

# Group Behavior Recognition in Videos based on Cam-Shift Tracking and Histogram Changing Rate

Shuang Liu<sup>a</sup>, Peng Chen<sup>b</sup>, Yanli Yu<sup>a</sup>, Xing Cui<sup>a,\*</sup>, and Denis Špelič<sup>c</sup>

<sup>a</sup>*School of Computer Science and Engineering, Dalian Minzu University, Dalian, 116605, China*

<sup>b</sup>*Department of Software Engineering, Dalian Neusoft University of Information, Dalian, 116030, China*

<sup>c</sup>*Faculty of Electrical Engineering and Computer Science, University of Maribor, Maribor, SI-2000, Slovenia*

---

## Abstract

With more and more cameras installed in public places, video surveillance systems play an increasingly important role in public safety. Research on intelligent video monitoring, especially activity recognition, is attracting increasing attention in the field of image processing. Unlike activity recognition of a single tracking object, group activity is more complex and difficult to recognize. To design a fast real-time group activity recognition algorithm without other auxiliary data, low computational cost is our focus. There are four steps for our group activity recognition system: preprocessing the captured videos, extracting foregrounds from backgrounds, tracking multiple objects and recognizing group activity. To remove noise in each frame image, the combination of the Gaussian filter algorithm and median filter algorithm is used in the preprocessing step. Then, the Gaussian mixture model is adopted to extract the foreground image. To ensure low computational cost, real-time Cam-Shift is chosen to track group activity with morphological operations in the tracking step. In the recognition step, the changing histogram rate is defined as the measure of identifying group behavior. Here, the changing histogram rate refers to the number of changing histograms and changing proportions. Experimental results show that the group activity recognition algorithm proposed in this paper is effective with low computational cost.

**Keywords:** video surveillance; group activity; Cam-Shift tracking; Gaussian mixture; histogram changing rate; filtering

(Submitted on March 11, 2018; Revised on April 24, 2018; Accepted on June 6, 2018)

© 2018 Totem Publisher, Inc. All rights reserved.

---

## 1. Introduction

With the rapid development of the economy and information technology in China, we are entering into an intelligent society. Among this, public safety is one pressing issue. Providing an early warning for dangerous accidents, even serious violent terrorist attacks such as looting and vandalism, is a primary focus. If violent accidents happen in public places, this will result in serious consequences and even casualties or property loss. With more and more cameras installed in public places, taking full advantage of these gapless 24-hour videos is very important in terms of public safety. If these videos can be processed and analyzed immediately to provide an early and real-time warning with appropriate emergency measures, injuries or more serious results for the public may be reduced or eliminated. From the perspective of public safety, we try to utilize video data captured by cameras installed in public areas and provide an early warning for abnormal or dangerous activities, providing service for public safety and social stability.

Research on intelligent video monitoring, especially activity recognition, is attracting increasing attention in the research field. In single object tracking, Arnold wrote one review paper in [1]. Compared to single object tracking and activity recognition, multiple object tracking or group activity is more difficult to detect. There are mainly three kinds of multi-object trackers, including pattern matching, object state estimation based on the Bayesian theory, and object detections [5]. There has been a fair amount of research conducted on action recognition. For the task of action recognition or classification, recent approaches use features based on optical flow [12], grammars [9], motion analysis [2,8], spatio-temporal interest points [10], or dense point trajectories extracting from regions [11], which are being tracked using optical flow across the frames. To improve processing efficiency for videos on a larger scale [13], deep learning tools such as convolutional neural networks [3,4,6,7] have been applied.

---

\* Corresponding author.

E-mail address: 390377951@qq.com

Our focus is slightly different from the methods mentioned above. Our goal is to design a fast real-time group activity recognition algorithm without other auxiliary data. Here, camera refers to a common camera widely used in our daily lives, not a professional camera. In order to run on simple computers with simple configurations, real-time with low computational cost is the primary requirement of our algorithm. With captured videos, the Gaussian filter and median filter are first called to remove noise in frame images. Then, the Gaussian mixture model algorithm is adopted to extract foregrounds. Cam-Shift tracking is then called after morphological operations for the foregrounds. Finally, the changing histogram is defined and computed to identify abnormal group behavior.

The rest of this paper is organized as follows. An overview of our whole design is given in Section 2. Based on a detailed analysis of the steps for video processing and foregrounds extraction in Section 3, a group activity recognition system based on Cam-Shift tracking and the histogram changing rate is put forward in Section 4. Some experimental results of the proposed algorithm are discussed in Section 5. The conclusions are given in Section 6.

## 2. Overview of the Proposed Recognition Flow

There are four steps for our group activity recognition system as shown in Figure 1.

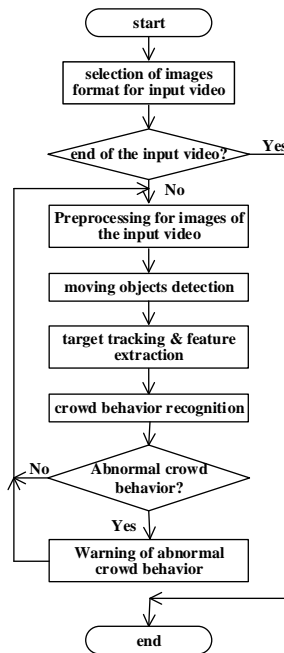


Figure 1. Flowchart of our group activity recognition system

Step 1 is preprocessing the captured videos. As the benchmark dataset published on the Internet, videos in .avi format is our video format. Different filtering methods have different advantages and disadvantages. To achieve an optimal filtering result, a combination of the Gaussian filter algorithm and median filter algorithm is adopted.

Step 2 is extracting foregrounds from backgrounds. Compared to the frame difference method, the Gaussian mixture model can be applied in many more scenes. To achieve a better recognition result, morphological operations are also adopted for the foregrounds.

Step 3 is multiple objects tracking. To get a real-time tracking result, we use Cam-Shift tracking. Color information is used in Cam-Shift tracking to identify each person as an individual, and each individual is marked with a rectangle. When there are occlusions in moving objects or the tracking object, the occlusions are considered as one whole and are bounded by one histogram rectangle.

Step 4 is recognizing group activity. Based on the obtained features from tracking methods, the changing histogram rate, which considers the number of changing histograms and changing proportions, is used to identify normal or abnormal group activity.

### 3. Preprocessing Step and Foreground Extraction

Image filtering is also called smoothing processing or fuzzy blurring. In image processing, some function of a local neighborhood of the pixels is applied on the original image to modify the pixels in an image. The goal of this is to obtain the main pattern in the data, removing noise, structural details, and instantaneous data. There are two common methods for image processing: linear filters and nonlinear filters. Nonlinear filters are more diverse and difficult to categorize, reducing noise levels without simultaneously blurring edges, such as histogram-based filters.

There are multiple options for our system: the Gaussian filter, the median filter, and the mean filter. There are some differences between these three methods. For example, the Gaussian filter is effective in removing Gaussian noise because it uses the current pixel value and the weighted average value of its surrounding pixels to replace the current value. The mean filter replaces the current pixel value with the mean. This method is also effective for Gaussian noise filtering, but the smoothing effect is not ideal when dealing with salt and pepper noise. The median filter is just the opposite. It replaces the current pixel value with a median value, ignoring the calculation of the maximum point and the minimum point, which is good for removing isolated noise. This method is not very effective for removing Gaussian noise, but it is effective for removing noise generated by pulse interference or image scanning. Because of the different calculation methods, the median filter, and Gaussian filter can be combined to complement each other. Our preprocessing step consists of these two methods of filtering to remove noise from the divided frame images.

Object detection or background subtraction is very important for detecting moving objects. Background subtraction compares current images with a background model. Anything that changes is considered to be an object. By using background subtraction, foreground objects can be detected in a frame image captured by a stationary camera. GMM is applicable to the case of light changing on the background. Because the outdoor scenes involved in this system are very complex, moving objects are detected using our improved GMM model, which is better for the latter target tracking algorithm. In the verification stage of the Cam-Shift tracking algorithm, the system uses the detection algorithm based on the Gaussian mixture model to obtain the background images of the moving objects in the video image. It then carries on the series of morphological operations and tracks the moving objects. The performance of our choice is better for recognition. Figure 2 shows the result of the moving object detected by the mixed Gaussian mixture model.

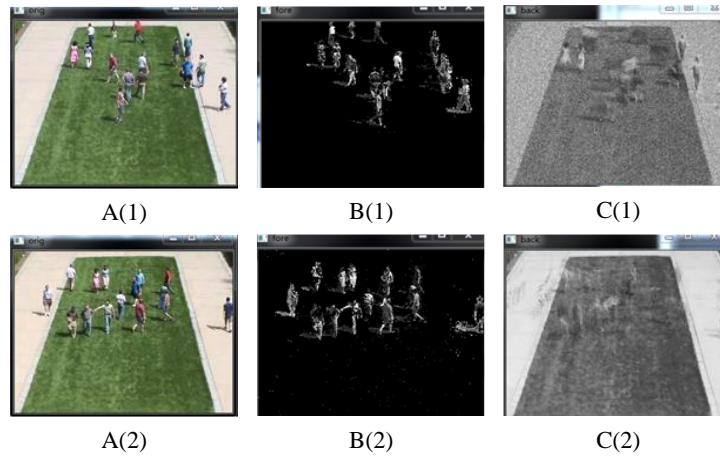


Figure 2. Moving object detected results of our improved Gaussian backgrounds mixture model

Here, A(1) and B(1) are the original frame images. B(1) and B(2) are the foreground images after adopting the Gaussian background mixture modeling, which is key for our group behavior recognition. C(1) and C(2) are the background images created by GMM. With analysis of the three filter methods and comparison of the smoothing results for the three methods, our preprocessing step consists of two methods, the median filter and Gaussian filter, to remove noise from the captured videos. To achieve an optimal recognition result, some morphological operations are also applied to the preprocessed smoothing images. For multi-objective motions in complicated scenes, the Gaussian backgrounds mixture model is adopted to detect moving objects for latter moving target tracking. The Gaussian backgrounds mixture model is a widely used model in video processing, and it will be improved in our system as an object detection tool.

### 4. Our Cam-shift Tracking Algorithm and Recognition Algorithm based on the Histogram Changing Rate

After preprocessing of the captured videos, the most important part of the tracking algorithm to find temporal and spatial features for action recognition. It is simple to track a single person or two persons. However, for group tracking, it is very

complicated because of occlusions and influences between different individuals. There are some special factors that need to be considered, such as the group density, the interception of the group and the relationship between individual behavior and overall behavior in the group. Group behavior recognition is based on features extracted from the moving target tracking. The moving target tracking method and feature extraction algorithm are directly related to the final recognition of the quality of the algorithm. Different tracking algorithms will lead to different recognition accuracy. Thus, target tracking and feature extraction are very critical steps in the group behavior recognition system. This will be discussed in the following sections.

#### 4.1. Cam-shift Tracking Algorithm

Cam-Shift (Continuously Adaptive Mean Shift) was first published by Gary Bradsky in his paper “Computer Vision Face Tracking for Use in a Perceptual User Interface” in 1988. It applies to mean shift first. Once mean shift converges, it updates the size of the window. It also calculates the orientation of the ellipse that best fits it. Again, it applies the mean shift with a new scaled search window and previous window location. The process is continued until the required accuracy is met. As an adaptation of the Mean Shift algorithm, the CamShift tracking algorithm is used for tracking regions of a video starting from any given frame throughout the next frames. It works well as an iterative convergent algorithm because of its efficiency and effectiveness in tracking targets in a sequence of video frames. The working process of our Cam-shift tracking algorithm is shown in Figure 3.

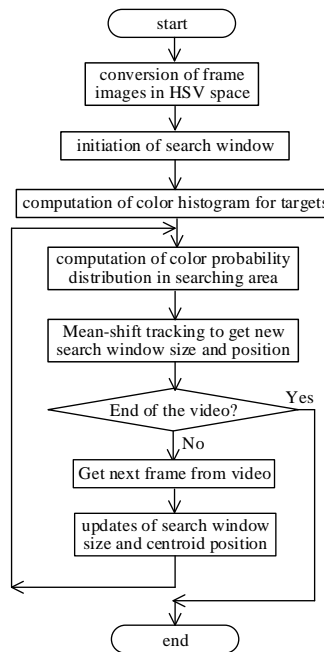


Figure 3. Flowchart of Cam-shift tracking in our system

The Cam-Shift algorithm works by matching the color histogram of the target to an image patch in each frame. That is, Cam-Shift uses color information in the video frame to find the location and size of the moving target. In the next frame of the continuous tracking, the next frame is searched by the size and position information of the current search window. The continuous tracking repeats this process. For the moving target with color information, the algorithm tracks the moving target through the color probability distribution information. Cam-Shift solves the problem of deformation of moving objects. Cam-Shift is not a demanding algorithm and is applicable for our real-time tracking with a low computational cost.

#### 4.2. Occlusion Processing

If one moving object is occluded by other objects, the contours, edges, and colors of the moving object are subject to some significant changes. There are two types of occlusions: the occlusions that occur between moving objects and the occlusions that occur between a moving object and the background. These two cases will lead to size changes of the moving targets to some extent. Here, the size of the moving object in one frame is considered as one decision rule whether occlusions exist. The size and color histograms are used to determine whether the occlusion is happening. The combination of several features to determine the presence of occlusions will avoid problems caused by the occlusions, greatly improving the performance of our recognition algorithm.

Our focus is to recognize behavior for multiple target movement tracking, considering occlusions between moving objects only and ignoring occlusions existing in the background. For occlusions that happen between moving targets, it is difficult to separate one moving target from another, and the occlusion time may be relatively long. To recognize the overall behavior for one group, two or three targets between which occlusions exist are considered to be one target. The combination of two features is used in this paper for latter recognition: the size feature and the color histogram. The size feature is computed based on area, and the color histogram is computed as Equation (1) and Equation (2).

$$M_{S_t} = \frac{S_t}{\frac{1}{n} \sum_{i=n}^{n+N} S_{t-i}} \quad (1)$$

$$M_{C_t} = \frac{\sum_{i=1}^{255} (C_t(i) - C_{t-1}(i))^2}{\sum_{i=1}^{255} C_t(i) \sum_{i=1}^{255} C_{t-1}(i)} \quad (2)$$

Here, the size is represented by  $M_{S_t}$  at time  $t$ , and the color histogram is represented by  $M_{C_t}$ . At time  $t$ , the size of the tracking area is denoted by  $S_t$ , and  $\sum_{i=n}^{n+N} S_{t-i}$  is used to represent the summation of the target area of inter-frames, utilizing the previous  $n$  frames to prevent errors caused by slow occlusions. At time  $t$ , color distribution of the gray value  $i$  is denoted as  $C_t(i)$ , and  $M_{C_t}$  is used to compare color histogram similarity of two consecutive frames. The decision rule is defined as Equation (3) when combining two features. Here,  $\tau_c$  is one defined threshold value with the range from 0.21 to 0.33.

$$M_t = M_{S_t} \times M_{C_t} \begin{cases} \text{occlusion happens, } M_t > \tau_c \\ \text{no occlusion, } M_t \leq \tau_c \end{cases} \quad (3)$$

#### 4.3. Recognition based on the Histogram Changing Rate

Our focus in this paper is to study group behavior based on multi-objects tracking. It is very difficult to obtain an accurate recognition result because of group density and complex scenes. Too many additional conditions will lead to complications and slow processing for the detection and tracking algorithms, so we try to use a common computer with a Windows operating system and ordinary camera without other conditions to achieve good real-time tracking. According to the common process, using a simple algorithm for detection and real-time tracking may not be satisfactory. Thus, the difficulty of our research in this paper is to determine what kind of algorithm to use and what kind of method to combine these algorithms to achieve an optimal tracking result with relatively low time-consuming cost. Through experimental verification, our choice is to detect the moving targets based on the Gaussian background mixture models and get better contours of the targets by adopting some morphological operations for foregrounds. Morphological operations also help to eliminate the adhesion of the moving target in the small area of the foreground image to set a good foundation for the subsequent tracking algorithm. Then, the real-time Cam-Shift tracking algorithm is used to track the moving objects. Because the group behavior may be obscured between certain moving objects, several targets are considered as one whole target if occlusion happens. Compared to the pyramid Lucas-Kanade algorithm, the Cam-Shift tracking algorithm does better in terms of real-time tracking performance, especially when tracking large amounts of consecutive videos. In the final group behavior recognition, histogram changes computed by the Cam-Shift algorithm are key factors, which is the number of changes of the histogram between frames in the video. Measuring speed changes and proportion changes in the histogram is used to determine whether the group behavior is normal or abnormal behavior. A flowchart of this system is shown in Figure 4.

This system uses the number of changes of the histogram to identify the behavior of the group. That is, the number of histograms in the current frame is compared with the number of histograms in the previous frame. Equation (4) is used to calculate whether the number of histograms entering the current frame is increased or decreased and whether the change rate of the histogram is increasing at a constant speed, suddenly increasing, decreasing at a constant speed, or suddenly decreasing. Group behavior is measured based on these two features as Equation (4) and Equation (5).

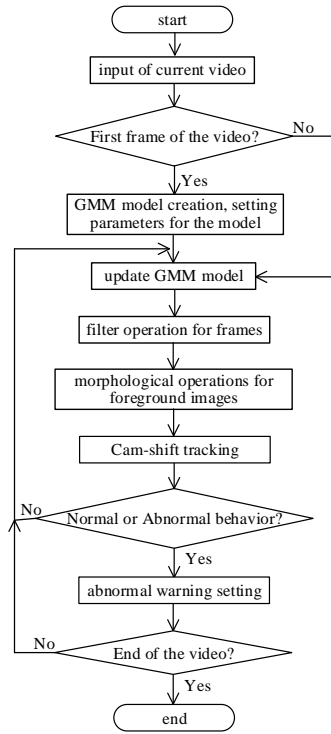


Figure 4. Flowchart of our system

$$\rho = \frac{W_{s_t}}{\frac{1}{n} \sum_{i=t-n}^t W_{s_i}} \quad (4)$$

$$\chi = \frac{|W_s|}{n} \quad (5)$$

Because of the multi-targets tracking in the group, it is possible to lose one tracking object by keeping two or three obscured objects as one object with one marking box to track one individual. Therefore, in the above equation, the average value of the number of histograms of the preceding  $n$  frames from the current time  $t$  is computed.  $W_{s_t}$  represents the tracking of the histogram box at current time  $t$ , and  $W_s$  represents the changing tracking box when there are new objects running into or out of the tracking box, that is,  $W_s = \sum_{i=t-n}^t (W_{s_i} - W_{s_{i-1}})$ . If  $W_s$  is a positive number, it indicates that the number of people in the frame of the video increases, and further decision is computed based on the change rate of the tracking box and the threshold. If  $W_s$  is negative, it indicates that the number of people in the frame of the video decreases, and further computation is completed to determine the escape event-based change rate of the tracking box and threshold.  $\tau_h$  is the high threshold for judging the existence ratio of the histogram, and  $\tau_l$  is the low threshold for judging the existence rate of the histogram.  $T_h$  is the fast threshold in the histogram change speed measure, and  $T_l$  is the slow threshold in the histogram change speed measure. Here,  $\tau_h$  and  $T_l$  thresholds are used to track the behavior of the population in the Time 14-5 video in PETS09, and  $\tau_l$ ,  $T_h$  are calculated using the UMN data video source 1.

## 5. Experimental Discussions

Applications of the Cam-Shift tracking result is shown in Figure 5 for the UMN data set and PETS2009 data set. The occlusion result is shown as (a). In the left part of frame image (a), occlusions exist between two persons. Our algorithm considers two persons as one object, marking them with one tracking box. Frame images (b), (c) and (d) are Cam-shift tracking results for different scenes.

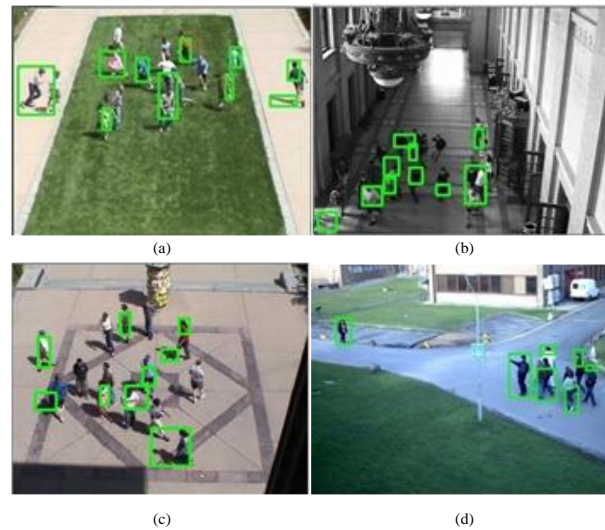


Figure 5. Cam-shift tracking results

Applications of the histogram change rate to identify the group behavior is shown in Figure 6.

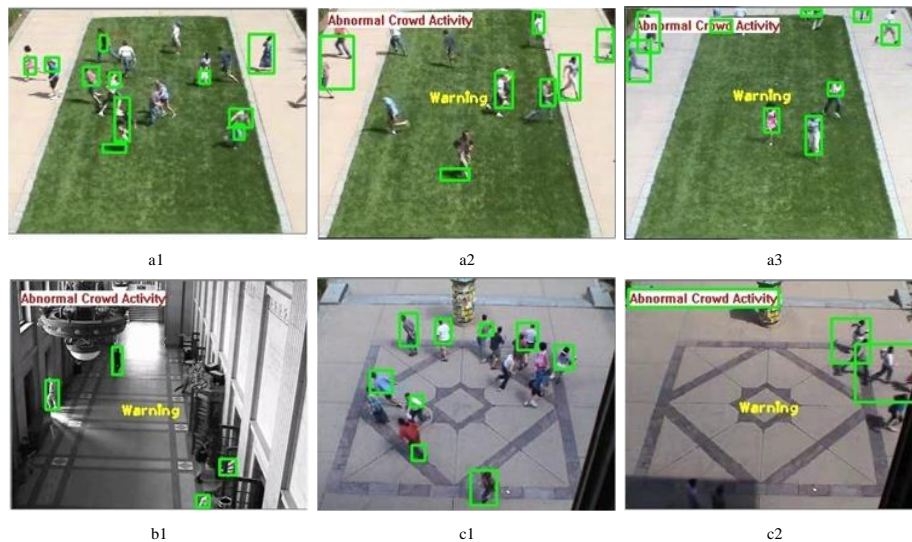


Figure 6. Recognition result for PETS09 data set

Because group behavior recognition is based on the histogram change rate, the histogram is not changed at the right moment when the group is about to change. Thus, with the frame that is currently changing as the starting frame, the fifth frame thereafter detects the alarm information. As shown in Figure 6, a1 is the second frame after the group behavior changes. No alarm information is detected because the histogram has not changed significantly in the current situation. a2 and a3 in Figure 6 detect the alarm information because there are some persons in the video running out of the frame image, resulting in the number of histogram and frequency changes of the histogram reducing or increasing. In this case, the alarm behavior information is detected. Similarly, c1 in Figure 6 does not detect the alarm information because the histogram has not changed, while c2 detects the alarm information because the histogram changes.

Now, some experimental results in data format are given to verify our algorithm based on the UMN data set and PETS2009 data set. For the UMN data set, one comprehensive video is downloaded from the website where the UMN data set is located, including three small videos in different environments. One video modification software is adopted to divide the source video into three small scenes videos. Then, experimental verification is completed on the three videos. Experimental results for our abnormal behavior recognition algorithm in the video is shown in Table 1. Results for the pyramid Lucas-Kanade algorithm is shown in Table 2 for the same data.



Table 1. Results of our algorithm

experimental video	Total number of video frames	Test the number of abnormal frames	Actual number of abnormal frames	Actual normal frame number	Test accuracy
video1	597	111	128	469	86.72%
video2	409	121	134	275	90.30%
video3	623	91	105	518	86.67%

Table 2. Results of the pyramid Lucas-Kanade algorithm

experimental video	Total number of video frames	Test the number of abnormal frames	Actual number of abnormal frames	Actual normal frame number	Test accuracy
video1	597	107	128	469	83.59%
video2	409	110	134	275	82.01%
video3	623	85	105	518	80.95%

Similar to the previous steps, the PETS2009 data set with the frame-by-frame collection is downloaded, including S1L1, S1L2, ..., S3HL and other video sets. Each video set is divided into several small video sets. We use small videos of different sports situations in three different scenes. Table 3 shows the statistical results of our Cam-Shift tracking combining histogram changing rate method. Table 4 shows the statistical results of the pyramid L-K energy method. It can be seen from Table 2 and Table 4 that the pyramid Lucas-Kanade algorithm uses the corners as the feature points to calculate the energy information of the group, utilizing the population energy to determine the accuracy of the group behavior. This is consistent with the Cam-Shift tracking algorithm using the histogram to recognize the group behavior.

Table 3. Results of our algorithm

experimental video	Total number of video frames	Test the number of abnormal frames	Actual number of abnormal frames	Actual normal frame number	Test accuracy
S1_L1T1357	220	0	0	220	100%
S1_L1_V001	89	31	34	55	91.18%
S3_HLT1416	222	110	130	92	84.62%

Table 4. Results of the pyramid Lucas-Kanade algorithm

experimental video	Total number of video frames	Test the number of abnormal frames	Actual number of abnormal frames	Actual normal frame number	Test accuracy
S1_L1T1357	220	3	0	220	-
S1_L1_V001	89	29	34	55	79.41%
S3_HLT1416	222	103	130	92	79.23%

However, for real-time tracking, abnormal group behavior recognition based on the Cam-Shift algorithm tracking and changing of the histogram is faster. Regardless of the type of video, the Cam-Shift tracking algorithm and the histogram changing rate recognition algorithm proposed in this paper are relatively stable. Therefore, the recognition algorithm based on our Cam-Shift tracking algorithm has the advantage of large-scale video processing, as shown in Table 5.

Table 5. Comparison results of processing speed

experimental video	L-K processing frames / s	Our Cam-Shift processing frames/s	Cam-Shift / L-K ratio
video 1	4.95	28	5.66
video 2	4.95	28	5.66
video 3	4.95	28	5.66
S1_L1T1357	4.91	28	5.73
S1_L1_V001	4.91	28	5.73
S3_HLT1416	4.91	28	5.73

## 6. Conclusions

Group behaviour recognition is very important for intelligence surveillance. Our goal is to design a fast real-time group activity recognition algorithm without other auxiliary data. In order to run on simple computers with simple configurations, real-time with low computational cost is the primary requirement of our algorithm. With captured videos, the Gaussian filter and median filter are first called to remove noise in frame images. Then, the Gaussian mixture model algorithm is adopted to extract foregrounds. Morphological operation is applied on the foregrounds to achieve optimal results. Because group behaviour may be obscured between certain moving objects, several targets are considered as one whole target if occlusion happens. Compared to the pyramid Lucas-Kanade algorithm, the Cam-Shift tracking algorithm performs better in real-time tracking, especially when tracking large amounts of consecutive videos. Measuring speed changes and proportion changes in the histogram is used to determine whether the group behaviour is normal or abnormal behaviour. Experimental results validated our group behaviour recognition algorithm.



## Acknowledgments

This work is partially supported by the National Nature Science Foundation of Liaoning Province with grant No. 2015020099 and National Natural Science Foundation with grant No.71303031.

## References

1. W. M. S. Arnold, M. Chu Dung, Cucchiara Rita, etc., "Visual Tracking: an Experimental Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.36, no.7, pp.1442-1468, 2014
2. G. T. Bae, S. Y. Kwak, H. R. Byun, "Motion Pattern Analysis Using Partial Trajectories for Abnormal Movement Detection in Crowded Scenes," *Electronics Letters*, vol. 49, no.3, pp.186-187, 2013
3. J. Donahue, L. A. Hendricks, S. Guadarrama, etc., "Long-term Recurrent Convolutional Networks for Visual Recognition and Description," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2625-2634, 2015
4. A. Elhayek, E. de Aguiar, A. Jain, etc., "Efficient ConvNet-based Marker-less Motion Capture in General Scenes with a Low Number of Cameras," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp.3810-3818, 2015
5. R. Girshick, J. Donahue, T. Darrell, etc., "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580-587, 2014
6. A. Karpathy, G. Toderici, S. Shetty, etc., "Large-Scale Video Classification with Convolutional Neural Networks," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1725-1732, 2014
7. S. Liu, X. D. Liang, L. Q. Liu, etc., "Matching-CNN Meets KNN: Quasi-Parametric Human Parsing," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1419-1427, 2015
8. B. B. Ni, P. Moulin, X. K. Yang, etc., "Motion Part Regularization: Improving Action Recognition via Trajectory Group Selection," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3698-3706, 2015
9. H. Pirsiavash, D. Ramanan, "Parsing Videos of Actions with Segmental Grammars," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 612-619, 2014
10. M. J. Roshtkhari, M. D. Levine, "An On-line, Real-time Learning Method for Detecting Anomalies in Videos Using Spatio-temporal Compositions," *Computer Vision and Image Understanding*, vol.117, no.10, pp.1436-1452, 2013
11. G. Shu, A. Dehghan, O. Oreifej, etc., "Part-based Multiple-person Tracking with Partial Occlusion Handling," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1815-1821, 2012
12. T. Wang, H. Snoussi, F. Smach, "Detection of Visual Abnormal Events via One-class SVM," *Proceedings of the 2012 International Conference on Image Processing, Computer Vision, and Pattern Recognition*, pp. 113-119, 2012
13. X. Y. Wang, Q. Ji, "Video Event Recognition with Deep Hierarchical Context Model," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4418-4427, 2015

**Shuang Liu** received her Ph.D. in Traffic Information Engineering and Control from Dalian Maritime University, China. She is currently an associate professor at Dalian Minzu University, China.

**Peng Chen** is currently a professor at Dalian Neusoft University of Information, China.

**Yanli Yu** is currently a researcher at Dalian Minzu University, China.

**Xing Cui** is currently a postgraduate at Dalian Minzu University, China.

**Denis Špelič** obtained his Ph.D. in Computer Science in 2011 from Maribor University. He is currently a researcher within the Faculty of Electrical Engineering and Computer Science at the University of Maribor, Slovenia.