Research on Multiple Abnormal Behaviour Detection Methods for Surveillance Video

Quange Tan and Yali Zhang*

People's Public Security University of China, Information Network Security Institute, Beijing 100038
*Correspondence Author Email: zhangyl_mail@163.com

Abstract. This paper uses the centroid acceleration of the moving target and the aspect ratio operator of the outer frame of the moving target to identify the human moving target. It uses inter-frame difference method to detect stagnation, uses 3D convolutional neural networks to detect fall, and analyzes the experimental results. Abnormal behaviour detection can assist in case investigation, large-scale activities, and security in key areas, etc., and help ensure social harmony and stability.

1. Introduction
Detection of abnormal behaviours based on video surveillance can enable the computer to automatically determine whether abnormal events have occurred in the picture, for example abnormal behaviours like cross-border, stagnation, hovering, falling, gathering, multiple occurrences, etc. The use of video structured technology can describe the characteristics of abnormal behaviours, mark and record the pictures, so as to achieve crime prevention, rapid alarm, and reduce the losses caused by emergencies.

This article will focus on the abnormal behaviour detection methods of crowding, staying and falling in different scenarios in video surveillance.

2. Moving Target Detection and Extraction
Moving object detection is to extract moving people or objects in a video sequence in order to track, classify, identify, and process the extracted moving objects. Moving target detection provides a binary image of moving targets to be identified for abnormal behaviour detection. There are two main methods for real-time moving target detection: background difference method and inter-frame difference method.

2.1. Background Difference Method
The background difference method is used for moving object detection of background still video images. The background difference method needs to first obtain the background image in the monitoring environment, and obtain the gray image of the moving object by performing a differential operation on the current frame and the background frame, which is called the foreground target.

The traditional background difference method is very sensitive to changes in lighting conditions, and will produce many wrong moving target points. This will affect the effect of target detection. Accumulated difference detection method can be adopted to reduce the situation of false detection. The specific steps are as follows:
Newly Acquired Image \[ \rightarrow^* a \rightarrow \text{Results} \rightarrow \text{Background} \]
\[ \downarrow^* (1-a) \]

**Figure 1.** Background difference method (Cumulative difference)

### 2.2. Feature Extraction

**2.2.1 Body identification**

By labeling the outer frame of a moving human body, the outer rectangular frame can be used instead of the outline of the human body. The center of mass of the binary image in the rectangle is used instead of the center of mass of the human body. At the same time, the human body identification facilitates the monitoring personnel to observe the characters in the video surveillance, and achieves the purpose of quickly finding and responding quickly.

**Figure 2.** Original image

**Figure 3.** Foreground binary image

**Figure 4.** Foreground border

In the study of the actors in the video, the study of the human center of mass can on the one hand simplify the complex human body. On the other hand, the human center of mass is an important attribute of human abnormal behaviour research. For example, the speed of movement of the center of
mass may represent the speed at which a person moves. The height of the center of mass can represent a person’s standing, squatting, and lying down states. When a person behaves abnormally, the center of mass may be too low or the center of mass may be outside the contour of the body. In the detection of abnormal behaviour of the crowd, it is also common to use the center of mass to study the behaviour of the group on behalf of the person.

Steps to calculate the gray level centroid of the outer border:
1) Calculate the sum of the gray level of all pixels in the outer border of the human body.
\[
\sum_{x_{\min} \leq x \leq x_{\max}, \ y_{\min} \leq y \leq y_{\max}} f(x, y)
\]
2) Calculate the sum of the product of the gray level of all the pixels in the bounding box of the human body and their abscissa.
\[
\sum_{x_{\min} \leq x \leq x_{\max}, \ y_{\min} \leq y \leq y_{\max}} f(x, y)x
\]
3) Calculate the sum of the product of the gray levels of all the pixels in the bounding box of the human body and their ordinates.
\[
\sum_{x_{\min} \leq x \leq x_{\max}, \ y_{\min} \leq y \leq y_{\max}} f(x, y)y
\]
4) The coordinates of the center of mass of the outer border of the human body are calculated as \((x_w, y_w)\),
\[
x_w = \frac{\sum_{x_{\min} \leq x \leq x_{\max}, \ y_{\min} \leq y \leq y_{\max}} f(x, y)x}{\sum_{x_{\min} \leq x \leq x_{\max}, \ y_{\min} \leq y \leq y_{\max}} f(x, y)}
\]
\[
y_w = \frac{\sum_{x_{\min} \leq x \leq x_{\max}, \ y_{\min} \leq y \leq y_{\max}} f(x, y)y}{\sum_{x_{\min} \leq x \leq x_{\max}, \ y_{\min} \leq y \leq y_{\max}} f(x, y)}
\]
This method can indicate the center of gravity of the target human body.

2.2.2 Feature operator
The characteristic operator of centroid acceleration and the aspect ratio of the bounding box is introduced as a simplified calculation method for studying the abnormal behaviour of the human body. Centroid motion acceleration:

When the moving target has abnormal movements such as rapid running or falling, the acceleration of the centroid movement will show a high value. Formula for calculating the acceleration of the center of mass:
\[
a = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}. \quad (x_1, \ y_1), \ (x_2, \ y_2)\]
Coordinates representing the centroid at the beginning and end of every second. In practical applications, the centroid acceleration threshold can be defined according to different video surveillance scenes, different camera distances, and ranges to classify the target's abnormal behaviour. According to the research by the US military doctor Paul Stapp, the maximum acceleration that the human body can withstand is 462 m/s². Therefore, if the acceleration of the center of mass exceeds 462 m/s² through real-time calculation, it is necessary to promptly warn and take relevant measures as soon as possible.

The outer frame aspect ratio:

The centroid motion acceleration reflects the overall motion state of the target person, and the
change in the aspect ratio of the outer border can reflect the change state of the target person. When the target task is squatting or falling, the outer frame aspect ratio will increase significantly in a short time. When a fight occurs, the width-to-height ratio of the outer frame will continue to vary greatly. The formula of the bounding box aspect ratio is $k = \frac{x_2 - x_1}{y_2 - y_1}$. In general, a person is walking or standing upright, with an aspect ratio $k < 1$. If the aspect ratio $k > 1$ occurs, the possibility of abnormal behaviour is greater.

3. Detection and Analysis of Common Abnormal Behaviour

3.1. Detention Detection and Analysis

Detention, especially in the crowd, can easily cause extremely bad effects, which can lead to vicious events. Therefore, the abnormal behaviour of the stay should be correctly identified and monitored.

Considering the behavioural characteristics of the stay, the occurrence of abnormal behaviour can be determined through the position detection at different time points in a period of time. Here we can use the inter-frame difference method. The inter-frame difference method is to extract a moving area of a detection object by making a difference between adjacent frame images in a comparison video. The specific process is as follows:

\[ D_t(x, y) = \left| f_t(x, y) - f_{t-1}(x, y) \right| \]

The gray level of the corresponding pixels of the two frames are used for differential calculation. When the lighting of the scene is stable, the inter-frame difference value $D_t(x, y)$ of the static area of the scene is zero. If in the area where the moving object is located, the inter-frame difference value $D_t(x, y)$ is not zero, this area is the area where the pixel changes, and it can be understood as the area where the motion occurs. However, due to the presence of noise in the surveillance video, the difference between frames in the static area is usually not equal to zero. This situation needs to be processed by threshold classification. The following formulas are used to classify and process the image thresholds to obtain differential images, which can effectively reduce the impact of noise and thus obtain the motion state of the target:

\[ R_k(x, y) = \begin{cases} 1 & D_k(x, y) > T \\ 0 & D_k(x, y) \leq T \end{cases} \]

$T$ is a custom threshold for classifying binary images.

In the experiment, after the differentially processed characters are continuously positioned, individuals who have been active for more than 30 minutes in a certain area are defined as detainees, and screenshots and calibration of the detained characters are performed. However, this method often calibrates the staff in the premises to reduce the recognition rate of staying. Information can be entered by staff who are often active in the area, and compared with the localized image information of the localized image using face recognition. Exclude images of staff and improve recognition efficiency.

3.2. Fall Detection and Analysis

Falls are extremely common anomalous behaviours, and fall detection plays an important role in
alerting to group events and preventing major risks. The following is a method for detecting fall of surveillance video based on 3D convolutional neural network. [9]

This method can extract the temporal and spatial characteristics of video images. \( v_{ij}^{xyz} \) is the eigenvalue of the i-th feature of the i-th layer at the (x, y, z) position, \( v_{ij}^{xyz} = F(b_j + \sum_m \sum_{p=0}^{P-1} \sum_{q=0}^{Q-1} \sum_{r=0}^{R-1} w_{ijm}^{pq} u_{(x+p)(y+q)(z+r)}) \). Where F is a non-linear function, P, R and Q are three dimensions for the convolution kernel. \( w_{ijm}^{pq} \) is the weight calculated on the convolution kernel (p, q, r) and the number m feature connection in the previous layer, and b is the bias.

The video sequence V with frame length L is input to a 3D convolutional neural network for recognition. Divide the video sequence V into video units with a frame length of l. Step size is \( \delta \), \( V = \{u_t\}_{t=1}^{T}, T = \left\lfloor \frac{L-1}{\delta} \right\rfloor \), where T is the total number of video units. Input video unit \( u_t \) to 3D convolutional neural network output \( y_t \). Calculate the probability \( p \) of n-type behaviours through the incentive function: \( P_t = \{P_t\}_{t=1}^n \), \( P_t = \frac{e^{y_t}}{\sum_{P=1}^{n} e^{y_t}} \). The mean value of the excitation function output of all units is the final result P. \( P = \frac{1}{T} \sum_{t=1}^{T} P_t \). The fall label corresponding to the largest probability \( P_{\text{max}} \) in P is the recognition result.

Falling behaviours are different from walking, standing, and sitting behaviours. Falling situations are complex and changeable. Therefore, sample data of fall behaviours in various situations should be added to the learning database for analysis. Traditional fall recognition requires a lot of preprocessing (such as extracting contours and background subtraction), which reduces the efficiency of recognition. In contrast, a 3-dimensional convolutional neural network can automatically extract the spatio-temporal features in the video and learn to achieve the role of classifying and predicting fall behaviour.

4. Summarization and Prospect
In this paper, targets are extracted by moving target detection. Background difference method are used for moving target detection. The human body is identified by the outline of the outline of the human body, and the centroid of the human body is calculated. The characteristic operators of the acceleration of the centroid movement and the aspect ratio of the outer border are used to assist the analysis of human behaviour. The specific detection methods of crowded, stranded, and fall abnormal behaviours are mainly studied.

At present, most of the research on abnormal behaviour detection stays at the experimental level, and very few systems are put into practical application. Because each video surveillance scene has its particularity, it is often necessary to measure the surveillance scene and perform monitoring experiments to determine the relevant detection threshold. With the rapid development of 5G technology and the emergence of high-end information technology, it is believed that the speed and volume of information transmission in the future will be greatly increased. The detection of abnormal behaviours in different scenarios may enable self-adaptation to different scenarios, thereby increasing the popularity of video surveillance abnormal behaviour recognition systems.

5. Acknowledgement
This paper is supported by People's Public Security University of China 2020 Basic Research Institute Focused on “Inherited Humans + Human Body Images for Automated Technology Research and Use” (2019LLYJGADX050). 2019 theory and soft science program of the Ministry of public security "Research on strategies to promote the training of professionals in police video image investigation technology in the era of big data” (2019llyjgadx050).

6. References
[1] Lijun Cui. Research on Detection Method of Human Behavior and Expression Abnormality Based on Video. Xian Province: Xi'an University of Technology, 2018: 1-25.
[2] Yukai Xiang. Review of human vision behavior recognition based on computer vision. Infrared, 2018, 39(11):1-6.

[3] Ximin Cai, Fangyu Hu. Detecting Abnormal Behavior in Examination Surveillance Video with 3D Convolutional Neural Networks. International Conference on Digital Home, 2016 6th:1-5.

[4] Lijun Wang, Ming Dong. REAL-TIME DETECTION OF ABNORMAL CROWD BEHAVIOR USING A MATRIX APPROXIMATION-BASED APPROACH[J]. IEEE, 2012:1-4.

[5] Min Qi. Introduction to Pattern Recognition. Beijing: Tsinghua University Press, 2007:42-45.

[6] Huaiqiang Li. Simulation of crowd anomaly detection method in sensitive area under behavior analysis. Computer Simulation, 2018, 35(8):1-4.

[7] Shishi Duan. Group abnormal behavior detection for video surveillance. Shanghai: Shanghai University, 2015:8-13.

[8] Huanli Feng, Hongyan Li. Intelligent detection simulation of crowded pedestrian abnormal behavior. Computer Simulation, 2018, 35(11):1-4.

[9] Xinwen Wang, Linbai Xie, Li peng. Residual network identification method for fall abnormal behavior. Journal of Frontiers of Computer Science and Technology. http://kns.cnki.net/kcms/detail/11.5602.tp.20190926.1324.004.html [10]

[10] Ling Bao. Research and System Design of Intelligent Number Statistics Technology Based on Deep Learning. Chengdu Sichuan Province: University of Electronic Science and Technology of China, 2015

[11] Jun Zhang. Anomaly detection of crowded people based on sports cricket. Modern computer, 2013

[12] Haiyang Yang. Crowd counting and short-term crowd flow prediction in crowded situations. Beijing: Beijing Jiaotong University, 2017

[13] Wanli Yu, Yumei Wei, Luqun Li. Human behavior recognition model based on multi-model fusion. Computer Engineering and Design, 2019, 40(10).