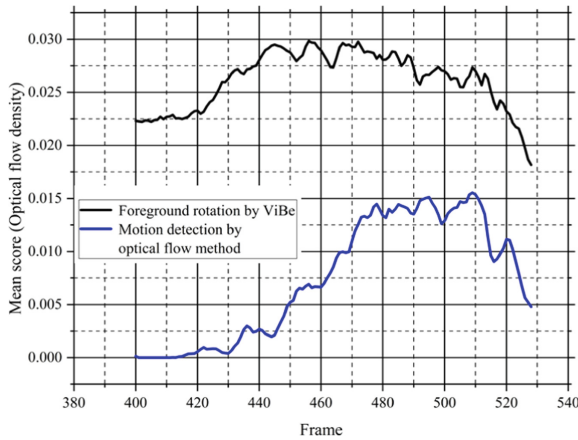

Fig. 4. Proportion of white pixels in a single frame image (Color figure online)

The abnormal behavior (high-speed part) is directly reflected as the white area in the binary image filtered by the optical flow method. Since the target with abnormal behavior moves faster, the white area should be larger in the optical flow image. However, because the optical flow image filters out the low-speed part and noise in the whole image, the proportion of the number of white pixels to the total pixels in the corresponding single frame image will be significantly reduced, but we can still judge whether abnormal events occur by the change trend of the proportion of white pixels after filtering (high-speed moving part), that is, the change of score.

It can be found that between frames 420 and 480, the score has a large continuous upward trend, which indicates that the number of high-speed moving pixels in the video begins to increase, and abnormal events occur at this time. Between frames 510 and 520, the score drops significantly, which indicates the end of abnormal events. This is consistent with the time point of abnormal behavior in the video, so it can be considered that the detection effect is good. Therefore, it can be concluded that the optical flow method can effectively identify the occurrence of abnormal events. When the abnormal time starts to occur, the scoring trend is large, but the overall score is in a low state.

5   Conclusions

In this paper, the ViBe background modeling method was used to extract the foreground, then, the optical flow method was used to extract the features of abnormal behaviors, so as to obtain a single frame binary image containing only the abnormal behaviors of the crowd. From these binary images, the abnormal behaviors of the crowd could be clearly identified.

The method proposed in this paper can be used to evaluate the detection effect quickly and accurately. However, for some other scenes, such as ghost probe, building collapse, explosion, etc., the detection result will still have deviation, especially under the situation of the video light changes strongly. The generality of this model in different scenarios will be further studied.

References

1. Zhang, Z.: A review of abnormal behavior detection techniques. Comput. Knowl. Technol. **16**(06), 199–201 (2020). (in Chinese)
2. Wu, X., Guo, H., Li, N.: Survey on the video-based abnormal event detection in crowd scenes. J. Electron. Measur. Instrum. **28**(06), 575–584 (2014). (in Chinese)
3. Wu, X., Ou, Y., Qian, H.: A detection system for human abnormal behavior. In: IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE (2005)

4. Nady, A., Atia, A., Abutabl, A.E.: Real-time abnormal event detection in crowded scenes. J. Theor. Appl. Inf. Technol. **96**(18), 6064–6075 (2018)
5. Chen, X., Lai, J.: Detecting abnormal crowd behaviors based on the div-curl characteristics of flow fields. Pattern Recogn. **88**, 342–355 (2019)
6. Rodriguez, M., Laptev, I., Sivic, J.: Density-aware person detection and tracking in crowds. In: IEEE International Conference on Computer Vision, ICCV 2011, Barcelona, Spain, 6–13 November 2011 (2011)
7. Wang, X., Song, H., Cui, H.: Pedestrian abnormal event detection based on mmulti-feature fusion in traffic video. J. Light Electronoptic **154**, 22–32 (2018)
8. Wang, Z., Zeng, H.: A new algorithm of moving vehicle detection and tracking based on combining frame difference method with optical flow technique. Comput. Appl. Softw. **29**(05), 117–120+129 (2012)
9. A, X.: Hybrid Gaussian Representation Model and its Application in Machine Industry Vision Inspection. Master's thesis, Kunming University of Science and Technology, Kunming, China (2021). (in Chinese)
10. Barnich, O., Van Droogenbroeck, M.: ViBE: A powerful random technique to estimate the background in video sequences. In: IEEE International Conference on Acoustics (2009)
11. Von Droogenbroeck, M., Paquot, O.: Background subtraction: experiments and improvements for ViBe. In: IEEE International Conference on Acoustics (2012)
12. Biao, W., Zheng, L.: Improvements on vibe algorithm for detecting foreground objects. In: International Conference on Computer Science and Network Technology (2017)
13. Gutchess, D., Trajkovics, M., Cohen-Solal, E.: A background model initialization algorithm for video surveillance. In: Proceedings IEEE International Conference on Computer Vision (2001)
14. Pless, R., John, L., Siebers, S.: Evaluation of local models of dynamic backgrounds. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2003)
15. Barron, J.L., Fleet, D.J., Beauchemin, S.S.: Performance of optical flow techniques. In: Proceedings Computer Vision and Pattern Recognition (1992)
16. Negahdaripour, S.: Revised definition of optical flow: integration of radiometric and geometric cues for dynamic. IEEE Trans. Pattern Anal. Mach. Intell. **20**(9), 961–979 (1998)
17. Beauchemin, S.S., Barron, J.L.: The computation of optical flow. ACM Comput. Surv. **27**(3), 433–466 (1995)
18. Liu, D.: Research of image denoising method based on wavelet transform. Mod. Electron. Technol. **36**(14), 93–95 (2013). (in Chinese)
19. Gao, X., Hu, Y., Du, W.: Application of decreasing noise of gray image by corrosion and expansion algorithm. J. Beijing Institute Graphic Commun. **22**(04), 63–65 (2014). (in Chinese)
20. Otsu, N.: A threshold selection method from gray-level histograms. Automatica **11**, 23–27 (1979)