The 12th International Conference on
Computer Science & Education (ICCSE 2017)
August 22-25, 2017. University of Houston, USA

Real-Time Detection Algorithm of Abnormal Behavior in Crowds Based on Gaussian Mixture Model

Zhaohui Luo[1],Weisheng He[1],Minghui Liwang[1],Lianfen Huang[1]*,Yifeng Zhao[1]
School of information science and engineering,
Xiamen University,
Xiamen, Fujian, China[1]
lfhuang@xmu.edu.cn

Jun Geng[2]
School of Information and Communication Engineering
Beijing University of Posts and Telecommunications
Beijing, China[2]

***Abstract*—Recently, abnormal evens detection in crowds has received considerable attention in the field of public safety. Most existing studies do not account for the processing time and the continuity of abnormal behavior characteristics. In this paper, we present a new motion feature descriptor, called the sensitive movement point (SMP). Gaussian Mixture Model (GMM) is used for modeling the abnormal crowd behavior with full consideration of the characteristics of crowd abnormal behavior. First, we analyze the video with GMM, to extract sensitive movement point in certain speed by setting update threshold value of GMM. Then, analyze the sensitive movement point of video frame with temporal and spatial modeling. Identify abnormal behavior through the analysis of mutation duration occurs in temporal and spatial model, and the density, distribution and mutative acceleration of sensitive movement point in blocks. The algorithm can be implemented with automatic adapt to environmental change and online learning, without tracking individuals of crowd and large scale training in detection process. Experiments involving the UMN datasets and the videos taken by us show that the proposed algorithm can real-time effectively identify various types of anomalies and that the recognition results and processing time are better than existing algorithms.**

***Index Terms*—video monitoring; crowd analysis; anomaly detection; online learning**

I. Introduction

Nowadays, in order to guarantee social safety, video surveillance systems whose hardware cost is low, have been widely applied in public places, such as airports, subway stations, traffic junctions and schools etc. However, most of the existing video surveillance systems are just used as a record systems, which can not detect and analyze an abnormal event automatically, and it is impossible for a person to watch the monitor at all times[1]. Because the monitoring facilities are typically mounted in public places, abnormal evens detection in crowds has been critical and has become a new area of interest in the intelligent monitoring system research community; the results could potentially be used for public security purposes[2]. On the other hand, traditional methods for individual behavior analysis cannot be used for crowded scenarios because of the occlusion phenomenon. Video surveillance in crowded areas is becoming increasingly significant for public security.

Recently, lots of efforts have been focused on intelligent surveillance video systems. In particular, abnormal event detection in crowded scenarios is a significant and challenging task in intelligent surveillance video systems. Abnormal event detection refers to detecting and responding to the abnormal changes or behaviors of humans or objects in videos. Currently, there are various abnormal detection algorithms proposed in crowded scenes, such as Mehran et al. [3] proposes the social force model, which uses interaction force between pedestrians as feature for abnormal event detection, and the result is good. Wang et al. [4] used an optical flow histogram (HOF) to express the statistics information of the motion velocity and direction. Cong et al. [5] proposed multi-scale HOF (MHOF) to express the statistics characteristics of the optical flow, and then they put forward using sparse representation to detect abnormal crowd behavior. These method has higher recognition accuracy, but it takes a lot of time to calculate optical flow and analyze statistic, it can't achieve Real-time monitoring. However, nowadays reasonable project is to achieve the real-time monitoring with the camera, the previous study can not achieve real-time monitoring will lose the practicability. In order to deal with the bottleneck, we use the Gaussian Mixture Model (GMM) to extract sensitive movement points as a feature description[6].

In this paper, we propose a new method of feature description. Base on getting foreground points by GMM, we can set different update threshold value of GMM, quantify the velocity of foreground points, then we can turn to analyze the number and distribution characteristics of the foreground points in certain speed. This is also the first time to create links among the definition of abnormal behavior and the concept of GMM. We define the abnormal behavior as a time varying and unstable state, and define the normal behavior as a stable behavior in time. In this paper, we apply the GMM to analyze crowd anomaly, the experimental results show that compared with the traditional classification method, this method has the advantages of high speed, high recognition rate and stable recognition affection. Experiments on different datasets and videos show that our method can effectively identify the abnormal behavior in a crowd scene and can even distinguish the types of anomalies, and the recognition results are better than those of existing algorithms.

The remaining of this paper is organized as follows. In Section II, we introduce the new feature descriptor of motion and show how to extract the sensitive movement point from the frames. We describe the method and process of behavior

978-1-5090-2508-4/17/$31.00 ©2017 IEEE

modeling in Section III. We classify the normal and abnormal behaviors and the types of abnormal behaviors in Section IV. The experiments and results are given in section V.

II. Feature Extraction

For dynamic feature description method involve expressing the object by direct extracting its motion information, and the most effective is the speed characteristic. Stauffer et al put forward Gaussian Mixture Model (GMM)[6].Based on the persistence and the variance of each of the Gaussians of the mixture, they determine which Gaussians may correspond to background colors. Pixel values that do not fit the background distributions are considered foreground until there is a Gaussian that includes them with sufficient, consistent evidence supporting it. In other words, GMM can remove repeated targets from the scene and detect the mutational target.

Also, the recently study shows that abnormal behavior often appears as objects suddenly moving rapidly such as people suddenly rush or a car passed quickly through the crowd, or appears as objects suddenly accelerate/decelerate such as people scatter in all directions, mass brawl and so on. This means the abnormal behaviors often appear with the abrupt change of movement velocity. In this paper, we first use the GMM method to analyze the surveillance video and detect the location of the pixels that do not appear frequently. Then we analyze the temporal and spatial motion information of abnormal behavior by modeling the abnormal behavior. At last, we can achieve real time abnormal behavior detection.

A. Mixed Gaussian Model

The recent history of each pixel is modeled by a mixture of *K* Gaussian distributions. The probability of observing the current pixel value is：

$$p(x_t) = \sum_{j=1}^{K} \pi_{j,t} N(x_t | u_{j,t}, \Sigma_{j,t}) \quad (1)$$

where x_t is the point's pixel value at time t, K is the number of models, $\pi_{j,\mathrm{t}}$ is the j-th Gaussian model's weight, $u_{j,\mathrm{t}}$ is the mean value of the *j*-th Gaussian in the mixture at time t, $\Sigma_{j,t}$ is the covariance matrix of the *j*-th Gaussian in the mixture at time *t*, and $N(x_t|u_{j,\mathrm{t}}, \Sigma_{j,\mathrm{t}})$ is *j* Gaussian distributions' probability density function at time *t*:

$$N(x_t|u_{j,\mathrm{t}}, \Sigma_{j,\mathrm{t}}) = \frac{1}{(2\pi)^{\frac{K}{2}} |\Sigma_{j,\mathrm{t}}|^{\frac{1}{2}}} e^{-\frac{1}{2}(x_t - u_{j,\mathrm{t}})^T \Sigma_{j,\mathrm{t}}^{-1} (x_t - u_{j,\mathrm{t}})} \quad (2)$$

To calculate the $p(x_t)$, we use GMM to extract foreground points, and judge whether its value is larger than a certain threshold. If it's larger, then we can judge the point can match current GMM, and this point is the background point, we can use this pixel to update the parameters of each Gaussian model[7]. If it isn't larger, we think this point can't match current model, we should delete the Gaussian model which occupies the smallest part and reinstitute a new Gaussian model. When the point matches the GMM, model's update includes weight update, mean update and variance update. Weight update formula can be expression as follow：

$$\pi_{j,\mathrm{t}} = (1 - \alpha)\pi_{j,\mathrm{t}-1} + \alpha(M_{j,\mathrm{t}}) \quad (3)$$

where α is the update rate of GMM, $M_{j,\mathrm{t}}$ is 1 when the point can match *j*-th Gaussian model, or is 0. Mean update formula can be expression as follow：

$$u_{j,\mathrm{t}} = (1 - \rho)u_{j,\mathrm{t}-1} + \rho(x_t) \quad (4)$$

Covariance update formula can be expression as follow：

$$\Sigma_{j,\mathrm{t}} = (1 - \rho)\Sigma_{j,\mathrm{t}-1} + \rho((x_t - u_{j,\mathrm{t}})^T (x_t - u_{j,\mathrm{t}})) \quad (5)$$

Update factor can be expression as follow：

$$\rho = \alpha N(x_t | u_{j,t}, \Sigma_{j,t}) \quad (6)$$

Judging from the formulas above, the value of update factor determine how fast can GMM rebuild a new model, If it's large, the new pixel value can be used to rebuild the new model by GMM quickly. In a large update rate, slowly changing objects will be modeled as the background by quickly updating GMM, as a result, only points can reach a certain velocity can be extracted as the foreground points. This means that the update rate of the GMM can be used to measure the velocity of the moving point[6].

B. Mixed Gaussian Model's Update Rate

In order to quantify the speed of crowd behavior, we should set the update rate α of GMM accurately, but the speed of crowd behavior is unknown, it's hard to set a stable update rate to quantify the speed of crowd behavior in all scenes. We use adaptive learning rate algorithm, first assume that the first 50 images of the video are normal behavior samples, we set it as the initialization video, then we use several GMM with different update rates to extract foreground points and pick out one GMM whose foreground points' quantity is closest to expected quantity to analyze current scene. In other words, when the quantity of foreground points is closest to the expected quantity, this model's update rate is similar to crowd's normal velocity, and if we use this update rate to analyze the crowd behavior, the quantity of foreground points we extract is almost stable. The formula of adaptive algorithm is as below (7):

$$\alpha = arg\min_{\alpha} |V_{\alpha i} - T_{GMM}| \quad (7)$$

$V_{\alpha i}$ is the quantity of SMPs we extract with the i-th update rate.

C. Extraction of sensitive moving points

After the adaptive selection of GMM update rates, we can get the suitable GMM in the current scene, and we can use it to analyze subsequent video frames and extract SMPs, we use V_{all} to stand for the quantity of SMPs (it has been transformed into a grayscale point). In other words, it's a faster moving point with respect to the current GMM, its value reflects the degree of significant movement. For example, the grayscale point areas are the sensitive motion areas in this video frame, and they remain as salient visual regions, as shown in Fig.1.

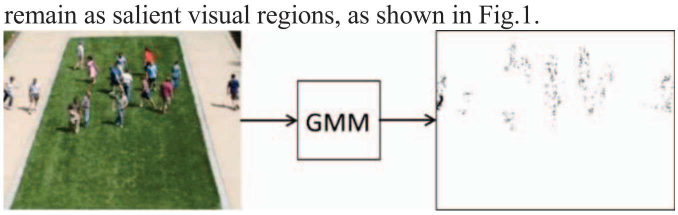

Fig.1. Extraction of SMPs by Gaussian Mixture Model

III. Crowd Abnormal Behavior Detection

Through the analysis, we can find that the abnormal behavior often has the following two main characteristics: one is the mutation of behavior and the other one is persistence of behavior time [11]. In this paper, we use the frame to detect the mutation of behavior, and set up the temporal and spatial model of abnormal behavior to analyze the temporal and spatial characteristics of behavior, as a result we can judge the crowd anomaly successfully.

A. Frames Detection

In our model, when the update rate α matches the normal crowd behavior, the number of SMP of the normal video frame is approaching a constant, otherwise, the number of SMP Greater than the constant, namely it is an abnormal frame. However, the crowd behavior is changing, it's unreasonable that only set a fixed threshold to analyze the quantity of SMPs is stable or not. In this paper, we calculate the normal behavior frames' average motion degree V_{mu}, and take it as the standard to determine the degree of current frame's movement, at the same time, we set an adaptive judgment threshold T. When it is detected that the pixel of the video frame exceeds the threshold value T, it is considered that it exceeds the range of the normal behavioral motion velocity, and when the number of SMP is greater than the threshold T, we can judge this frame is an abnormal frame. But whether it is an abnormal behavior that should be further analyzed, because a behavior is composed of multiple consecutive frames. If it is observed that the current velocity is within the range of normal behavior fluctuation, it is considered to be a normal frame and the threshold T is updated online using the value of current frames' pixels. The update formula of T is as below:

$$T = \gamma * V_{mu} + C \tag{8}$$

where C is a constant, γ is a constant variable.

$$V_{mu} = V_{mu} + \beta * (V_{all} - V_{mu}) \tag{9}$$

where β is the frame detection threshold T's online update rate.

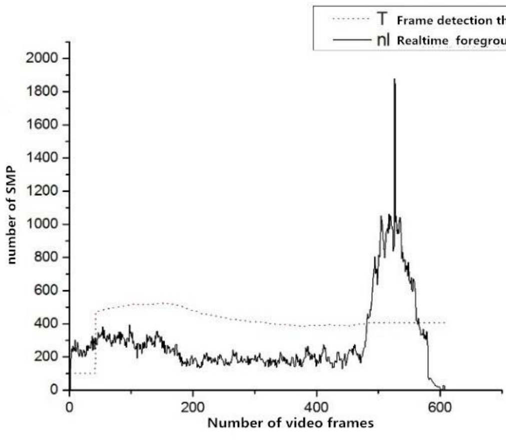

Fig.2. The Frame Detection in video frames

Fig.2 shows the frame detection results on the a fraction of UMN dataset.

B. Spatio-temporal Detection

Firstly, we use the GMM method to detect abnormal frame, once detect the first abnormal frame we should set a space stack which can store a continuous abnormal behavior and estimate the duration of abnormal behavior, then we push this first abnormal frame into the space stack and begin to count the abnormal frame for a continuous time[14] (i.e. video frame rate Fs is known， so that the duration of the behavior can be obtained from the frame rate), if the number of abnormal frames exceeds one frame rate *Fs* (for example, if the current frame rate is 25, the number of consecutive abnormal frames exceeds 25), this indicates that the abnormal behavior lasts for 1 second, and the behavior is judged to be an abnormal behavior and continues to push the subsequent abnormal frame into the abnormal space-time stack until the abnormal behavior terminates.

Set $\Lambda_t{}^k = \{(x,y)_t, P_{x,y} = 1\}$ represents the coordinates of the t-th SMP of the k-th frame in the spatio-temporal model; only when a pixel point (x,y) is a SMP, then $P_{x,y} = 1$, otherwise it is 0. Fig.3 shows the spatio-temporal model.

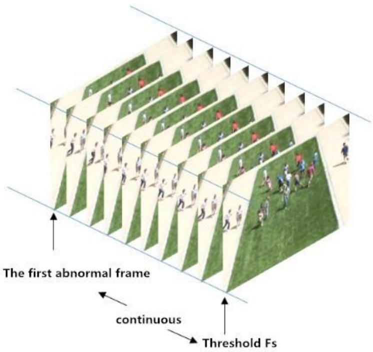

Fig.3. Space-time diagram

IV. Abnormal Behavior Classification

In this paper, we combine frame detection and Space - time modeling to detect crowd anomaly and the types of abnormal bevavior. In order to realize the recognition of abnormal behavior categories, this paper analyzes the spatio-temporal model of abnormal behavior, and realizes the recognition of abnormal categories by analyzing the change law of sensitive moving points in the spatio-temporal model.

A. Calculation of the anomaly's starting point

In the previous section, the anomaly's spatio-temporal model we extract contains all abnormal behavior which are from the beginning to the end of all the frames, and the first frame of the spatio-temporal model also represents the first frame of the abnormal behavior. So, we only need to calculate the center point of sensitive moving points in the first frame, and this point is the anomaly's starting center position $P_s\,(\bar{x},\bar{y})$

$$P_s\,(\bar{x},\bar{y}) \rightarrow \begin{cases} \bar{x} = \sum_{t=1}^{V_{mu}^1} x_t\,, \ \ x_t \in \mathbf{\Lambda}^1 \\ \bar{y} = \sum_{t=1}^{V_{mu}^1} y_t\,, \ \ y_t \in \mathbf{\Lambda}^1 \end{cases} \tag{10}$$

where $\mathbf{\Lambda}^1$ is the set of the first frame's SMPs, V_{mu}^1 is the quantity of first abnormal frame's foreground points, x_t ,

y_t represent the x, y coordinate values of the t-point in the foreground point set $\boldsymbol{\Lambda}^1$.

B. The Mean Distance of Abnormal frame's SMPs

In order to analyze the trend of abnormal behavior in spatio-temporal model, it is necessary to calculate the distance of each frame relative to the starting center, D^k represents the average distance of the k-th frame from the starting center $P_s\ (\bar{x}, \bar{y})$.

$$D^k = \frac{\sum_{t=1}^{v_{mu}^k} \sqrt{(x_t - \bar{x})^2 + (y_t - \bar{y})^2}}{v_{mu}^k} \tag{11}$$

C. Analysis of Abnormal Behavior Dispersion

The distance of the scattered behavior relative to the center point changes rapidly, and the change of the gang fights behavior relative to the center point is relatively slow. Analyze the dispersion S of the spatio-temporal model's distance set D^k, if S is large then the distance change is rapidly and it's scattered behavior, if S is small then the distance change is slow and it's gang fights behavior.

$$S = \sqrt{\frac{\sum_{k=2}^{Fn} (D^k - \bar{D})^2}{Fn-1}}, \tag{12}$$

where S is the distance dispersion of the abnormal behavior, Fn is the total number of frames of the spatio-temporal model. $\bar{D} = \sum_{k=1}^{Fn} D^k / Fn$ is the average distance of the spatio-temporal model.

V. Experimental Results

In order to verify the effectiveness of the algorithm, we experiment with the algorithm under VS2013 and OPENCV2.4.8, and test the performance of the algorithm. Experiment is acted in windows 10 64bit system platform, the computer configuration is: Intel (R) Xeon (R) E3 CPU - 1230 V2 3.3GHz, 16.00GB memory. In our experiments, two datasets are used to test the proposed algorithm: University of Minnesota (UMN) Dataset and the crowd abnormal video made by us. The UMN dataset is the most widely used test dataset for the research area of the crowd abnormal behavior detection.

The receiver operating characteristic curve (ROC) is used to represent the recognition effect at different thresholds. The ROC curve consists of true positive rate (TPR) and false positive rate (FPR). The ROC space defines the pseudo-normal rate (FPR) as the X-axis, and the true class rate (TPR) is defined as the Y-axis. In this paper, true positive (TP) refers to the abnormal frames that are detected correctly, whereas fault positive (FP) refers to the normal frames that are labeled as being abnormal. True negative (TN) refers to the normal frames that are detected correctly, and fault negative (FN) refers to the abnormal frames that are labeled as normal.

TPR: the rate of the abnormal frames that are detected correctly.

$$TPR = \frac{TP}{TP+FN} \tag{13}$$

FPR: the rate of the normal frames that are labeled as abnormal.

$$FPR = \frac{FP}{FP+TN} \tag{14}$$

A. Abnormal Behavior Detection of UMN:

UMN Dataset[9] has 3 different scenes, totaling 11 videos with a resolution of 240 * 320. Scene 1 contains 1450 frames, scene 2 contains 4415 frames, and scene 3 contains 2145 frames. Figure 4 shows the normal walking events and abnormal run events detected by this algorithm

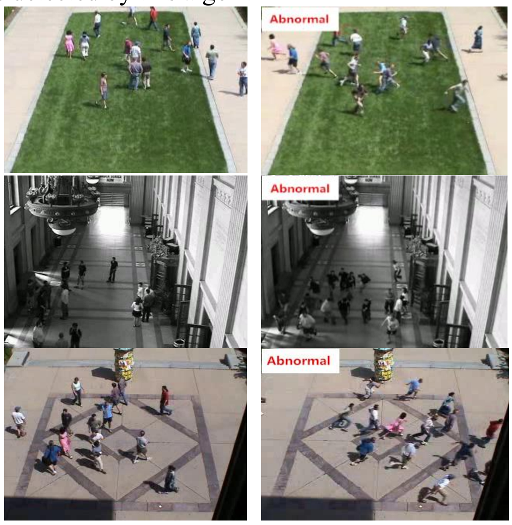

Fig.4. The normal and anomalous figure detected by UMN data set

Fig.4 shows the testing results on the UMN dataset[9] for three different Scenes.

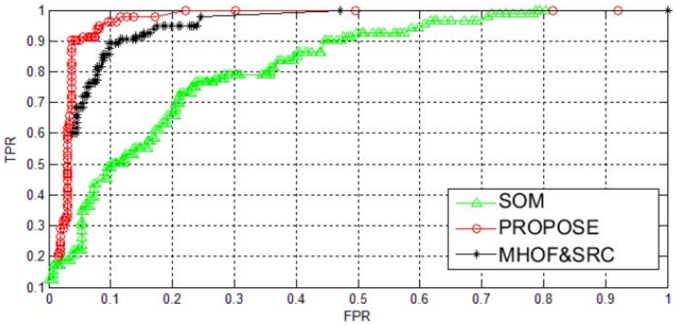

Fig.5. Experiment results for the UMN dataset

Fig.5 shows the testing results on the UMN dataset for three different methods: the SOM[8], the MHOF and SRC [5], and the proposed method. The TPR and FPR for the three different methods are shown in TABLE I.

TABLE I. THE EXPERIMENTAL RESULT OF UMN DATA SET

Abnormal detection method	FPR	TPR	operation time
SOM[8]	0.1	0.51	1.61927s
MHOF&SRC[5]	0.1	0.89	2.87934s
PROPOSE	0.1	0.98	0.050696s

Experimental results show that this method can effectively identify a variety of abnormal events (such as running, scattered in the crowd, etc.). The SOM algorithm needs to extract the image optical flow characteristics, consuming a lot of time. MHOF & SRC algorithm not only needs to extract the optical flow characteristics of video frames, but also requires sparse approximation and calculate sparse reconstruction cost. Therefore, although the algorithm has high recognition rate, it takes too much time. Our method does not need to extract the optical flow characteristics, which greatly simplifies the time of feature extraction. The experiment proves that the method can detect at a faster speed.

B. Abnormal Behavior Type Recognition

The experiment uses UMN data sets and homemade video to complete, homemade video has 6 video, which contains three scattered (4117 frames) and three groups of gangs (3982 frames) abnormal behavior. Resolution is 180 * 320.

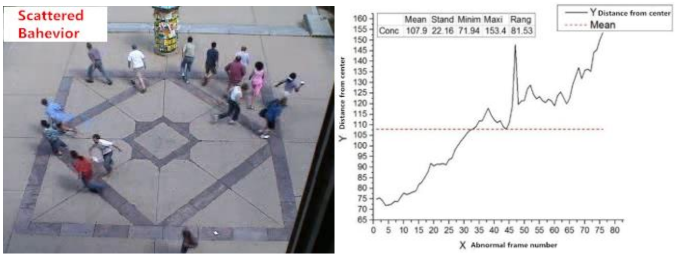

Fig.6(a) UMN database for scattered behavior scene.

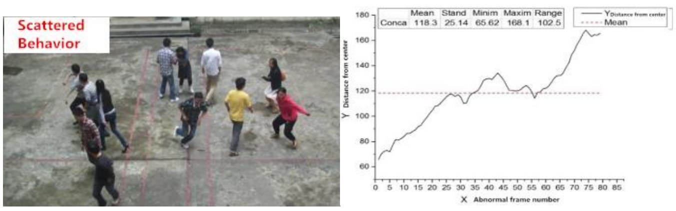

Fig.6(b) Our database for scattered behavior scene.

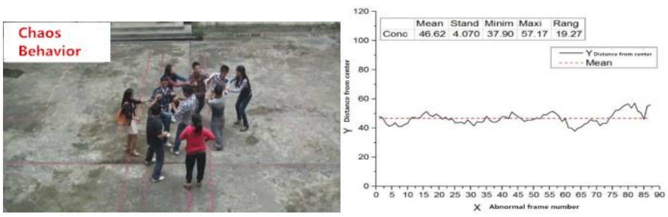

Fig.6(c) Our database for chaos behavior scene.

Fig.6. The two types abnormal behavior in different database

From the graph of Fig.6 and the reference index on the map can be seen, chaos and scattered behavior are very different, the type is significant. The mean and dispersion for the two different types of abnormal behavior are shown in TABLE II.

TABLE II. THE EXPERIMENTAL RESULT OF DATA SET

data set	Mean	Dispersion	FPR	TPR
UMN database, scattered behavior	107.9	22.16	0.1	0.992
Homemade database, scattered behavior	118.3	25.14	0.1	0.989
Homemade database, chaos behavior	46.6	4.07	0.1	0.994

C. Conclusion

In this paper, a new feature descriptor called the SMP, which is the significant moving points' characteristics obtained by GMM, which is introduced for crowd abnormal behavior detection for video frames. Then we combine the spatio-temporal model to determine the crowd abnormal behavior. Proposed algorithm can adapt to the environment, such as lighting and other background changes, it does not need a lot of training data, it can update parameters according to the slow changes of the scene, and it can achieve real-time update effect. The experimental results from the UMN dataset and the videos taken by us show that the proposed algorithm is successful in detecting the different types of abnormal crowd behavior and superior to the state-of-the-art algorithms in terms of its performance, it can realize not only the fast identification of abnormal behavior at low false detection rate, but also the requirement of real-time recognition, which is used for public security purposes.

ACKNOWLEDGMENT

The work presented in this paper was partially supported by National High Technology Research and Development Program(Grant number 2015AA01A707),2014 National Natural Science Foundation of China (Grant number 61371081), 2015 National Natural Young Science Foundation of China (Grant number 61401381), Fujian Educational and Research Projects for Youth and Mid-aged Teacher (JA14002).

Corresponding author: Lianfen Huang lfhuang@xmu.edu.cn

REFERENCES

[1] Q. Wang, Q. Ma, CH. L，HY. L, and CY. Z "Hybrid Histogram of Oriented Optical Flow for Abnormal Behavior Detection in Crowd Scenes," IJPRAI vol. 30, pp. 165–179, February, 2016.

[2] Z. Fang, F. Fei, Y. F, C. L, and N. X, "Abnormal event detection in crowded scenes based on deep learning," Multimedia Tools & Applications, USA, vol.75, pp.14617-14639, November, 2016.

[3] R. Mehran, A. Oyama, and M. Shah, "Abnormal crowd behavior detection using social force model," Computer Vision and Pattern Recognition, vol. 2, pp. 935-942, June, 2009.

[4] T. Wang, and H. Snoussi, "Histograms of optical flow orientation for visual abnormal events detection," Advanced Video and Signal-Based Surveillance, vol.18, pp. 13-18, September 2012.

[5] Y. Cong, J. Yuan, and J. Liu, "Abnormal event detection in crowded scenes using sparse representation," Pattern Recogn , vol. 46, pp. 1851-1864, July, 2013.

[6] C. Stauffer, and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," Computer Vision and Pattern Recognition, vol.2, pp.22-29, June, 1999.

[7] El. Attar, A, R. Khatoun, and M. Lemercier, "A Gaussian mixture model for dynamic detection of abnormal behavior in smartphone applications," Global Information Infrastructure and Networking Symposium, vol. 60, pp.1-6, October, 2014.

[8] Jie Feng, Chao Zhang, and Pengwei Hao, "Online learning with self-organizing maps for anomaly detection in crowd scenes," Pattern Recognition, vol.23, pp. 3599-3602, August 2010.

[9] UMN, "Unusual crowd activity dataset of University of Minnesota, " available from http://mhacsumnedu/movies/crowd-activity-allavi.