

Violent Behavior Detection Based on SVM in the Elevator

Guang Shu, Gaojing Fu, Peng Li and Haiyu Geng

Department of Computer Science and Technology, Harbin University of Science and Technology, Harbin, 150080, China

Email: 125322223@qq.com, zsg0601@gmail.com

Abstract

To avoid fighting and violence occurred in the elevator, this paper proposed an abnormal behavior detection method based on SVM to achieve real-time monitoring. Firstly, the corners of the video sequences were detected and the Lucas-Kanade algorithm was used to calculate the optical flow to obtain velocity vector information. Secondly, this algorithm established a feature vector combining the corner kinetic energy with movement characteristics of targets (including change rate of area, change rate of external rectangle length-width ratio, distance between the targets and the angle difference of target movement direction) as the basis of violent behavior detection. Finally, SVM classifier was constructed to identify the violent behavior. The experiment results showed that the method could detect violent behavior in the elevator effectively and the algorithm was with less complex calculation and higher detection rate thus it could alarm real-time.

Keywords: *video sequence, corner kinetic energy, velocity vector information, SVM*

1. Introduction

As the popularization of vertical transportation vehicle, elevator brings convenience and quickness to people, while its safety problems have caused widespread concern within the elevator manufacturers and the public. Due to the closure environment in the elevator, it is extremely easy to occur to women's sexual harassment and violence, such as fighting and robbery.

At present, the studies on video abnormal detection are still at the primary stage all over the world. The commonly used methods are as follows: (1) The method based on matching the template [1]. The Literature [2] extracted motion descriptors from a lot of static pictures of human behavior firstly, then quantified and split, built an offline classifier based on the inductive characteristics as a N-dimensional feature vector, finally, made a comparison between the video sequence and the template classifier to obtain the classification results. This method has a better recognition rate to the static objects and some simple behavior, while templates are complex to build. (2) The method based on state space model [3, 4]. The literature [5] taken the static characteristics of each person's status as a state node in the state graph, identified human behavior by traversing the state nodes in the graph. It is difficult for the algorithm to apply to the actual real-time monitoring for its complexity and large amount of calculation. (3) The method based on the dynamic characteristics of the image [6]. The literature [7] proposed an overall assessment of the image information by analyzing video sequences of dynamic information such as trajectory and optical flow in an unknown application scene. The method estimated the information by calculating the energy of the whole image with the large amount of calculation.

2. System Fundamentals

This paper proposed a feature extraction method of human motion behavior on the combination of corner kinetic energy and dynamic characteristics. The algorithm mainly includes the following sections:

(1) Extract the foreground image and judge the number of people in the elevator: The Gaussian mixture background model and background subtraction method are used to extract the target prospects. Each separated connected region is identified by labeling algorithm of the connected domains. This paper adopts the method of statistics the number of pixels in each connected region to obtain the number of people in the elevator. Finally, the algorithm takes the different approaches to detect abnormal behavior for single and multiplayer situations. In the case of multiplayer situation, the main unusual behavior to be detected is like fighting, looting and robbery.

(2) Extract dynamic features: Extract the movement characteristics of target mass as the basis for the identification of violent behavior.

(3) Calculate optical flow of the corner and the corner kinetic energy: Using the Harris operator to detect the corners, through Lucas-Kanade local optical flow method to calculate the optical flow of the corner, energy value of the current frame can be obtained according to the optical flow information of all corners. The main purpose is to reflect drastic and chaos degree of the violent motion according to the changes of corner kinetic energy.

(4) Establish an abnormal behavior detection model and make judgments: A feature vector of the corner kinetic energy modal and movement characteristics of target prospects (including change rate of area, change rate of external rectangle length-width ratio, distance between the targets and angle difference of target movement direction) is established to be the basis of violent behavior detection. After inputting the feature vector into the SVM classifier for training, the algorithm implements the decision of violence.

2.1. Foreground Extraction and Identify the Number of People

At present, the typical methods of foreground extraction are as follows: the inter-frame difference method, the optical flow method and the background subtraction method. Because of the environment is relatively stable in elevator. It is not sensitive on the light, rain, snow and other natural factors. This paper takes the Gaussian mixture method to build background model because of its strong ability to adapt to the changes of environment. The main idea of it is to establish the κ (usually 3~5) multidimensional Gaussian distribution to mixed simulate background value for each pixel point. To the pixel (x_0, y_0) , $x_t = (r, g, b)^T$ is observation value when time is t , $P(x)$ is the probability of belonging to the background :

$$P(x) = \sum_{i=1} \omega_{i,t} \eta(x_t, \mu_{i,t}, \sigma_{i,t}^2)$$

Among, $\eta(x_t, \mu_{i,t}, \sigma_{i,t}^2)$ represents the i^{th} Gaussian distribution model when time is t . $\omega_{i,t}$ ($\sum \omega_{i,t} = 1$) is the weight corresponding to each model of distribution. When detecting foreground, the method matches the current frame pixel values with the background model according to the inequality $|\mu_{i,t} - x_t| < 2.5\sigma$.

The results of using the Gaussian mixture model and the background subtraction method to extract the foreground are shown in Figure 1.

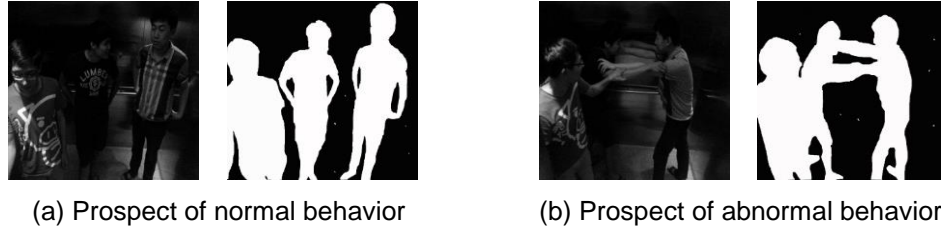


Figure 1. Foreground Extraction Results

This paper manually sets up an effective interested detection area. The next step is to identify each separate connected region by the connected domain recognition algorithm.

It is easy to get the information about each connected region after detecting connected domains. The algorithm is going to identify the number of people. This paper adopts the method that based on statistics the number of outlook connected region pixels to distinguish both single-player and multiplayer which is described in the literature [8], the main idea is to make statistics about total numbers of the pixels in the connected regions which including different numbers of the human and summarize the number of pixels when different numbers of people appearing in the connected regions.

2.2. Characteristics Extraction

A. Extract Movement Characteristics

The existing surveillance monitoring scenes are commonly large and restricted by the resolution ratio of the device. It is difficult to build a model of the human body because of the detailed characteristics because of human bodies are not obvious in the video. Considering that fighting and robbery these two kinds of abnormal interaction, the movement is intense and the motion characteristics have great differences when they happen. So the motion information of the video sequences is applied in this paper as a feature vector, it not only shows the human behavior characteristics completely, but also avoids the complexity of calculation effectively. The selected movement characteristics in this article are as follows:

(1)The change rate of area $r(s_i^t)$.

$$r(s_i^t) = \frac{|area(s_i^t) - area(s_i^{t-1})|}{area(s_i^{t-1})}$$

In this formula, $area(s_i^t)$ represents the area of the i^{th} target block when time is t .

(2)The change rate of external rectangle length-width ratio $r(c_i^t)$

$$r(c_i^t) = \frac{|r(c_i^t) - r(c_i^{t-1})|}{r(c_i^{t-1})}, r(c_i^t) = \frac{L_i^t}{W_i^t}$$

Among, L_i^t and w_i^t , respectively represents the length and width of the minimum external rectangle, the length-width ratio of the i^{th} target block is represented by $r(c_i^t)$ when time is t .

(3)Distance between the targets $d_{i,j}^t$

$$d_{ij}^t = |P_i^t - P_j^t|$$

In this formula, p_i^t and p_j^t respectively means the i^{th} target block and the j^{th} target block coordinate of the center position when time is t .

(4)The angle difference of target movement direction ϕ'_{ij}

$$\phi'_{ij} = \angle (\theta'_i, \theta'_j)$$

In this formula, θ'_i and θ'_j respectively represents angle difference of movement direction of the i^{th} target block and the j^{th} target block when time is t , and they can be obtained by optical flow method.

B. Calculation of Corner Kinetic Energy

The corner can effectively reduce the amount of data information and meanwhile keep the important characteristics of graphic, so the way of calculating the optical flow of corner is used in this paper, it can remarkable reduce the amount of data, and improve the operation speed of the algorithm.

(1) Corner Detection

Corner is an important feature of the image, it is the larger point of gray scale edge curvature, may reflect the outline information of the image, the corner detection method based on Harris operator is applied in this paper.

(2) Calculate the optical flow of corner

The changes of the image could be reflected in the optical flow, it contains the target motion information, thus can be used to determine the movement of the objective for observers.

Lucas-Kanade algorithm is a calculation method of sparse optical flow, it is a kind of local method, can set optical flow computation window according to the size of the image with smaller calculation, and is not sensitive to noise. Lucas-Kanade algorithm assumes that in a small neighborhood area Ω , the motion vector remains constant, then using the weighted least squares to estimate the optical flow. Through the Lucas-Kanade algorithm can calculate the optical flow of the corners. The optical flow contains velocity vector information.

Videos are not just seen as the accumulation of image frames in the time domain, the links and changes between the pixels of frame sequences also imply abundant energy information. In this paper, starting from the idea that the moving objects have kinetic energy, each corner is viewed as a particle. The method first calculates the kinetic energy of each corner, and then summarizes all the kinetic energy values of each corner in the target blocks to get the total kinetic energy of the whole target. Inside the elevator, when the behavior is normal, the total kinetic energy will be generally stable, and there is little change, but when fighting, robbery, such abnormal behavior happens, the total kinetic energy will increase and there will change obvious. Therefore, detecting abnormal behavior with the change of the kinetic energy is effective.

As is known in kinematics knowledge, the kinetic energy formula for the object is $E = \frac{1}{2} m v^2$. In this paper, each corner is regarded as a particle, and we can get the kinetic energy formula of the i^{th} corner in a frame by imitating the kinetic energy formula:

$$E_i(x, y) = w_i(x, y) \cdot V_i^2(x, y)$$

Among, $E_i(x, y)$ refers to the kinetic energy of the i^{th} corner; $w_i(x, y)$ is the weight of the i^{th} corner; $V_i^2(x, y)$ is the velocity model value of the i^{th} corner. So there comes the formula of the target total kinetic energy:

$$E(n) = \sum_{i=1}^m E_i(x, y) \quad (1)$$

$E(n)$ is the target total kinetic energy in the N^{th} frame. m is the number of effective corners of the target in the N^{th} frame.

The messier the velocity direction of pixel is, the more likely abnormal behavior happens. The weight $w_i(x, y)$ is calculated by the change angel of the corner velocity vector in two adjacent frames, the bigger change it is, the bigger weight there will be. Given $angle_i^{k-1}(x, y)$ is the velocity direction value of the i^{th} corner in $k-1^{th}$ frame, the weight is:

$$w_i(x, y) = \frac{|angle_i^k(x, y) - angle_i^{k-1}(x, y)|}{\pi} \quad (2)$$

Combining (1) with (2), the kinetic energy calculation formula of the n^{th} frame is:

$$E(n) = \sum_{i=1}^m \frac{|angle_i^k(x, y) - angle_i^{k-1}(x, y)|}{\pi} \cdot V_i^2(x, y)$$

The above four kinds of different motion characteristics and the corner kinetic energy of the targets are used to build the feature matrix F_i^m :

$$F_i^m = \begin{bmatrix} E_i^1 & E_i^2 & \cdots & E_i^n \\ r(s_i^1) & r(s_i^2) & \cdots & r(s_i^n) \\ r(c_i^1) & r(c_i^2) & \cdots & r(c_i^n) \\ d_{ij}^1 & d_{ij}^2 & \cdots & d_{ij}^n \\ \phi_{ij}^1 & \phi_{ij}^2 & \cdots & \phi_{ij}^n \end{bmatrix}$$

Each column of the matrix respective represents characteristics of the i^{th} object in a frame.

3. Support Vector Machine

SVM (Support Vector Machine) is a new classifier of machine learning methods based on statistical learning theory and pattern recognition. This model has a significant effect on dealing with the small samples, high-dimension, nonlinear and local minima problems.

Through nonlinear mapping ϕ , the inputted low-dimensional Euclidean space is mapped to a high-dimensional inner product linear feature space Z by SVM classifier to make the data set linear and separable in space Z . The mapping formula is $\phi(X) = [\phi_1(x), \phi_2(x), \dots, \phi_n(x)]^T$.

By the principle of Hibert, it could define a inner product function $K(x_i, x_j)$ as the kernel function which satisfies the Mercer conditions:

$$K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$$

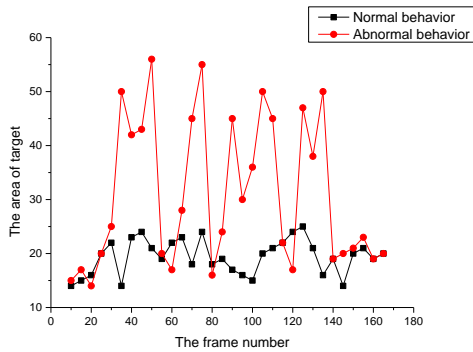
Then, the nonlinear problem in the low dimensional space is converted to the linear classification problem in the high dimensional space. The commonly used kernel functions include polynomial kernel function, the radial basis kernel function, the sigmoid kernel function. By choosing different kernel functions, the SVM turns undivided linear sample space to the high dimensional feature space, and different kernel functions can be used to construct different SVM models.

4. Experimental Results and Analysis

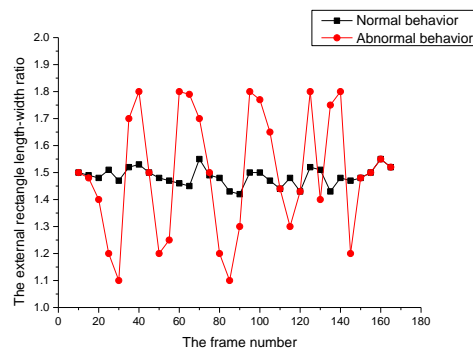
In this paper, the data sets come from video surveillance segments of the elevator monitoring and the fighting videos are collected by the research team. The experiment

environment is a laptop with 1.60GHz CPU, 1GB SDRAM Pentium5, Windows 7 operation system. The camera is 300 thousands pixels of CCD camera. LIBSVM software development kit is used in this paper to train the classifier and classification. There are 200 positive samples which category label is +1, there also have 60 negative samples which category label is -1, all of the videos are divided into training set and test set according to the ratio of 2:1.

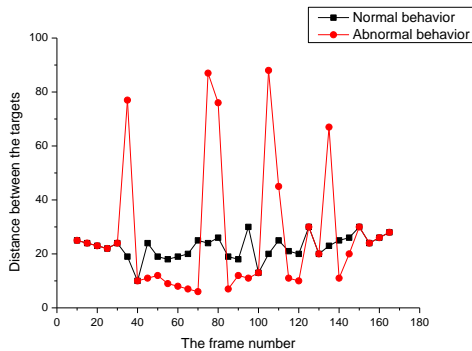
The experiment used 5 groups of positive and negative samples data, extracted the features from the 10th frame to the 165th frame in each sample. Figure 2 shows the changing trends of various kinds of characteristics under the condition of different behavior. From the 40th frame to the 140th frame, when the normal behavior happens, the states of various characteristics tend to be stable, however, when the abnormal behavior happens, the area of the target appears huge fluctuations, the length-width ratio has a dramatic changes, the distance between the connected regions spreads sometimes compact and sometimes disperse, the movement direction of the unusual behavior is very messy and there are no rules in it, there is also a drastic increase and a large variation range in kinetic energy values during this time.



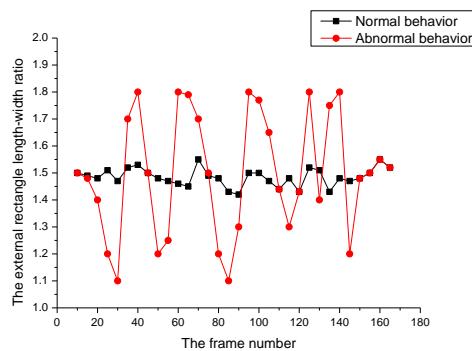
(a) The area changes



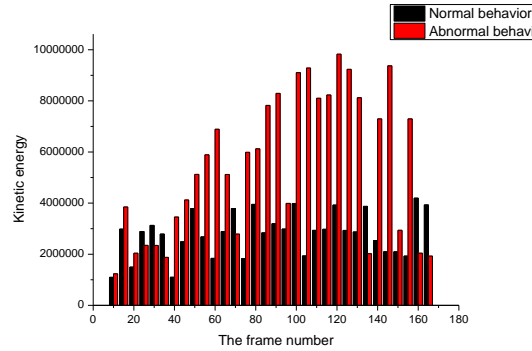
(b) The changes of target external rectangle



(c) The distance



(d) The angle difference



(e) The changes of corner kinetic energy

Figure 2. The Five Characteristics of Comparison Results between Different Behavior

As can be seen from the below images, these five groups training characteristics are from different points of view to distinguish the differences of normal behavior and abnormal behavior in the elevator effectively.

After extracting the motion features and obtaining corner kinetic energy, the given normal video sequences and analog abnormal video sequences can be used for training the SVM model by constantly adjusting the parameters. The algorithm for training and recognition of SVM is as follows:

(1)The sample data will be converted into the format which required by the SVM software tools.

(2)After extracting the movement features of human, in order to recognize conveniently, each column data of the feature vector must be normalized:

$$X = (x - \min(X)) / (\max(X) - \min(X))$$

(3)Determine the kernel function: The human motion feature extracted above is a fusion of five different characteristics, which have different behavioral attributes and belong to the non-linear separable characteristics. Thus Gaussian radial basis function is selected in this paper to convert the extracted human motion features in space, which could convert motion characteristics of human from the low-dimensional non-linear separable space to the high-dimensional linear separable space. The RBF is shown as formula (3):

$$K(x, x_i) = \exp \left\{ -\gamma \frac{\|x - x_i\|^2}{\sigma^2} \right\} \quad \gamma > 0 \quad (3)$$

(4)The optimal parameters are selected through the cross-validation method. The SVM model is obtained by training the entire training sets with the best parameters. The selections of the kernel function parameters are shown in Table 1.

Table 1. The comparison of Recognition Rate with Different Kernel Parameters

The kernel parameters	$C=0.5, \gamma=0.05$	$C=2, \gamma=0.35$	$C=4, \gamma=0.15$
Recognition rate (%)	87.15	89.35	90.15

C is the penalty coefficient, which determines the loss degree of the outliers in sample. The category information of the sample is known in the training process, it participates in training the classifier together with the training samples.

In order to verify the accuracy of algorithm in this paper, a number of experiments were conducted in the elevator platform in the laboratory, comparing with the manually setting thresholds method in literature [9] as well as Markov method in literature [10], the correct rate and processing speed are shown in Table 2.

Table 2. The Comparison Results of All Kinds of Arithmetic

Arithmetic	Number of test videos	Correct rate (%)	Process frames per unit time(fps)
Threshold comparison	100	83	27
Markov	100	89	30
Arithmetic in this paper	100	91	28

It is shown in Table 2, that the abnormal behavior recognition algorithm of this article has a significant advantage than the simple thresholds comparison algorithm and the Markov method in the higher recognition rate and processing speed.

5. Conclusion

According to the special environment in the elevator, the characteristic differences of normal behavior and abnormal behavior are analyzed in this paper. By extracting a series of features in the elevator, the modal combined a variety of motion characteristics with corner kinetic energy is established as the recognition basis of SVM. The Lucas-Kanade algorithm is used to calculate the optical flow to obtain velocity vector information and corner kinetic energy. This method not only ensures identification accuracy but also greatly reduces the data scale and accelerates the processing speed. The algorithm is with less complex calculation and higher detection rate. In future work, the author will further improve the algorithm, and apply to the other video surveillance in public places.

Acknowledgements

The paper is supported by the National Natural Science Foundation (61103149) of China.

References

- [1] C. Thureau and V. Hlavac, "Pose Primitive Based Human Action Recognition in Videos or Still Images", *Computer Vision and Pattern Recognition*, (2008), pp. 1-8.
- [2] C. Thureau, "Behavior Histograms for Action Recognition and Human Detection", *Human Motion—Understanding, Modeling, Capture and Animation*. Springer Berlin Heidelberg, (2007), pp. 299-312.
- [3] J. C. Niebles and F. F. Li, "A Hierarchical Model of Shape and Appearance for Human Action Classification", *Computer Vision and Pattern Recognition*, (2007), pp. 1-8.
- [4] J. Candamo, M. Shreve and D. B. Goldgof, "Understanding Transit Scenes: A Survey on Human Behavior-Recognition Algorithms", *Intelligent Transportation Systems*, vol. 11, no. 1, (2010), pp. 206-224.
- [5] M. Maierdan, K. Watanabe and S. Maeyama, "Human Behavior Recognition System Based on 3-Dimensional Clustering Methods", *Control, Automation and Systems*, (2013), pp. 1133-1137.
- [6] H. Bouma, G. Burghouts and L. Penning, "Recognition and Localization of Relevant Human Behavior in Videos", *SPIE Defense, Security, and Sensing. International Society for Optics and Photonics*, (2013), pp. 2-10.
- [7] N. D. Doulamis, A. S. Voulodimos and D. I. Kosmopoulos, "Enhanced Human Behavior Recognition Using HMM and Evaluative Rectification", *Proceedings of the First ACM International Workshop on Analysis and Retrieval of Tracked Events and Motion in Imagery Streams*, (2010), pp. 39-44.
- [8] T. H. Chen, T. Y. Chen and Z. X. Chen, "An Intelligent People-Flow Counting Method for Passing Through a Gate", *Robotics, Automation and Mechatronics*, (2006), pp. 1-6.
- [9] Y. P. Tang and H. F. Lu, "The Elevator Violence Prevention Based on Computer Vision Intelligent Video Surveillance", *Journal of Zhengjiang University of technology*, vol. 37, no. 6, (2009), pp.15-19.

- [10] H. J. Yuan and C. R. Wang, "Human Behavior Identification Based on a Semi-supervised Learning Method of Markov", *Journal of Computer Aided Design and Graphics*, vol. 23, no. 10, (2011), pp. 1749-1757.

