

Research Article

A Novel Detection Framework for Detecting Abnormal Human Behavior

Chengfei Wu ¹ and Zixuan Cheng ²

¹Center of Information Development and Management, Soochow University, Suzhou 215006, China

²School of Electronic and Information Engineering, Changchun University, Changchun 130022, China

Correspondence should be addressed to Chengfei Wu; wuchengfei@suda.edu.cn

Received 23 October 2020; Revised 13 November 2020; Accepted 17 November 2020; Published 30 November 2020

Academic Editor: Yi-Zhang Jiang

Copyright © 2020 Chengfei Wu and Zixuan Cheng. 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.

Public safety issues have always been the focus of widespread concern of people from all walks of life. With the development of video detection technology, the detection of abnormal human behavior in videos has become the key to preventing public safety issues. Particularly, in student groups, the detection of abnormal human behavior is very important. Most existing abnormal human behavior detection algorithms are aimed at outdoor activity detection, and the indoor detection effects of these algorithms are not ideal. Students spend most of their time indoors, and modern classrooms are mostly equipped with monitoring equipment. This study focuses on the detection of abnormal behaviors of indoor humans and uses a new abnormal behavior detection framework to realize the detection of abnormal behaviors of indoor personnel. First, a background modeling method based on a Gaussian mixture model is used to segment the background image of each image frame in the video. Second, block processing is performed on the image after segmenting the background to obtain the space-time block of each frame of the image, and this block is used as the basic representation of the detection object. Third, the foreground image features of each space-time block are extracted. Fourth, fuzzy C-means clustering (FCM) is used to detect outliers in the data sample. The contribution of this paper is (1) the use of an abnormal human behavior detection framework that is effective indoors. Compared with the existing abnormal human behavior detection methods, the detection framework in this paper has a little difference in terms of its outdoor detection effects. (2) Compared with other detection methods, the detection framework used in this paper has a better detection effect for abnormal human behavior indoors, and the detection performance is greatly improved. (3) The detection framework used in this paper is easy to implement and has low time complexity. Through the experimental results obtained on public and manually created data sets, it can be demonstrated that the performance of the detection framework used in this paper is similar to those of the compared methods in outdoor detection scenarios. It has a strong advantage in terms of indoor detection. In summary, the proposed detection framework has a good practical application value.

1. Introduction

In recent years, with the frequent occurrences of abnormal group events such as fights, stampedes, riots, and demonstrations, video surveillance equipment has been widely used in public places such as railway stations, streets, campuses, and banks. Abnormal behavior detection using traditional video surveillance is mainly realized by manual methods. However, a long-term continuous observation often leads to staff fatigue, making them prone to missed inspections. The emergence of machine learning [1–3] can be exploited to

realize the automatic detection of abnormal human behaviors. Compared with traditional manual methods, this method can save manpower and reduce missed detections.

With the rapid development of machine learning and deep learning, intelligent video surveillance is becoming increasingly mature. Some well-known video surveillance systems [4–6] have been proposed one after another and have been put on the market. The detection of human abnormal behavior using video surveillance has become an important research topic in the field of computer vision in recent years, and it has attracted widespread attention.

Reference [7] summarized the recognition of different postures of the human body during activities. Reference [8] proposed a new technology and applied it to human behavior analysis and motion detection in video surveillance scenarios. Reference [9] provided a comprehensive overview of the current automated monitoring systems for abnormal behavior detection. Reference [10] proposed a method for estimating abnormal human behavior in different environments based on video surveillance. Reference [11] used a set of data codebooks constructed with two-level trees to calculate the similarity between space-time cubes at two scales and then used the linear discriminant analysis model to represent the topics of these scenes. If the test sample does not belong to these topics, it is considered abnormal. The hidden Markov model [12] is a typical inference model that can be applied to abnormal behavior detection in video scenes. Reference [13] used an independent hidden Markov model to construct sparse features in the training phase, and the model could adapt to scene changes by transforming into the most representative model. Reference [14] used a fully convolutional neural network (FCNN) for fast abnormal behavior detection. Reference [15] proposed the Appearance and Motion Deep Net (AMDN) and applied it to abnormal behavior detection in videos.

The above research is mainly based on the detection of abnormal human behavior in outdoor scenes. In campus scenes, students spend most of their time in indoor venues such as classrooms. The effect obtained when the above studies are applied to the detection of abnormal behavior by indoor personnel is not ideal. Moreover, although the detection method based on deep learning is convenient for feature extraction, its detection time cannot meet the needs of realistic applications. Abnormal behavior detection requires high efficiency in terms of detection time. In contrast, the time consumption of machine learning algorithms [16–25] has improved. Based on this fact, this paper uses a concise detection framework based on machine learning algorithms. First, a background modeling method based on a Gaussian mixture is used to segment the background image of each image frame in a video. Second, block processing is performed on the image after dividing the background to obtain the space-time block image of each picture. Third, the foreground image features of each space-time block image are extracted. Fourth, FCM is used to detect the data set. The main work of this paper is summarized as follows:

- (1) An abnormal human body behavior detection framework based on machine learning algorithms is used. The current popular deep learning algorithms are also applied to the detection and tracking of abnormal pedestrian behavior. However, when the crowd density is high or the color and apparent texture are similar, the detection may fail. On the other hand, when the probability of running behavior by a pedestrian is small, a large number of training samples cannot be collected. In this situation, deep learning algorithms are not suitable, and the supervised machine learning algorithm used in this article is more applicable.
- (2) The detection framework takes the space-time block as the basic representation of a detection object and extracts the foreground motion features of the space-time block image. FCM is used to detect outliers in the data sample and finally realize the detection of abnormal behavior in the crowd.
- (3) To demonstrate the effectiveness of the detection framework used, this paper conducts experiments on public data sets and newly constructed data sets. The experimental results show that the detection effect of the proposed detection framework in indoor environments is significantly better than those of other comparison methods, and the detection effect in outdoor environments is similar to those of the comparison methods. In summary, for places with more indoor scenes, the detection framework used in this paper has better application value.

The organizational structure of the remaining part is as follows: the second section is related knowledge, which mainly includes Abnormal Behavior Characteristics, Human Abnormal Behavior Detection Process, and Typical Public Data Set. The third section is Abnormal Human Behavior Detection Framework, which mainly includes Inspection Framework Description, Representation of Detection Objects, Feature Extraction, FCM, and Detection Framework Execution Steps. The fourth section is Experimental Results and Analysis. The fifth section is the conclusion.

2. Related Information

2.1. Abnormal Behavior Characteristics. In the process of human movement, normal behavior usually refers to the time and space states that exhibit a certain repetitiveness and regularity. These states include walking speed, walking posture, and spatial position. For example, walking and running at a constant speed are considered normal behaviors. However, there is no unified standard for the definition of abnormal behavior. Some scholars believe that all behaviors that do not match predefined normal behaviors are abnormal behaviors. Some scholars believe that behaviors that rarely occur or have short durations are abnormal behaviors. According to people's daily behaviors and living habits, it has been found that people's walking, running, and other behaviors have certain periodic laws. Therefore, for campus monitoring, we define the following characteristics as abnormal behaviors. These abnormal characteristics are as follows:

- (1) Abnormal walking trajectory: the walking motion trajectory of a person can indicate the purpose of his/her motion, and it is of great significance for safety monitoring systems to accurately detect this behavioral feature. For example, during a morning class, all students walk from the cafeteria or dormitory to the classroom, but some students walk backward. It is generally believed that people's destinations should be clear. Therefore, the definition of an abnormal walking trajectory in this article is as follows: within the camera's field of view, when

the walking path of a moving target over a certain period of time appears as a loop, stop-and-go, and return pattern, it is said that the target exhibits abnormal behavior.

- (2) Abnormal walking posture: walking upright is one of the signs of normal human appearance. In daily behavior, most people walk upright, but they also have other behaviors. For example, when an event such as stomach pain occurs, a person bends over and walks slowly. Therefore, we define an abnormal walking posture as follows: when the tested target bends over and suddenly squats during walking, we say that the target's behavior is abnormal.
- (3) Abnormal head rotation: according to people's cognitive habits, when a normal person walks, he/she usually looks ahead. Therefore, this paper provides the following definition of abnormal head rotation: in the course of a person's travel, when we detect frequent changes in the orientation of the person's head, it is considered that his/her behavior is abnormal.

2.2. Human Abnormal Behavior Detection Process. The detection process for abnormal human behavior is shown in Figure 1. First, spatiotemporal segmentation is performed on the video to extract features that can describe the characteristics of the target area. Then, during the training phase, normal events are modeled. In the testing phase, the abnormality of the test features is calculated for the normal event model that has been learned. In addition, it is judged whether a given behavior is abnormal according to the set abnormality threshold. The two steps of feature extraction and abnormal behavior detection by the model have a great impact on the detection effect with regard to abnormal behavior.

- (1) Pretreatment: during the preprocessing stage, the rough video data are prescreened through preprocessing operations. For example, normalization of video data and foreground extraction are performed to eliminate useless information in the image and enhance the detectability of valuable information. The data are simplified to the greatest extent possible, and their reliability is enhanced for subsequent processing and analysis.
- (2) Feature extraction: the features used to describe the characteristics of crowd movement are mainly divided into two categories: trajectory segmentation features and local temporal and spatial features. Most studies in the literature use local spatiotemporal features to extract other features, and the extracted motion information mainly includes gradient features and optical flow motion features.
- (3) Anomaly detection model: during the training process, the normal event modeling and testing phases make abnormality judgments on the test features based on the model. The detection model commonly used at this stage is the abnormal behavior detection

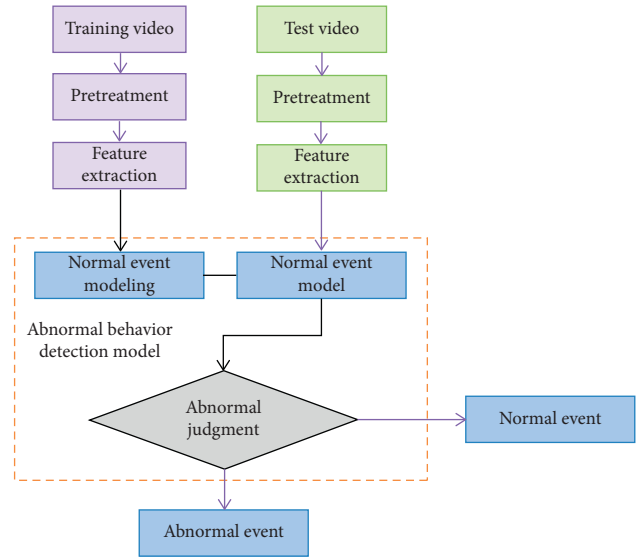


FIGURE 1: Abnormal human behavior detection process.

model based on reconstruction. This type of method mainly uses normal behavior feature training to obtain a model for normal behavior. In the normal model, the reconstruction error corresponding to anomalies is large, and the reconstruction error corresponding to normal features is small. Therefore, a feature is judged based on a comparison between the test behavior characteristics and the reconstruction error of the normal behavior model with regard to the abnormal threshold. This type of method can be subdivided into two categories: reconstruction models based on sparse coding and reconstruction models based on deep learning.

2.3. Typical Public Data Set. Recently, the number of publicly available data sets for abnormal behavior detection has increased. The typical public data sets are introduced in Table 1.

3. Abnormal Human Behavior Detection Framework

3.1. Inspection Framework Description. The detection framework used in this paper is shown in Figure 2. As shown in the figure, the video data are preprocessed first. In this stage, the mixed Gaussian background modeling method [30] is mainly used to eliminate the background of each frame of a given image. In the second stage, the foreground image features of each space-time block image are extracted. In the third stage, FCM is used to cluster the extracted feature data sets. In the fourth stage, outliers are obtained according to the clustering results to determine whether the image frame is an abnormal frame.

3.2. Representation of Detection Objects. Taking the space-time block as the detection object, the video sequence frame is first divided into $W * H$ spatial blocks, and the size of each

TABLE 1: Introduction of data sets.

Data set	Details
UCSD [26]	The database contains two video sets, Ped1 and Ped2. The Ped1 video set contains a total of 34 training video sequences and 36 test video sequences, each with a resolution of $238 * 158$. Each video sequence contains 200 frames of images. The Ped2 data set contains a total of 16 training videos and 12 test videos, each with a resolution of $320 * 240$.
UMN [27]	The database contains 3 crowded scenes; the total numbers of frames in the three scenes are 1450, 4415, and 2145, and the resolution of each is $320 * 240$.
Subway [28]	The database contains the subway-entrance and subway-exit subsets. The video lasts 1 hour and includes abnormal behaviors such as retrograde and fare evasion at the ticket gate.
Avenue [29]	The data set has 16 training videos and 21 test videos, each with a resolution of $640 * 360$. Abnormal behaviors include throwing paper everywhere and running.

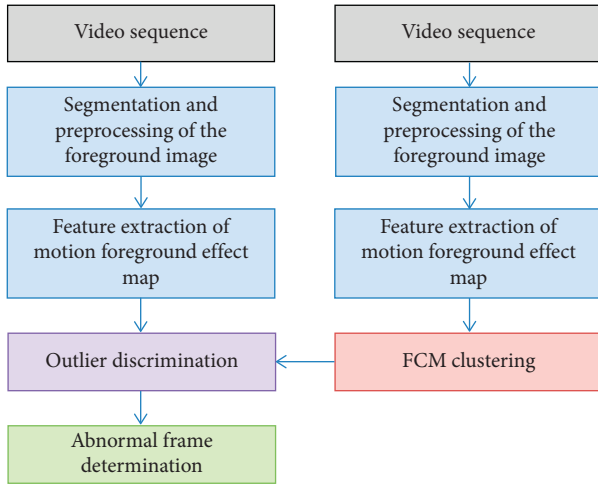


FIGURE 2: Flowchart of the detection framework used.

spatial block is $L * L$. Because the focus is on analyzing the various parts of the movements of pedestrians, such as the impact of human legs running, the criterion for determining L is as follows: a single spatial block can be expressed as a certain part of a moving target, and a space-time block is composed of consecutive multiple frame spatial blocks at the same position. The basic representation method of using the space-time block as the detection object is not done to directly extract motion information from the current space-time block but rather to further analyze the motion effect of a space-time block with rich foreground information on the surrounding space-time block.

3.3. Feature Extraction. The description of the features of a space-time block can be obtained by using histogram statistics of low-level visual information such as the gray gradient, optical flow, and texture of the space-time block [31]. Pedestrians with the “panic running” behavior have the characteristics of fast movement and great kinetic energy. Their effect on the surrounding space environment is more significant than that of a pedestrian walking normally. The speed of pedestrian movement, the range of influence of pedestrian movement, and the distance between pedestrians and space are the main factors that determine the effects of pedestrians on space. The feature extraction process of running foreground images based on space-time blocks is shown in Figure 3.

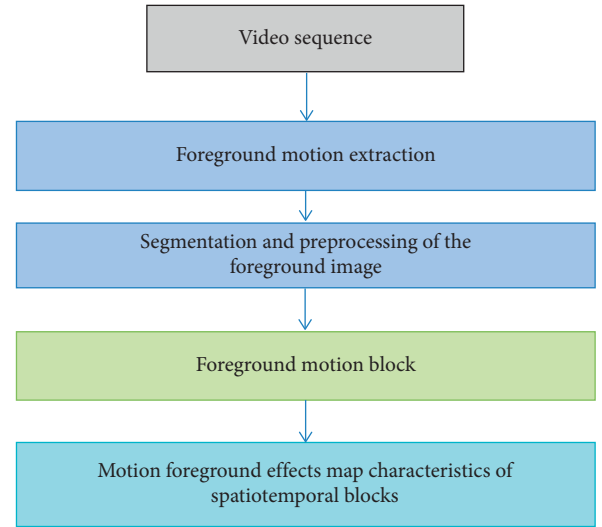


FIGURE 3: Feature extraction flowchart.

As shown in Figure 3, first, the foreground motion image is extracted from the video frame sequence by the adaptive Gaussian mixture model, and the spatial block is obtained. The spatial block is combined with the foreground image and then preprocessed to obtain the foreground motion block. Then, the motion representation of the foreground motion block is obtained according to the dense optical flow of the video frame. Finally, the effect weights vector for all foreground motion blocks is calculated for each spatial block. The effect weights vectors of consecutive multiple frames of space blocks are averaged to obtain a feature description of a given space-time block. The detailed description of each step is as follows.

3.3.1. Foreground Motion Block. A foreground motion block can effectively represent the movement information of pedestrians. Set the j -th space block to B_j , $1 \leq j \leq W \times H$. The foreground motion block refers to the space block where the premovement scenic spot appears. The foreground information contained in some spatial blocks in the foreground motion block may be information such as noise, which cannot correctly characterize the motion behavior of an object. To determine the existence of such motion blocks, this paper preprocesses all spatial blocks. Suppose b_j is the number of front spots in a block; only when equation (1) is satisfied, can it be retained as a foreground motion block:

$$C_i = B_j, \quad \text{if } \frac{b_j}{L \times L} > \lambda. \quad (1)$$

C_i ($1 \leq i \leq S$) indicates that an image block B_j can be used as the j -th foreground motion block when the above conditions are met. λ ($0.1 \leq \lambda \leq 0.3$) is the comparison threshold for the previous scenic spot. The above operation is the preprocessing of the foreground image, and foreground motion blocks can be extracted from the preprocessed foreground image.

3.3.2. Foreground Motion Effects Map. The optical flow vector of all pixels in a preprocessed spatial block is extracted, and the average value is used as the optical flow vector of the current block. This is used as the motion representation of the foreground motion block, as follows:

$$c_i = \frac{1}{J} \sum_j g_i^j. \quad (2)$$

c_i represents the optical flow vector of the i -th foreground motion block. J is the number of all pixels in the foreground block. g_i^j represents the optical flow vector of the j -th pixel in the i -th foreground motion block. $\|c_i\|$ and Γc_i represent the magnitude and direction, respectively, of the optical flow of the i -th foreground motion block.

3.3.3. Feature Extraction. According to the obtained sub-block effect measurement, the subblock effect of normal behavior and the subblock effect of abnormal behavior can be effectively distinguished. Therefore, this paper uses a foreground motion effects map feature to characterize the motion effects of neighboring foreground blocks on space-time blocks.

For a spatial block B_j and a foreground motion block C_i , to measure whether the foreground motion block has an effect on the surrounding space block, two index variables are defined as follows:

$$\begin{aligned} \varphi_{ij}^d &= \begin{cases} 1, & \text{if } \text{dist}(i, j) < \delta_d, \\ 0, & \text{else,} \end{cases} \\ \varphi_{ij}^\theta &= \begin{cases} 1, & \text{if } -\frac{\pi}{2} < \theta_{ij} < \frac{\pi}{2}, \\ 0, & \text{else.} \end{cases} \end{aligned} \quad (3)$$

$\text{dist}(i, j)$ represents the Euclidean distance between the foreground motion block C_i and the space block B_j . δ_d is the distance threshold. θ_{ij} is the angle between the vector from the foreground block C_i to the space block B_j and the optical flow of C_i . $(-\pi/2, \pi/2)$ represents the field of view when the foreground motion block is in motion. These two index variables measure whether the space block B_j is in the influence range of the foreground block C_i .

The effect weight of the foreground motion block C_i on the spatial block B_j is defined as

$$w_j = \begin{cases} \varphi_{ij}^d \varphi_{ij}^\theta \exp\left(-\frac{\text{dist}(i, j)}{\|c_i\|}\right), & \text{if } C_i \neq B_j, \\ \varphi_{ij}^d \varphi_{ij}^\theta \exp\left(-\frac{1}{\|c_i\|}\right), & \text{else.} \end{cases} \quad (4)$$

When the space block B_j is in the influence range of the foreground block C_i , the weight of the effect of B_j received by C_i is inversely proportional to the distance between the two and directly proportional to the magnitude of the optical flow of C_i . C_i represents a pedestrian. When the pedestrian runs vigorously, the weight w_{ij} increases with the magnitude of the optical flow. Multiple foreground blocks with different motion directions have differently weighted weights on B_j , so all effect weights of B_j must be counted to form an effective feature representation. To increase the degree of discrimination of the extracted features for the purpose of calculating the efficiency of the model, the moving direction Γc_i of the foreground motion block C_i is quantized as follows:

$$q(\Gamma c_i) = k_i, \quad \text{if } (k_i - 1) \times \frac{2\pi}{p} < \Gamma c_i \leq k_i \times \frac{2\pi}{p}, \quad (5)$$

where $k_i \in \{1, 2, \dots, p\}$ and p is the total number of quantization direction intervals. k_i represents the quantization direction index value of the optical flow of the i -th foreground motion block. The histogram statistics of the effect weights generated by the foreground block for the spatial block B_j are based on the quantization direction of the optical flow of the foreground motion block:

$$f_j(k_i) = \sum w_{ij}, \quad \text{if } q(\Gamma c_i) = k_i. \quad (6)$$

Equations (4) and (5) are incorporated into equation (6) and the histogram statistics f_j are calculated. To obtain the long-term statistics of the foreground motion effect and make the extracted features highly discriminative, the spatial blocks $\{B_j, B_{j+1}, \dots, B_{j+m-1}\}$ with m consecutive frames are taken as a space-time partition \bar{B}_j . According to equation (6), the characteristic description of each space block in \bar{B}_j can be calculated. The corresponding weight vector of a given space-time block can be expressed as $\{h_j, h_{j+1}, \dots, h_{j+m-1}\}$, $1 \leq j \leq W \times H$. The feature description of the m -frame space block takes the mean value as the running foreground effect feature $\tilde{\mathbf{h}}_j$ ($1 \leq j \leq W \times H$) of the space-time block \bar{B}_j . The calculation formula is as follows:

$$\tilde{\mathbf{h}}_j = \frac{1}{m} \sum_{i=0}^{m-1} \mathbf{h}_{j+i}, \quad \tilde{\mathbf{h}}_j \in R^p. \quad (7)$$

3.4. FCM. FCM is a classic clustering algorithm based on distance. Given a data set $X = \{x_1, x_2, \dots, x_n\}$ consisting of n p -dimensional samples, the data are grouped into c ($c \in [2, n]$) categories. The center of each category is v_c . $V = \{v_1, v_2, \dots, v_c\}$ is the cluster center matrix. $U = \{u_{ki}\} \in R^{n \times c}$ is the membership matrix, which satisfies $\sum_{k=1}^c u_{ki} = 1$. The objective function of FCM is as follows:

$$J = \sum_{k=1}^c \sum_{i=1}^n u_{ki}^m \|x_i - v_k\|^2, \quad (8)$$

where m is the fuzzy coefficient, $m=2$. Minimizing J can complete the division of the sample set. Using the Lagrange multiplier method to iteratively update J , the calculation formulas for the membership and clustering centers are obtained as follows:

$$u_{ki} = \frac{\|x_i - v_k\|^{-2/(m-1)}}{\sum_{j=1}^c \|x_i - v_j\|^{-2/(m-1)}}, \quad (9)$$

$$v_k = \frac{\sum_{i=1}^n u_{ki}^m x_i}{\sum_{i=1}^n u_{ki}^m}. \quad (10)$$

Equations (9) and (10) are iterated repeatedly until the FCM algorithm converges. FCM is used to distinguish the outliers among the characteristic sample points of the space-time block to be detected. For the video frame sequence $\{F_1, F_2, \dots, F_T\}$, the feature sample set $\{\tilde{\mathbf{h}}_{(i,j)}^1, \tilde{\mathbf{h}}_{(i,j)}^2, \dots, \tilde{\mathbf{h}}_{(i,j)}^{T-m}\}$ of all space-time blocks located in (i, j) , $i \in [1, W]$, $j \in [1, H]$, can be extracted. FCM is used to cluster the sample set located at (i, j) , the optimal number of clusters for the optimization algorithm is K , and the best performance is obtained when the cluster center is set as $\{v_{(i,j)}^l\}$, $K \leq l \leq 1$.

For the image frame T to be detected, the space-time block response feature map $\mathbf{e}_{(i,j)}$ located at (i, j) must be obtained. Through equation (11), the distance between it and the cluster center set $\{v_{(i,j)}^l\}$, $K \leq l \leq 1$ can be calculated:

$$d(i, j) = \min_l \|e_{(i,j)} - v_{(i,j)}^l\|, \quad K \leq l \leq 1. \quad (11)$$

The threshold ε_1 is set when $d(i, j) > \varepsilon_1$, $\mathbf{e}_{(i,j)}$ is an outlier. Then, the empty time block at (i, j) is marked as abnormal. When the number of abnormal points in a frame of an image is greater than the total number of foreground blocks, the frame of the image is considered to be an abnormal frame.

3.5. Detection Framework Execution Steps. The execution steps of the detection framework used in this paper are as follows:

- (1) A video sequence $\{F_1, F_2, \dots, F_T\}$ is given. The feature set $\{\tilde{\mathbf{h}}_{(i,j)}^1, \tilde{\mathbf{h}}_{(i,j)}^2, \dots, \tilde{\mathbf{h}}_{(i,j)}^{T-m}\}$, $i \in [1, W]$, $j \in [1, H]$, is obtained by extracting the empty block at position (i, j) .
- (2) The FCM algorithm is used to cluster the feature sample set to obtain the cluster center set $\{v_{(i,j)}^l\}$, $K \leq l \leq 1$.
- (3) The foreground motion image feature $\mathbf{e}_{(i,j)}$ of the space-time block at (i, j) in frame T to be detected is calculated. The distance $d(i, j)$ from the cluster center set to $\mathbf{e}_{(i,j)}$ is then calculated.

- (4) When $d(i, j) > \varepsilon_1$, $\mathbf{e}_{(i,j)}$ is regarded as an outlier.
- (5) The number of outliers in the frame is counted. If the number is greater than the total number of foreground blocks, the frame is determined to be an abnormal frame.

The flowchart of the detection framework used is shown in Figure 4.

4. Experimental Results and Analysis

4.1. Experiment-Related Instructions. To demonstrate the feasibility and superiority of the proposed detection model, the public UMN data set and manually generated data set are used in the experiment. The UMN data set is a public data set used for anomaly detection research that mainly includes sudden movements, crowd appearances and disappearances, and aggregations. The data set has a total of 7739 frames of images, each image size is 320×240 , and the resolution is low. There are three types of scenes. The first type of scene contains 2 sets of complete anomalies. The second category contains 6 complete abnormal data points. The third category contains 3 sets of complete abnormal data points. The comparison methods are taken from [32], [33], and [34]. The settings of the parameters for each comparison method are consistent with those in the respective references. The evaluation indicators are the detection accuracy [35] and Area under the ROC curve (AUC) [36].

4.2. UMN Data Set Experiment. The first 180 frames of 9 video clips in three scenes are extracted as the training set, and the rest are extracted as the test set. The detection accuracy and AUC values of each method under different scenarios are shown in Table 2.

The experimental data in Table 2 show that different detection methods work in different scenarios, and the performance of each is quite different. The common point is that all algorithms have their best detection performances on Scene 1. From the average results of the three scenes detected by each method, it can be inferred that the detection accuracy and AUC obtained by the method in this paper are both greater than those of other comparison methods. The detection accuracy obtained by the method in this paper is improved by 5.2% over that of SF, 1.3% over that of STC, and 2.0% over that of MA. The AUC is increased by 2.0% compared with that of SF, 1.9% compared with that of STC, and 1.5% compared with that of MA. Through the comparison with the two above indicators, it is demonstrated that the method in this paper performs better than other methods in detecting abnormal human behavior on the UMN data set.

To explore whether the selection of the clustering algorithm in the detection framework is appropriate, we replace the FCM at the end of the framework with the classic K -means clustering algorithm [37]. The impacts of these two different clustering algorithms on the detection performance of this detection framework are compared. Two different clustering algorithms are applied to the detection framework

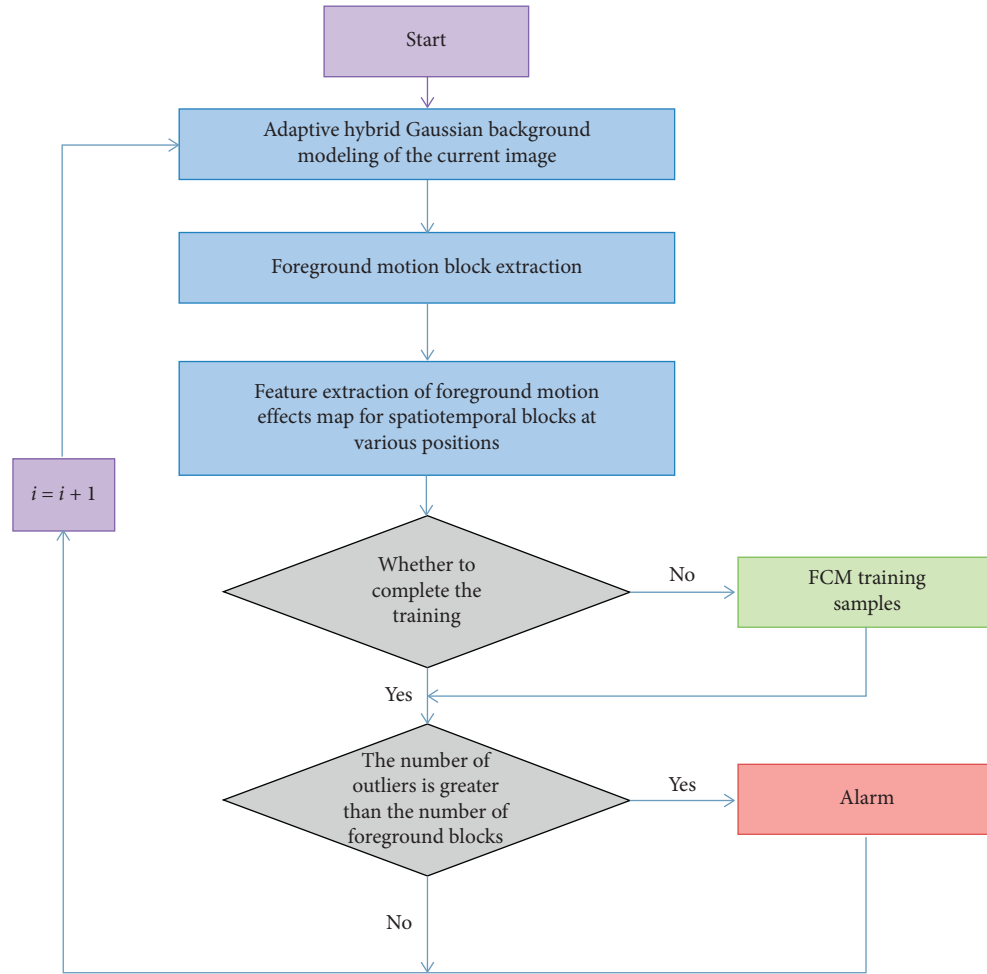


FIGURE 4: Detection flowchart.

TABLE 2: Comparison of experimental results obtained by different detection methods on the UMN data set.

Method	Scenes	Accuracy	AUC
Reference [32]	Scene 1	0.8212	0.9878
	Scene 2	0.8145	0.9653
	Scene 3	0.8187	0.9692
	Mean	0.8181	0.9741
Reference [33]	Scene 1	0.8568	0.9908
	Scene 2	0.8476	0.9697
	Scene 3	0.8442	0.9639
	Mean	0.8495	0.9748
Reference [34]	Scene 1	0.8678	0.9901
	Scene 2	0.8422	0.9624
	Scene 3	0.8208	0.9830
	Mean	0.8436	0.9785
Our method	Scene 1	0.8717	0.9974
	Scene 2	0.8312	0.9853
	Scene 3	0.8780	0.9981
	Mean	0.8603	0.9936

to detect human abnormalities in the UMN data set. The experimental results are shown in Table 3.

TABLE 3: Performance comparison when different clustering algorithms are applied to the detection framework.

Method	Scenes	Accuracy	AUC
<i>K</i> -means	Scene 1	0.8689	0.9895
	Scene 2	0.8276	0.9757
	Scene 3	0.8701	0.9923
	Mean	0.8555	0.9858
FCM	Scene 1	0.8717	0.9974
	Scene 2	0.8312	0.9853
	Scene 3	0.8780	0.9981
	Mean	0.8603	0.9936

The experimental results show that the detection effect obtained by using FCM in the detection framework is better than that obtained by using *K*-means. This is because FCM is not sensitive to noise in the data, so the detection result is not easily affected by noise either, thereby improving the detection accuracy. This is the reason why this paper chooses FCM for use in the detection framework.

4.3. *Self-Made Data Set Experiment.* To further verify the robustness of the algorithm in this paper, we download

TABLE 4: Comparison of experimental results obtained by different detection methods on self-made data sets.

Method	Scenes	Accuracy	AUC
Reference [32]	Indoor	0.7993	0.9075
	Outdoor	0.8078	0.9288
Reference [33]	Indoor	0.8097	0.9323
	Outdoor	0.7989	0.9224
Reference [34]	Indoor	0.8001	0.9230
	Outdoor	0.8212	0.9456
Our method	Indoor	0.8214	0.9712
	Outdoor	0.8115	0.9697

multiple normal and abnormal group behavior videos under different scenarios from the Internet. The downloaded videos are randomly combined to construct 35 video data points. Each video includes anomalous group incidents such as indoor pushing, trampling, outdoor robberies, and terrorist attacks. Among the 35 videos, 25 videos are used as training data sets, and 10 videos are used as test data sets. The experimental results on the manually created data set are shown in Table 4.

In indoor scenes, the detection method used on the manually created data set has the best detection accuracy and AUC value. This fully demonstrates the superiority of the proposed detection method when used for indoor scene detection. In outdoor scenes, the MA algorithm has the best detection performance, but the detection effect of the proposed method is close behind, and the performance gap is not large. Moreover, the detection performance of this method in outdoor scenes is better than those of SF and STC. This shows that the indoor detection performance of this method is significantly improved, the outdoor detection performance is not poor, and the method can fully meet the needs for actual application scenarios.

5. Conclusion

Abnormal behavior detection in videos is a research hotspot in the field of smart security. According to different surveillance video scenes and surveillance objects, abnormal behaviors have different definitions. The detection of abnormal behaviors among a crowd in public places has high research value. This research focuses on the analysis and detection of the abnormal behaviors of crowds in videos, focusing on "panic running" and other behaviors. For these abnormal behaviors, a detection framework based on machine learning algorithms is used. First, a space-time block is used as the basic representation of the detection object. Second, a background modeling algorithm is used to analyze the effect of the foreground motion block on the surrounding space blocks and calculate the characteristics of the foreground motion effects map of the space-time block. Finally, the FCM algorithm is used for clustering, training, and outlier detection. When the number of detected outliers is greater than the number of foreground blocks, it is determined that the detected frame is an abnormal frame. Compared with existing detection algorithms, the experimental results show that the method used has better

detection performance. However, there are many parameters that need to be manually determined in the proposed detection framework, and this is a drawback of this research. Improving upon this drawback is also a future research direction.

Data Availability

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

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This study was supported by the Scientific Research Project of Jilin Education Department.

References

- [1] N. Hong, H. Park, and Y. Rhee, "Machine learning applications in endocrinology and metabolism research: an overview," *Endocrinology and Metabolism*, vol. 35, no. 1, pp. 71–84, 2020.
- [2] William Trung Le, F. Maleki, F. P. Romero, R. Forghani, and S. Kadoury, "Overview of machine learning: part 2: deep learning for medical image analysis," *Neuroimaging Clinics of North America*, vol. 40, no. 3, pp. 417–431, 2020.
- [3] H. Andy, R. Terry Van Vleet, R. R. Gupta, M. J. Liguori, and M. S. Rao, "An overview of machine learning and big data for drug toxicity evaluation," *Chemical Research in Toxicology*, vol. 33, no. 1, pp. 20–37, 2020.
- [4] I. Haritaoglu, D. Harwood, and L. S. Davis, "W/sup 4/: real-time surveillance of people and their activities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 809–830, 2000.
- [5] Y.-li Tian, L. Brown, A. Hampapur, M. Lu, A. Senior, and C. Shu, "IBM smart surveillance system (S3): event based video surveillance system with an open and extensible framework," in *Proceedings of the Advanced Video and Signal Based Surveillance Conference*, pp. 318–323, IEEE, Como, Italy, 2005.
- [6] K. H. Tan, "Vs-star: a visual interpretation system for visual surveillance," *Pattern Recognition Letters*, vol. 31, no. 14, pp. 2265–2285, 2010.
- [7] A. Bux, P. Angelov, and Z. Habib, "Vision based human activity recognition: a review," *Multimedia Tools and Applications*, vol. 79, 2017.
- [8] M. S. K. Mishra and K. S. Bhagat, "A survey on human motion detection and surveillance," *International Journal of Advanced Research in Electronics and Communication Engineering*, vol. 10, 2015.
- [9] M. Valera and S. A. Velastin, "Intelligent distributed surveillance systems: a review," *Vision, Image and Signal Processing*, vol. 152, no. 2, pp. 192–204, 2005.
- [10] P. Popoola Oluwatoyin, "Wang kejun. Video-based abnormal human behavior recognition review," *IEEE Transactions on Systems Man & Cybernetics Part C*, vol. 42, no. 6, pp. 865–878, 2012.

- [11] B. Wang, M. Ye, X. Li, and F. Zhao, "Abnormal crowd behavior detection using size-adapted spatio-temporal features," *International Journal of Control, Automation and Systems*, vol. 9, no. 5, pp. 905–912, 2011.
- [12] L. R. abiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Readings in Speech Recognition*, vol. 77, no. 2, pp. 267–296, 1990.
- [13] A. E. Gunduz, C. Ongun, T. T. Temizel et al., "Density aware anomaly detection in crowded scenes," *Iet Computer Vision*, vol. 10, no. 5, pp. 374–381, 2016.
- [14] M. Sabokrou, M. Fayyaz, M. Fathy et al., "Deep-anomaly: fully convolutional neural network for fast anomaly detection in crowded scenes," *Computer Vision and Image Understanding*, vol. 172, 2017.
- [15] D. Xu, E. Ricci, Y. Yan et al., "Learning deep representations of appearance and motion for anomalous event detection," in *Proceedings of the British Machine Vision Conference*, pp. 8.1–8.12, London, UK, 2015.
- [16] P. Qian, Y. Chen, J.-W. Kuo et al., "mDixon-based synthetic CT generation for PET attenuation correction on abdomen and pelvis jointly using transfer fuzzy clustering and active learning-based classification," *IEEE Transactions on Medical Imaging*, vol. 39, no. 4, pp. 819–832, 2020.
- [17] Y. Jiang, K. Zhao, K. Xia et al., "A novel distributed multitask fuzzy clustering algorithm for automatic MR brain image segmentation," *Journal of Medical Systems*, vol. 43, no. 5, p. 118, 2019.
- [18] P. Qian, C. Xi, M. Xu et al., "SSC-EKE: semi-supervised classification with extensive knowledge exploitation," *Information Sciences*, vol. 422, pp. 51–76, 2018.
- [19] Y. Jiang, Z. Deng, F.-L. Chung et al., "Recognition of epileptic EEG signals using a novel multiview TSK fuzzy system," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 1, pp. 3–20, 2017.
- [20] P. Qian, Y. Jiang, S. Wang et al., "Affinity and penalty jointly constrained spectral clustering with all-compatibility, flexibility, and robustness," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 5, pp. 1123–1138, 2017.
- [21] P. Qian, Y. Jiang, Z. Deng et al., "Cluster prototypes and fuzzy memberships jointly leveraged cross-domain maximum entropy clustering," *IEEE Transactions on Cybernetics*, vol. 46, no. 1, pp. 181–193, 2016.
- [22] K.-J. Xia, X. Zhong, Li Zhang, and J. Wang, "Optimization of diagnosis and treatment of chronic diseases based on association analysis under the background of regional integration," *Journal Medical Systems*, vol. 43, no. 3, 2019.
- [23] P. Qian, S. Sun, Y. Jiang et al., "Cross-domain, soft-partition clustering with diversity measure and knowledge reference," *Pattern Recognition*, vol. 50, pp. 155–177, 2016.
- [24] Y. Jiang, D. Wu, Z. Deng et al., "Seizure classification from EEG Signals using transfer learning, semi-supervised learning and TSK fuzzy system," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 12, pp. 2270–2284, 2017.
- [25] K. Xia, H. Yin, P. Qian, Y. Jiang, and S. Wang, "Liver semantic segmentation algorithm based on improved deep adversarial networks in combination of weighted loss function on abdominal CT images," *IEEE Access*, vol. 7, pp. 96349–96358, 2019.
- [26] T. Wang, M. Qiao, Z. Lin et al., "Generative neural networks for anomaly detection in crowded scenes," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 5, pp. 1390–1399, 2019.
- [27] N. Bird, S. Atev, N. Caramelli et al., "Real time, online detection of abandoned objects in public areas," in *Proceedings of the 2006 IEEE International Conference on Robotics and Automation*, pp. 3775–3780, Orlando, FL, USA, 2006.
- [28] A. Adam, E. Rivlin, I. Shimshoni, and D. Reinitz, "Robust real-time unusual event detection using multiple fixed-location monitors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 3, pp. 555–560, 2008.
- [29] X. Li, W. Li, B. Liu, and N. Yu, "Object and patch based anomaly detection and localization in crowded scenes," *Multimedia Tools and Applications*, vol. 78, no. 15, pp. 21375–21390, 2019.
- [30] Z. Zivkovic, "Improved adaptive Gaussian mixture model for background subtraction," in *Proceedings of the International Conference On Pattern Recognition*, pp. 28–31, IEEE Computer Society, Stockholm, Sweden, 2004.
- [31] T. Wang and H. Snoussi, "Histograms of optical flow orientation for visual abnormal events detection," in *Proceedings of the IEEE-Ninth International Conference on Advanced Video and Signal-Based Surveillance*, pp. 13–18, IEEE Computer Society, Beijing, China, 2012.
- [32] R. Raghavendra, A. Del Bue, M. Cristani, and V. Murino, "Optimizing interaction force for global anomaly detection in crowded scenes," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 136–143, 2011.
- [33] N. Li, X. Wu, D. Xu, H. Guo, and W. Feng, "Spatio-temporal context analysis within video volumes for anomalous-event detection and localization," *Neurocomputing*, vol. 155, no. C, pp. 309–319, 2015.
- [34] Y. Zhang, H. Lu, L. Zhang, and X. Ruan, "Combining motion and appearance cues for anomaly detection," *Pattern Recognition*, vol. 51, no. C, pp. 443–452, 2016.
- [35] A. Al-Dhamari, R. Sudirman, and N.N. H. Mahmood, "Transfer deep learning along with binary support vector machine for abnormal behavior detection," *IEEE Access*, vol. 8, pp. 61085–61095, 2020.
- [36] Y. Cai, H. Jiang, H. Wang, and X. Chen, "Trajectory-based anomalous behaviour detection for intelligent traffic surveillance," *IET Intelligent Transport Systems*, vol. 9, no. 8, pp. 810–816, 2015.
- [37] P. Kristina, "Sinaga; miin-shen yang, unsupervised K-means clustering algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020.