



Video crowd detection and abnormal behavior model detection based on machine learning method

Shaoci Xie¹ · Xiaohong Zhang¹ · Jing Cai²

Received: 17 May 2018 / Accepted: 9 August 2018 / Published online: 16 August 2018
© The Natural Computing Applications Forum 2018

Abstract

Pedestrian detection and abnormal behavior detection is the computer for a given image and video, to determine whether there are pedestrians and their behavior is normal. Pedestrian detection is the basis and premise of pedestrian tracking, behavior analysis, gait analysis, pedestrian identity recognition and so on. A good pedestrian detection algorithm can provide strong support and guarantee for the latter. The overall goal of this project is to learn different data mining methods and try to improve the detection accuracy of video crowd machine abnormal behavior. Aiming at the shortage of user behavior anomaly detection model proposed by Lane et al., a new IDS anomaly detection model is proposed. This model improves the representation of user behavior patterns and behavior profiles and adopts a new similarity assignment method. Experiments based on Unix user shell command data show that the detection model proposed in this paper has higher detection performance.

Keywords Machine learning · Image and video · Pedestrian detection · Anomaly detection

1 Introduction

There are two basic types of network intrusion detection technology, that is, anomaly detection and misuse detection. At present, anomaly detection is the main direction of intrusion detection research. This detection technology establishes the normal behavior pattern of the system or the user and detects the intrusion by the comparison and matching between the monitoring system or the actual behavior pattern of the user and the normal mode [1, 2]. It is characterized by the lack of excessive knowledge of the system defects, strong adaptability, and the ability to detect the unknown intrusion modes. The high probability of false alarm is the main factor that restricts the application of abnormal detection at present. The key problem of

anomaly detection is the establishment of normal behavior pattern and how to compare and judge the current behavior by using normal behavior pattern [3, 4].

The application of intelligent technology such as neural network and machine learning in abnormal detection has been carried out at home and abroad. The research goal is to improve the accuracy, real time, efficiency, and adaptability of the detection system. Some of the research results have been close to or reached the level of practicality in detection performance and operability [5]. This paper first introduces the user behavior anomaly detection model based on machine learning proposed by Lane et al. On the basis of which a new detection model is proposed, which is mainly used for the human intrusion detection system (IDS) with the shell command as the audit data [6, 7]. The model uses a variety of shell command sequences with different length to express the user's behavior pattern and sets up a number of sample sequences to describe the normal behavior contour of the legitimate users in the network. When the detection is detected, the similarity degree is assigned by the sequence of variable length commands, and the user behavior is judged by the similarity degree after the noise filtering. Experiments using

✉ Shaoci Xie
xiesc@cqupt.edu.cn

¹ School of Big Data and Software Engineering, Chongqing University, Chongqing 400030, China

² School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

Unix user shell command data show that the new detection model has high detection performance and strong operability.

2 A fixed length command sequence detection model for machine learning

2.1 Basic principle of machine learning

Machine learning is a new branch of artificial intelligence. Through the study of human cognitive mechanism, it establishes various learning models with the aid of machine (computer system) and gives the ability of machine learning. On this basis, it builds a task-oriented learning system with specific application. A machine learning system consists mainly of a learning unit, a knowledge base, and an executive unit, in which a learning unit builds a knowledge base by using information provided by an external source and makes improvements to it (adding new knowledge or reorganizing the existing knowledge). The execution unit performs tasks by using knowledge in the knowledge base, and the information after task execution is fed back to the learning unit as input for further learning. The learning unit is the core of the learning function of the machine learning system. It involves the way of dealing with the outside information and the methods used in the process of acquiring new knowledge. The knowledge base is used to store knowledge, including the original domain knowledge of the system (this knowledge is long term, relatively stable), and the various new knowledge acquired through learning (this knowledge is short term and varied). The choice of knowledge representation plays a very important role in the design of learning system. Execution unit is a key part that makes learning system practical and at the same time able to evaluate learning methods.

A large part of machine learning research focuses on two areas: classification and problem solving. After more than 30 years of development, there have been a lot of learning methods, such as inductive learning, case learning, genetic learning, etc., but these methods have their limitations. It is the mainstream of current research to explore new learning methods and algorithms in combination with specific application fields.

2.2 Description of sequence detection model of fixed length command

Lane of Purdue University has proposed an anomaly detection model of user behavior based on machine learning and conducted in-depth research and experiments on the model. The model uses the fixed length shell command sequence to represent the user's behavior pattern and

sets up the sample sequence library to describe the normal behavior profile of the legitimate users. When working, the current command sequence of the monitored user is compared and matched to the sequence library of the command and the behavior of the monitored user is judged according to the similarity of the two. The main points of the model are as follows:

1. The shell command sequence with fixed length is used as the minimum data unit to describe the user's behavior pattern, and an example learning method is used to establish the sample sequence library of the legitimate user.
2. A similarity degree between two sequences is defined, which is used to represent the similarity between the behavior patterns represented by the two sequences. On this basis, a similarity between a sequence and a sample sequence library is defined, which is used to represent the maximum similarity between the behavior patterns represented by the sequence and the various normal behavior patterns.
3. When the model works, it calculates the similarity between each sequence in the monitored user sequence stream and the sample sequence library. Then, the similarity degree is processed with window filtering, and the similarity judgment value arranged in time sequence is obtained. According to the size of the judgment, the behavior of the monitored user is judged in real time.

There are several key problems in the model:

1. The selection of optimal sequence length;
2. The extraction of sample sequence;
3. The definition of similarity function;
4. Selection of noise filtering algorithm.

Lane et al. performed a large number of experiments on UNIX user shell command data for the above problems. The following are the experimental results:

1. The optimal sequence length is related to the behavior characteristics of specific users. As the length of the sequence increases (from 1 to 15), the detection performance of the model changes with different users.
2. In a variety of similarity functions, the similarity function corresponding to the correlation between adjacent commands is superior to the similarity function without considering correlation. Mean noise filtering and median filtering algorithms have little difference in detection performance.
3. The clustering method is more adaptable to different users in the sampling method such as clustering, occurrence probability extraction, time sequence interception, random selection and so on.

3 Dense crowd state perception model

3.1 Technology roadmap

The population state perception process mainly consists of four processes, namely video acquisition and processing, optical flow calculation, vector light flow mapping to geographic space, population state detection and analysis, as shown in Fig. 1.

The concrete steps are as follows:

1. First, through the mobile phone or camera in the shopping mall, hospital, station, bank and other people crowded and easy to happen public events in a place to gather experimental data. It mainly includes normal crowd behavior video and abnormal behavior video, such as fighting, running, gathering, abrupt, retrograde and so on, providing test data for the detection of abnormal behavior of the crowd. Based on high-definition remote sensing image or GPS, the location of the center point of camera is collected. Based on camera calibration or direct setting, the camera's height, tilt angle, azimuth, and main distance are obtained [7].
2. Second, the optical flow field of each monitor video is calculated. Based on the two front and back images in the video, the real-time optical flow field of the current video is obtained by using the relevant optical flow calculation method.
3. Third, map the optical flow fields in each region to geospatial space. The optical flow field calculated by video image space is pixel coordinates. It is difficult to locate, measure or analyze in the geographic space. It needs to be projected into the map by the mapping method of image space and geospatial space.
4. Fourth, in geographic space, any vector light flow has a clear spatial reference. It is very easy to understand the position, direction, speed, acceleration, and other motion parameters of the crowd. It is an important basis for the detection and early warning of the crowd.

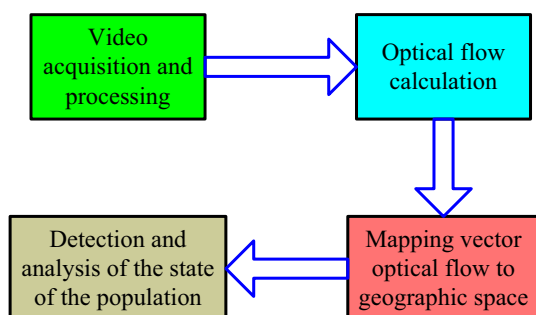


Fig. 1 Population state perception technical route

3.2 Key technology

The motion vector field based on video extraction is mainly dependent on the image space. The optical flow field is based on the image space in the direction of pixels, and the real measurement of the dynamic target in the geographic scene cannot be obtained, including the position, direction, and speed [8]. In order to make the state of the population be located, measurable, and analytical, and to have the advantage of reading, it is necessary to focus on the extraction of optical flow field and the mapping of the optical flow field to the geographic space.

3.2.1 Optical flow calculation method of Lucas–Kanade

Optical flow field computation is one of the most important key technologies of crowd state perception. The calculation of optical flow field is first proposed by American scholars HOM and SCHUNCK, and the improvement and optimization of optical flow field correlation algorithms are constantly promoted in the study. The optical flow field is widely applied to pedestrian detection, pedestrian gait analysis, and motion segmentation. Some scholars also evaluate the efficiency of optical flow method [9].

In this paper, the Lucas–Kanade optical flow method is used to obtain the motion state of the crowd. The method first converts video into sequence image and transforms it into sequence gray scale. In a short time, the gray value of the same point of moving target will not change and then the equation is constructed and the optical flow vector is calculated [10]. The change in crowd state can be expressed by field model. The change in sequence grayscale image is the expression of crowd movement. The faster the grayscale changes, the faster the crowd will move. A vector field can be formed by the constraint of the same gray value of the feature points in the near time. The magnitude and direction of each vector in vector field are the speed and direction of crowd movement. It should be noted that the vector field is not the real velocity of the crowd, but the perspective view of the crowd movement.

The Lucas–Kanade method searches the motion vectors of each pixel according to the change in the gray gradient of the adjacent frames. Based on the research of HORN and SCHUNCK, the Lucas–Kanade optical flow algorithm is effective. The method is a gradient-based locally parameterized optical flow estimation method. The method assumes that the optical flow vector in a small window in the image space remains unchanged, and the optical flow vector can be computed by the least square method.

It is assumed that the brightness of one pixel point (x, y) in the image plane is $E(x, y, t)$ at the t moment, $u(x_0, y_0)$ is

the horizontal variation component of the pixel point, and $v(x, y)$ is the vertical component of the pixel point.

$$u = dx/dt \quad (1)$$

$$v = dy/dt \quad (2)$$

After a time interval of Δt , the point corresponds to the brightness of $E(x + \Delta x, y + \Delta y, t + \Delta t)$, and when the Δt is very small to 0, it is assumed that the brightness remains unchanged, such as formula (3).

$$E(x, y, t) = E(x + \Delta x, y + \Delta y, t + \Delta t) \quad (3)$$

When the brightness of the point changes, the brightness of the moving point is developed by the Taylor formula as shown in Eq. (4).

$$E(x + \Delta x, y + \Delta y, t + \Delta t) = E(x, y, t) + \frac{\partial E}{\partial x} \Delta x + \frac{\partial E}{\partial y} \Delta y + \frac{\partial E}{\partial t} \Delta t + \varepsilon \quad (4)$$

If the derivative of the two order is infinitesimal, the Δt delta T is 0, as shown in formula (5).

$$-\frac{\partial E}{\partial t} = \frac{\partial E}{\partial x} \frac{dx}{dt} + \frac{\partial E}{\partial y} \frac{dy}{dt} = \nabla E g w \quad (5)$$

The formula $w = (u, v)$ is the basic optical flow constraint equation.

$E_x = \partial E / \partial x$, $E_y = \partial E / \partial y$, $E_t = \partial E / \partial t$ shows the gradient of the pixel gray level in the x, y, t direction in the image and can be rewritten as: $E_x u + E_y v + E_t = 0$, that is, gray consistency constraint conditions. The input video size is $X \times Y$; all the Harris corners of the current image are detected as the feature points; then, the Pyramid LK optical flow method is used to take the characteristic window of $w \times w$ and to solve the optical flow UV of the feature points based on the constraint conditions of the optical flow, such as formula (6).

$$\begin{bmatrix} E_{x1} & E_{y1} \\ E_{x2} & E_{y2} \\ \dots & \dots \\ E_{xn} & E_{yn} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} -E_{t1} \\ -E_{t2} \\ \dots \\ -E_{tm} \end{bmatrix} \quad (6)$$

That is $Av = -b$.

Then, the least squares method is used, such as formulas (7) and (8).

$$A^T A v = A^T (-b) \quad (7)$$

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \sum E_{xi}^2 & \sum E_{xi} \sum E_{yi} \\ \sum E_{xi} \sum E_{yi} & \sum E_{yi}^2 \end{bmatrix} \begin{bmatrix} -\sum E_{xi} & \sum E_{ti} \\ -\sum E_{yi} & \sum E_{ti} \end{bmatrix} \quad (8)$$

The experimental video contains a total of n_{Frame} frames, and the time is t_{Sum} seconds. The time required for each frame can be calculated by formula (9).

$$t_{\text{Each}} = t_{\text{Sum}} / n_{\text{Frame}} \quad (9)$$

The summation is calculated from 1 to n and calculated by image derivative and accumulated separately. At the same time, the weight function $W(i, j, k)$, $i, j, k \in [1, m]$ is needed to highlight the coordinates of the central point of the window, and the Gauss function is used to make the calculation effect obvious, and the real-time optical flow field of the current video can be obtained. The algorithm has better monitoring effect for specific application scenarios, such as station crowd, and has broad application prospects.

3.2.2 Mapping the optical field to 2D geospatial

The optical flow field obtained by the video has its coordinates in pixel coordinates, which is difficult to locate, measure, or analyze in the geographic space. It needs to be projected onto the map through the mapping method of image space and geographic space. The clear spatial reference and macroscopic observation perspectives of the map will help to grasp the movement of the large-scale population as a whole. Mapping optical flow field to 2D geospatial is another key technique of this paper [11].

The traditional method needs to calculate the 2D homography matrix. Firstly, three or more control points are selected in the image to obtain the image coordinates of the control points; then, corresponding points are selected in the high-definition remote sensing image or map to obtain the spatial right-angle coordinates of the control points; finally, the 2D homography matrix H is calculated. This method not only requires interactive selection, but also requires high-definition remote sensing images or topographic maps. It is suitable for small-scale video surveillance scenarios and is difficult to use in large-scale video surveillance systems. In particular, the 2D homography approach is ineffective when acquiring 3D geometric information of the target. In this context, Xingguo et al. proposed a mutual mapping model for surveillance video and 2D geospatial space, which can better achieve a one-to-one correspondence between image space and two-dimensional geospatial space.

In order to obtain the real three-dimensional information of the crowd, taking into account the PTZ camera, this article has introduced the mutual mapping model of surveillance video mapping to 2D geospatial. This model requires the camera internal and external parameters to be known for ease of calculation. The camera model is represented by a homogeneous equation, such as formula (10).

$$\lambda \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & s & u_0 & 0 \\ & f_y & v_0 & 0 \\ & & 1 & 0 \end{bmatrix} \begin{bmatrix} R & -R\bar{c} \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (10)$$

In the formula, (X, Y, Z) is the space right-angle coordinates in GIS, \bar{c} is the position of camera center in the space right-angle coordinate system, (x, y) is the image space coordinate, (u_0, v_0) is the main point coordinates, $f_x = f/p_x, f_y = f/p_y$ is the equivalent focal length, s is the non-orthogonal coordinate axis, and R is rotation transformation.

In the formula (10), the camera’s internal and external parameters are hidden in the matrix. In order to facilitate the use of the camera model, a camera model which is easy to use, such as formula (11), is obtained through comparative analysis of camera models of photogrammetry and computer vision.

$$\lambda \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & u_0 \\ & fv_0 \\ & 0 \end{bmatrix} \begin{bmatrix} a_1 & b_1 & c_1 \\ -a_2 & -b_2 & -c_2 \\ -a_3 & -b_3 & -c_3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -X_x \\ 0 & 1 & 0 & -Y_y \\ 0 & 0 & 1 & -Z_z \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (11)$$

The parameter $u_0 = x_{0ph}, v_0 = H_{pic} - y_{0ph}$ is the coordinates of the principal points, and H_{pic} is the image height, which is consistent with the meaning of the internal and external parameters in photogrammetry.

3.3 Experiment and analysis

3.3.1 Experimental platform and data

Based on MATLAB, OpenCV, VSC#2012 and ArcEngine as an experimental platform, this paper integrates the algorithms implemented under various environments and develops a prototype system for the state perception of a group of individuals. The basic library of computer vision under MATLAB and OpenCV has many basic functions, such as video real-time access, video file playback, video frame acquisition, image size adjustment and so on. Based on this, the proposed algorithm is implemented and encapsulated as DLL. The related DLL is invoked under VSC#2012, and the prototype system is developed combined with the two development component ArcEngine of GIS.

In this paper, a number of video data containing people are collected from the restaurant and teaching building in a university, and QuickBird remote sensing images and high precision vector maps are collected in this area [12]. In order to obtain the internal and external reference of the camera, the method of extermination and extermination is

used to calculate the main distance, and the external reference of the camera is obtained by combining high-definition remote sensing images, and the constraint rotation angle is 0. Thus, the mapping from video data to geospatial space is realized.

3.3.2 Experimental results and analysis

The experiment mainly includes three aspects: optical flow extraction, optical flow to geospatial mapping, and spatial analysis of optical flow. The video data of the restaurant and the teaching building are extracted from the restaurant and the teaching building, which reflects the trend of the crowd movement. As shown in Fig. 2, the arrow points to the direction of the optical flow, and its length represents the value of the optical flow. When someone moves normally in the crowd, the arrow tends to converge in a similar direction; otherwise, the arrow will be cluttered and the length will be longer.

Based on the mutual mapping model between video and image space and geospatial, the optical flow extracted from the video can be mapped to GIS in real time. In the form of arrows, GIS has clear spatial reference and measurement information, which can grasp the motion state of the regional population from the global perspective. As shown in Fig. 3, the prototype system displays the optical flow field at a certain time.

Based on GIS, spatial analysis of optical flow field can be carried out. Scattered point interpolation and contour generation experiments are carried out in this paper. The IDW method is selected for the scatter point interpolation. As shown in Fig. 4, we can see that the darker the color is, the faster we can understand the velocity field of the crowd movement according to the depth of the color. Interpolation-based DSM can automatically generate contour isoline distribution and also reflects the movement state of the crowd. The smaller the change in the value, the greater the



Fig. 2 Detection of abnormal behavior in people

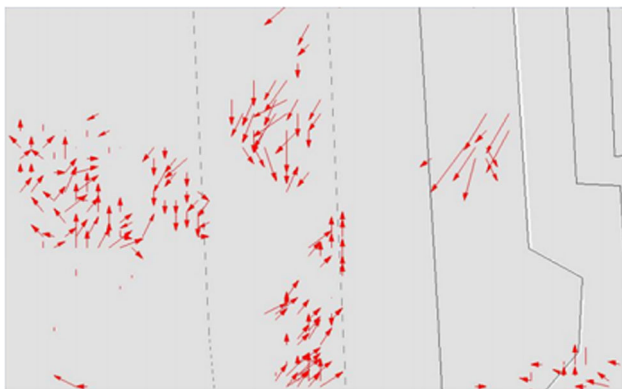


Fig. 3 Optical flow field in a map

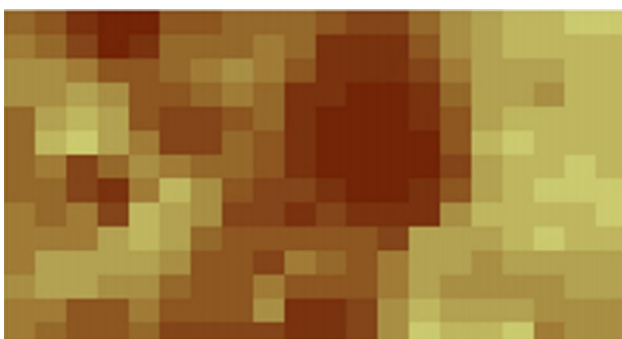


Fig. 4 Optical flow interpolation

distance between the lines. The faster the value goes up or down, the smaller the line spacing.

4 Variable length command sequence behavior detection model based on machine learning

The detection of abnormal events can be divided into two categories: one is the local anomaly detection, and an abnormal event is detected in a certain area of a given frame in a given video. The two is global anomaly detection. In a given video, it detects abnormal events at the beginning of a frame, aiming at the whole frame, rather than a certain area [13]. In general, the global anomaly detection means that each frame in the video is judged to be normal or abnormal. Since an early warning signal for one frame in the video can meet the requirements of video surveillance, this paper mainly studies the second methods, that is, global anomaly detection.

4.1 Description of variable length command sequence detection model

Lane et al. have two disadvantages: the fixed length command sequence detection model.

It lacks flexibility and adaptability in the representation of user behavior patterns. Behavior pattern refers to a certain regularity embodied in the process of user operation: in practice, the behavior patterns of different users are different, and the number of orders executed by the same user when completing different behavior patterns is not the same [14]. Therefore, the command sequence with fixed length is difficult to represent the overall behavior profile of users in a comprehensive and accurate way.

1. It is not easy to estimate the optimal sequence length for specific users. Lane et al. mainly use experimental methods to determine the optimal sequence length [15]. This method requires a large amount of computation, and its performance lacks stability.
2. We improved and corrected the above shortcomings of the fixed length command sequence detection model and proposed a variable length command sequence detection model, which is described as follows.

4.2 Algorithm flow

Figure 5 illustrates the process of crowd abnormal event detection in this paper. The purpose of crowd event detection is to classify each frame as normal or abnormal [16].

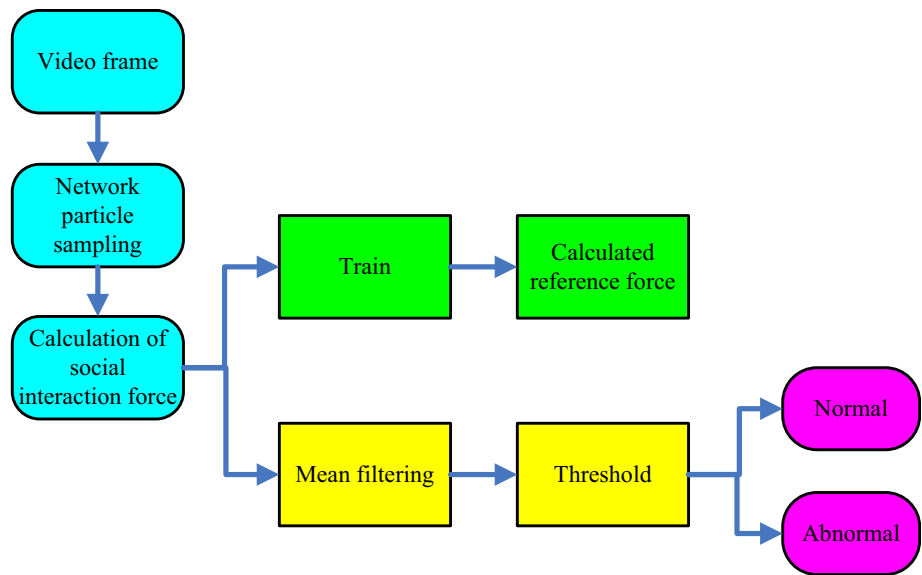
The algorithm flow is as follows:

1. First, a grid segmentation is performed for each video frame to extract a series of particle points and each particle point is treated as an entity.
2. Calculate the reference interaction force of the training set. In the data set, a part of the normal video frames of the crowd is extracted and used as a training set. In most of the standard data sets, there will be an initial video segment that is normal pedestrian behavior [17]. This segment can be used as a reference. Let N be the number of samples of the training set, which is the number of particles extracted in each frame, which is the reference to the social interaction under normal conditions:

$$F_r = \frac{1}{N} \sum_{i=1}^k F_{\text{int}}(x_i) \quad (12)$$

3. For each frame of the test set, calculate the sum of the social forces of all particles, as shown in Eq. (13), and then calculate the distance between the social force of

Fig. 5 Crowd anomaly detection flowchart



the current frame and the reference force, as shown in Eq. (14):

$$F_t = \sum_{i=1}^k F_{int}(x_i^{new}) \tag{13}$$

$$C_t = |F_t - F_r| \tag{14}$$

Finally, we use mean filtering to reduce the abrupt jump of a frame and get the filtered result after the difference of the magnitude of the interaction force calculated by formula (14). Next, each frame is divided into abnormal or normal classes according to the threshold value, as shown in formula 15:

$$L_t = \begin{cases} \text{Abnormal,} & \text{if } C_t > th \\ \text{Normal,} & \text{otherwise} \end{cases} \tag{15}$$

The whole process of crowd abnormal event detection algorithm based on variable length command sequence detection model is shown Fig. 6:

Calculating the interaction force between crowds and shading is difficult. The particle flow calculation method provides another feasible method for the interaction force of the crowd. In densely populated individuals, the movement of individuals is similar to the gradual movement of a moving particle [18]. The optical flow in the crowd shows the direction of the crowd movement.

1. Crowd movements are captured by optical flow grids.
2. In the crowd, people tend to move smoothly, so the optical flow mean can be used to detect the convection of particles frequently.

A normal person in the crowd of interaction force diagram is shown in Fig. 7, in which red represents the calculated interaction force, yellow indicates the optical flow.

5 Experimental results and analysis

The video data set selected in this paper comes from a square denser data set, which includes video frames for normal and abnormal behavior of the crowd. The data set has two different scenes, while the left of Fig. 8 represents the behavior of the normal crowd, and the left of Fig. 9 represents abnormal behavior. Normal and abnormal behaviors were tested, including normal walking and standing, abnormal behavior, and sudden agitation [19]. In order to simplify the calculation model, the grid particle resolution in the interactive force flow field is set to 20% of the total number of pixels, that is, a grid particle is extracted for every 5 pixels, and the panic factor is fixed to 0.25.

First, for each video frame, the social interaction force is calculated. Figure 8 is an interactive force flow map under the normal behavior of the crowd. Blue indicates that the interaction force is smaller in the region, and the red indicates that the interaction force is larger. It can be seen that under normal conditions, the force flow is mainly distributed in one or two regions with low interaction force, which indicates that the social interaction force of the population is at a lower level.

Figure 8 shows a social interaction force flow diagram of one of the abnormal behaviors in the crowd video. It can be seen that the red area in the thermogram is more than that in Fig. 9. The red region indicates that the social interaction force is larger, that is, the activity of the crowd is more intense and frequent, and that the probability of the abnormal behavior of this video image is larger.

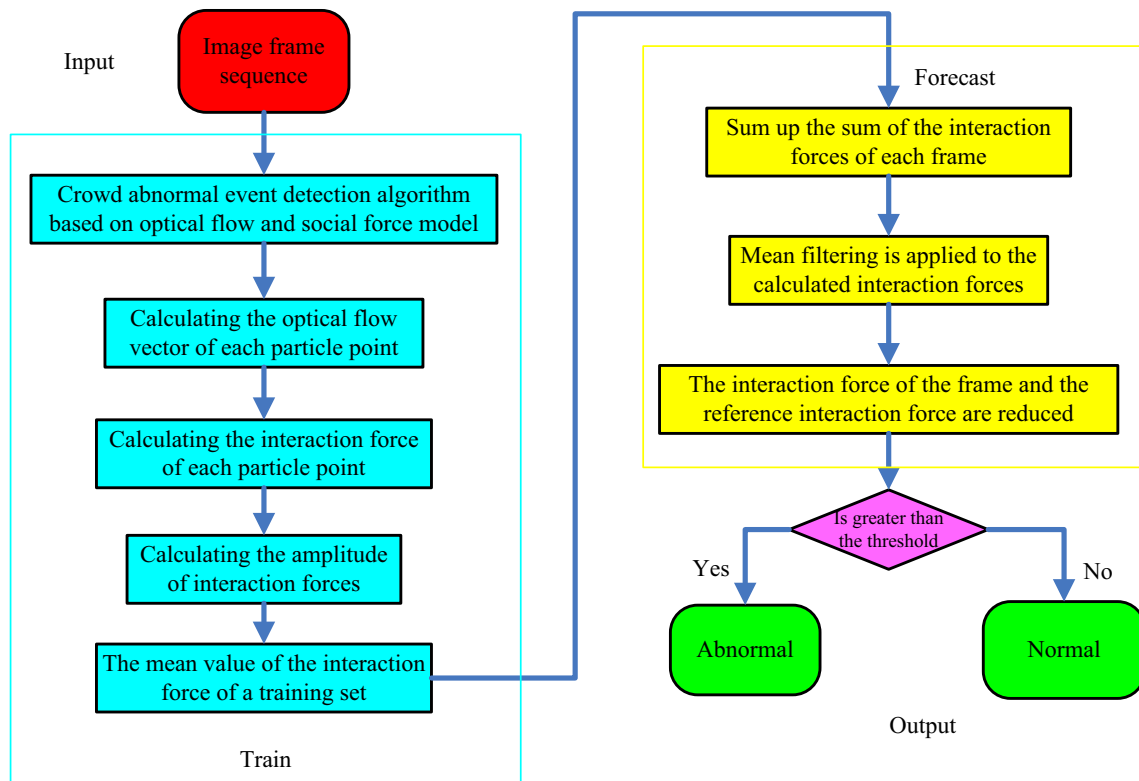


Fig. 6 Anomaly behavior detection algorithm based on variable length command sequence detection model

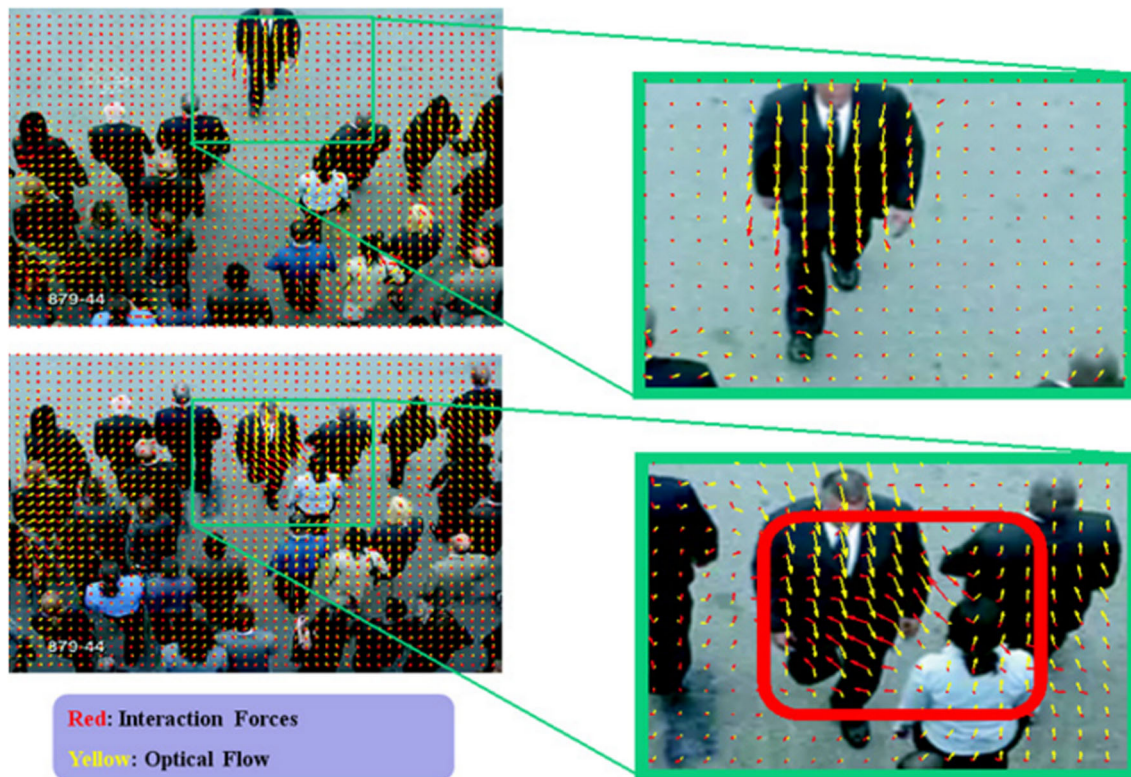


Fig. 7 Interaction force of normal people in the population

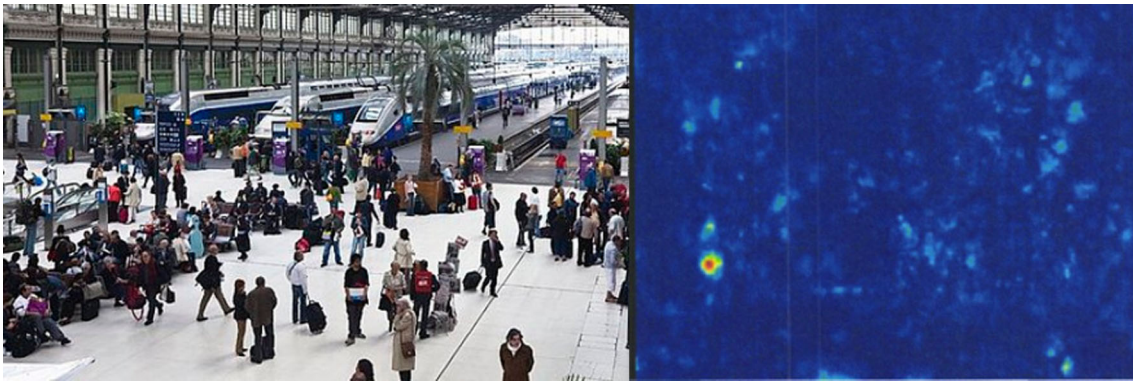


Fig. 8 Social interaction force flow diagram under normal behavior of the population

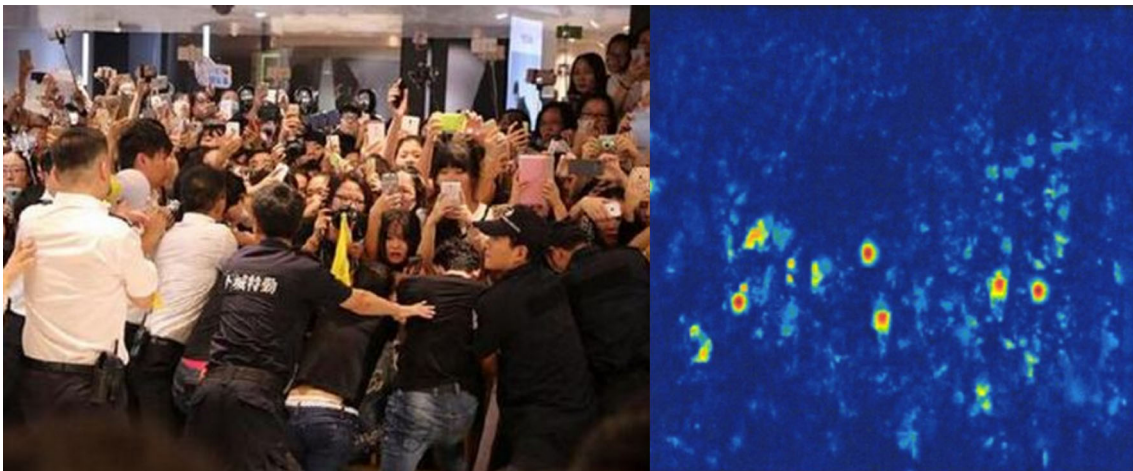


Fig. 9 Social interaction force flowchart under abnormal crowd behavior

6 Conclusion

This paper proposes a new machine learning-based IDS video crowd and its behavior anomaly detection model, and uses Unix user's shell command data to conduct experiments. Experimental results show that the detection performance of the new model is much better than that of Lane et al. Because the learning method and detection algorithm in the model have some adaptability to different detection data, this model can also be used for other data types (such as system call) outside the shell command (such as system call), but the specific application scope and detection performance need further research and experiment. How to further improve the speed, real-time immunity and robustness of the system is the focus of further research.

Acknowledgements This paper is supported by the Science and Technology Research Project of Chongqing Municipal Education Committee (Grant: KJ1704089).

References

1. Mohammadi S, Galoogahi HK, Perina A, Murino V (2017) Physics-inspired models for detecting abnormal behaviors in crowded scenes. *Group Crowd Behav Comput Vis* 1(12):253–272
2. Rabiee H, Mousavi H, Nabi M, Ravanbakhsh M (2017) Detection and localization of crowd behavior using a novel tracklet-based model. *Int J Mach Learn Cybern* 2(2):1–12
3. Wang X, Gao M, He X et al (2014) An abnormal crowd behavior detection algorithm based on fluid mechanics. *J Comput* 9(5):1144–1149
4. Cui J, Liu W, Xing W (2014) Crowd behaviors analysis and abnormal detection based on surveillance data. *J Vis Lang Comput* 25(6):628–636
5. Roubtsova ANS, With PHND (2013) Group localisation and unsupervised detection and classification of basic crowd behaviour events for surveillance applications. *Proc SPIE Int Soc Opt Eng* 8663(3):135–145
6. Zhang X, Wang M, Zuo J et al (2015) Abnormal crowd behavior detection based on motion clustering of mesoscopic group. *Yi Qi Yi Biao Xue Bao/Chin J Sci Instrum* 36(5):1106–1114
7. Abardeig J, Cai J, Zhu Z (2016) Study on the method of detecting the crowd abnormality in the sensitive media image based on behavior analysis. *Recent Adv Electr Electron Eng* 9(1):29–33

8. Zhu S, Hu J, Shi Z (2016) Local abnormal behavior detection based on optical flow and spatio-temporal gradient. *Multimed Tools Appl* 75(15):9445–9459
9. Alvar M, Torsello A, Sanchez-Mirallas A et al (2014) Abnormal behavior detection using dominant sets. *Mach Vis Appl* 25(5):1351–1368
10. Zhang X, Zhang Q, Hu S et al (2018) Energy level-based abnormal crowd behavior detection. *Sensors* 18(2):423
11. Chan YT (2017) Extracting foreground ensemble features to detect abnormal crowd behavior in intelligent video-surveillance systems. *J Electron Imaging* 26(5):051402
12. Balasubramanian Y (2015) Human crowd behavior analysis based on graph modeling and matching in a synoptic video. *Volume 3(3):1050–1056*
13. Zhao F, Li J (2014) Pedestrian motion tracking and crowd abnormal behavior detection based on intelligent video surveillance. *J Netw* 9(10):2598
14. Sang HF, Yu C, Da-HE K (2016) Crowd gathering and running behavior detection based on overall features. *J Optoelectron Laser* 27(1):52–60
15. Pan S, Sun W, Zheng Z (2017) Video segmentation algorithm based on superpixel link weight model. *Multimed Tools Appl* 76(19):19741–19760
16. Shehab A, Elhoseny M, Muhammad K, Sangaiah AK, Yang P, Huang H, Hou G (2018) Secure and robust fragile watermarking scheme for medical images. *IEEE Access* 6(1):10269–10278
17. Zheng Z, Huang T, Zhang H et al (2016) Towards a resource migration method in cloud computing based on node failure rule. *J Intell Fuzzy Syst* 31(5):2611–2618
18. Tharwat A, Mahdi H, Elhoseny M, Hassanien AE (2018) Recognizing human activity in mobile crowdsensing environment using optimized k-NN algorithm. *Expert Syst Appl* 107(1):32–44
19. Zheng Z, Jeong HY, Huang T et al (2017) KDE based outlier detection on distributed data streams in multimedia network. *Multimed Tools Appl* 76(17):18027–18045