

# Detection and Analysis of Human Behavior in Video Monitoring

Dongze Li<sup>1</sup>, Liangheng Qian<sup>1</sup>, Ling Peng<sup>1</sup>, Zigui Zhu<sup>1</sup>, Peng Bai<sup>1</sup> and Xin Wang<sup>1,\*</sup>  
<sup>1</sup>College of Computer Science and Technology, Jilin University, Changchun 130012, China  
 \*Corresponding author

**Abstract**—The research content of this paper is mainly aimed at the automatic recognition of human behavior in video surveillance. It usually contains two main steps: foreground detection and behavior recognition. In the foreground detection step, this paper studied related algorithms and theories, briefly analyzed the principles and the advantages and disadvantages of various algorithms, finally choose the Vibe algorithm to extract the motive targets, and made some improvements to the original version of Vibe, which can get a more satisfied result. In behavior recognition step, we roughly classified the existing algorithms into three categories: template matching method, state space method, and semantic description method. After analyzing these three types of algorithms, we decided to select the Support Vector Machine (SVM) for training. We used the Star model for the feature collection. After the simulation experiment, the recognition result is not bad.

**Keywords**—behavior recognition; star model; vibe foreground detection; SVM

## I. INTRODUCTION

In recent years, with the rapid growth of national economy, social progress and the steady improvement of national strength, video surveillance has become a necessary facility for banks, transportation, railway stations, airports and military bases. Naturally, the demand for video surveillance is rising. But monitoring at this stage is a waste of resources and inefficient, because it relies on human resources. Moreover, in many cases, it is impossible to make timely response when an incident occurs, but only to obtain evidence afterwards. And it is a very arduous task to find the effective information in a large number of video. So intelligent video surveillance has gradually become a research hotspot. It is to use computers to process the input video stream, analyze and understand the processing results. When there is improper behavior, it will trigger a warning. This can liberate manpower, reduce the omission of events and enhance regional security.

Intelligent surveillance is also a hot technology at home and abroad. For now, there are many studies on it, and got some achievements. For example, VSAM(Visual Surveillance and Monitoring) developed by Carnegie Mellon University in the early years [1];EU's ADV ISOR system[2];IBM developed SSS(IBM developed SSS system) system[3]; Vstar system developed by Institute of Automation, Chinese Academy of Sciences[4], etc. There are also many corresponding products in the market. There are many foreign manufacturers providing intelligent video analysis software products, such as Westec, Interactive, Verint, Vidient, Visual Defence, Nextiva, Ivbox, etc. It can be seen that intelligent video surveillance has formed

an industry, and there are many technologies waiting for people to study and improve.

Although intelligent surveillance has been applied in some fields for many years, there are still many problems that need to be solved urgently. Otherwise, it will hinder the promotion of intelligent surveillance based on visual recognition technology. How to quickly extract moving objects and accurately recognize human behavior is the most urgent problem to be solved. This project focuses on these two issues, and puts forward some improvement methods on the existing methods and strategies. Fig. 1 is the flow chart of this project, in which the dashed rectangle is the method improved or newly proposed in this paper.

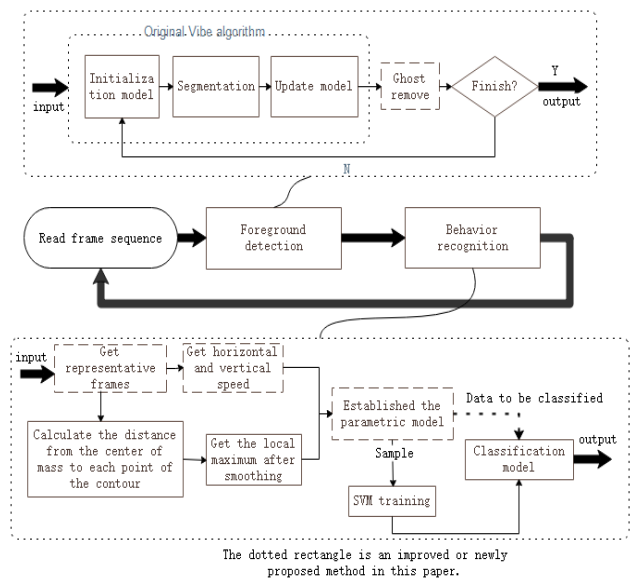


FIGURE 1. FLOW CHAT

## II. BASICS

In the field of human behavior recognition in video, we call the area of human motion as the foreground. Compared with the whole picture, foreground recognition is easier to process, so foreground extraction is necessary. There are three traditional detection methods: background elimination method[5], which is fast and easy to implement, but it is difficult to obtain an accurate background model in practical application. Optical flow method[6][7] requires a long calculation time and cannot meet the real-time requirements; The frame difference method[8], which is fast in calculation,

has large errors. Vibe algorithm[9] is based on background subtraction, the main steps are as follows:

- Initialization: Only one frame is used to initialize the background model.
- Foreground segmentation: Comparing the current pixel with the sample set corresponding to the background model, calculating the distance between the sample point and the current point, counting the numbers of the points which distance are close enough. Then segment the point according to the number reaches the threshold or not.
- Model updating: Updating background model based on segmentation results, it has certain randomness.

On this basis, this paper improves some shortcomings of the original Vibe algorithm. See the next chapter for details.

Human behavior recognition refers to the analysis and recognition of human motion state. Star model[10] is a parametric model proposed by Hironobu Fujiyoshi and Alan J. Lipton to extract a ten-dimensional vector from complex data as a feature. This model is used to reduce the computational complexity of subsequent recognition algorithms. Support Vector Machine (SVM) [11] is a classification algorithm based on state space. This paper presents a behavior recognition method based on star model and SVM.

### III. IMPROVEMENT STRATEGY

#### A. Ghost Remove-Our Improvement for Vibe

Although the original Vibe algorithm has many advantages, but since only one frame is used to initialize the model, however, since only one frame is used to initialize the model, ghosts appear in subsequent foreground segmentation : Background areas that covered by the foreground at the beginning remain as foreground after the foreground moves to other place. In order to remove ghost more quickly for a better foreground and background segmentation effect, this article improves the original vibe algorithm ,that is , the ghost can be removed after the foreground area is separated from the ghost area.

Because ghost areas are actually part of the background, they should have similar characteristics to the adjacent background areas. According to this characteristic of ghost, this article identifies and removes ghosts. The specific steps are as follows:

- For each foreground area, find out its outer rectangle, delete the foreground area in the rectangle, leave the background area as the adjacent background area of the foreground area.
- Calculate the histograms of foreground area and its adjacent background area respectively.
- Normalize the Histogram and then calculate the correlation coefficient as similarity degree.
- If the correlation coefficient is greater than the threshold value, the foreground area is determined as ghost and removed from the foreground segmentation map.

Through experiments, when the threshold value is 0.7, it works better.

#### B. Behavior Recognition

In this paper, Star model is used for behavior recognition, which classifies behavior by analyzing representative frames. The results of Konrad Schindler and Luc van Gool[12] show that in a complete human motion cycle, only 1-10 frames can be extracted to identify an action, and even one frame can achieve 90% of the discrimination effect. Therefore, it is very important to select the most representative number of frames in the action flow. This section also proposes a method for selecting the most representative frames:

$$f = \arg \min (H_t / L_f) \quad (1)$$

Among them,  $f$  is the selected frame number,  $L_f$  is the width of the human peripheral contour in a frame, and  $H_f$  is the height of the human peripheral contour in a frame. This method ensures the maximum expansion of the limbs and reduces the degree of occlusion of the limbs.

The star model has parameterized the single frame action, but discarded the process data. So the process data is compensated here. The compensation method is as follows: extracting the first and last ten frames of the representative frame, then adding the representative frame itself, at last using formula 2 to calculate the speed in pixels/frames:

$$\begin{aligned} v_x &= \frac{1}{20} \sum_{t=1}^{20} |x_c(t) - x_c(t+1)| \\ v_y &= \frac{1}{20} \sum_{t=1}^{20} |y_c(t) - y_c(t+1)| \end{aligned} \quad (2)$$

Among them,  $x_c$  and  $y_c$  are the X coordinates and Y coordinates of the center of mass of the person in frame  $t$ . Considering the distance between the person and the camera, it will lead to the same displacement in reality but different pixels in the camera. Therefore, it is more reasonable to regard the person as the displacement unit than to use the displacement of the pixel as the parameter. Formulas 3 for calculating displacement units are as follows:

$$\begin{aligned} m &= \arg \min(L_m) \quad n = \arg \min(H_n) \\ v'_x &= \frac{v_x}{L_m} \quad v'_y = \frac{v_y}{H_n} \end{aligned} \quad (3)$$

### IV. RESULT

The following figure is a comparison of the experimental results before and after the improvement of Vibe algorithm. It can be seen clearly that ghosts are eliminated.



FIGURE II. COMPARISON

The first one is before processing, the second one is the result of VIBE algorithm, and the third one is the result of ghost elimination method proposed in this paper.

In this paper, we construct a one-to-many classifier, that is, a classifier is used to screen whether a sample belongs to a certain behavior, and a total of 10 classifiers are constructed. In the simulation test, one material is selected at a time, and the remaining material is used as the sample set to train the classifier. Finally, the selected material is put into the trained classifier to estimate and obtain the estimated results. This operation is performed on each material, that is, one material is used as the test sample, the other material is used as the training set, and then a separate material is input into the classifier trained. The results are shown in Table 1.

TABLE I. CLASSIFICATION RESULT

action	Number of correct classifications
bend	8
jack	4
jump	7
pjump	7
run	6
side	5
skip	6
walk	7
wave1	7
wave2	5

The average success rate was 78%.

V. SUMMARY

In summary, this paper presents a behavior recognition method, which is based on Vibe foreground extraction algorithm and SVM classification algorithm. Based on the original vibe algorithm, this paper adds a new step, ghost elimination, and extracts a clear prospect. Then, in order to reduce the computational complexity of subsequent behavior recognition, a method of selecting representative frames is proposed. Then, based on the star model, a star model with process compensation is proposed, and feature extraction is carried out. Then SVM classification training is used to get the classification model which can detect human behavior. The experimental results show that the proposed method can achieve a satisfactory success rate of behavior recognition.

REFERENCES

[1] Collins R T, Lipton A J, Kanade T, et al. A system for video surveillance and monitoring: VSAM final report. Robotics Institute, Carnegie Mellon University: Technical Report CMU-RI-TR-00-12, 2000.

[2] Na Ylor M, Attwood C I, Annotated digital video for intelligent surveillance and optimized retrieval: final report[R] [S.I.] : ADV ISOR Conortium, 2003.

[3] Shu C F, Hampapur A, Lu M, Brown L, et al. IBM smart surveillance system (S3) :An open and extensible framework for event based surveillance//Proceedings of the IEEE International Conference on Advanced Video and Signal Based Surveillance. Como, Italy, 2005:318-323.

[4] Huang Kaiqi, Tan Tieniu. Vs-star: A visual interpretation system for visual surveillance .Pattern Recognition Letters, 2010, 31 (14):2265-2285.

[5] Elhabian Y. Shireen;;El-Sayed M. Khaled;;Ahmed H. Sumaya. Moving Object Detection in Spatial Domain using Background Removal Techniques - State-of-ArtRecent Patents on Computer Science, 2008.

[6] Bardow P, Davison A J , Leutenegger S . Simultaneous Optical Flow and Intensity Estimation from an Event Camera[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016.

[7] LIU Hong-bin, CHANG Fa-liang. Moving object detection by optical flow method based on adaptive weight coefficient, Optics and Precision Engineering,2016,24(02):460-468.

[8] Jian Zhao. The Research of Moving Target Detection Method Based on Three-frame Difference[D].Xidian University,2013.

[9] M. Van Droogenbroeck and O. Barnich.ViBe: A Disruptive Method for Background Subtraction. In T. Bouwmans, F. Porikli, B. Hoferlin, and A. Vacavant, editors, Background Modeling and Foreground Detection for Video Surveillance, chapter 7. Chapman and Hall/CRC, June 2014.

[10] Fujiiyoshi, H., Lipton, A.J.. Real-time human motion analysis by image skeletonization[P]. Applications of Computer Vision, 1998. WACV '98. Proceedings, Fourth IEEE Workshop on,1998.

[11] Suykens J A K. Support vector machines: a nonlinear modelling and control perspective.[J]. European Journal of Control, 2001, 7(2-3):311-327.

[12] Schindler, K., van Gool, L. Action snippets: How many frames does human action recognition require?[P]. Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on,2008.