Crowd escape event detection based on Direction-Collectiveness Model

Mengdi Wang, Faliang Chang and Youmei Zhang
School of Control Science and Engineering
Shandong University,17923 Jingshi Road, Jinan 250061, PR China
[e-mail: flchang@sdu.edu.cn]
*Corresponding author: Faliang Chang

Received July 30, 2017; revised November 28, 2017; accepted April 11, 2018; published September 30, 2018

Abstract

Crowd escape event detection has become one of the hottest problems in intelligent surveillance filed. When the ‘escape event’ occurs, pedestrians will escape in a disordered way with different velocities and directions. Based on these characteristics, this paper proposes a Direction-Collectiveness Model to detect escape event in crowd scenes. First, we extract a set of trajectories from video sequences by using generalized Kanade-Lucas-Tomasi key point tracker (gKLT). Second, a Direction-Collectiveness Model is built based on the randomness of velocity and orientation calculated from the trajectories to express the movement of the crowd. This model can describe the movement of the crowd adequately. To obtain a generalized crowd escape event detector, we adopt an adaptive threshold according to the Direction-Collectiveness index. Experiments conducted on two widely used datasets demonstrate that the proposed model can detect the escape events more effectively from dense crowd.

Keywords: Video Surveillance, Direction-Collectiveness Model, Crowd Escape Event Detection, gKLT Key Point Tracker

This work is supported by the National Natural Science Foundation of China (No.61673244, No.61273277), the Science and Technology Development Plan Project of Shandong Province of China(2017GGX10137)

1. Introduction

With the development of digital information, internet technology has evolved by leaps and bounds. In recent years, an increasing number of CCTV cameras have been installed in the public place, which is to capture and detect the occurrence of abnormal events, such as mass brawl, traffic accidents and stampede, and then send out forewarning in time. Since recognizing the abnormal events from massive videos by hand is unrealistic, it is an urgent requirement to make abnormal event detection automatically. Many researchers have contributed a lot of work in this field, but the automation process still remains immature and complex.

Anomaly event, which can be interpreted as the undesirable event, is opposite to the normal or regular event. Under the normal/regular scenarios, pedestrians prefer to follow their neighbors and share similar direction or speed. But when the anomaly event occurs, shown in **Fig. 1**, out of fear, people would escape in relatively chaotic directions and run as quickly as possible to avoid the potential danger [1], which increases the possibility of stampede death to some extent. As for detecting the anomaly events, it can be divided into two categories. First, it is the single-object detection [2], such as pedestrian fainting detection and pedestrian retrograding detection. Second, it is the multi-objects detection [3-4], which is to detect the crowd abnormal activities.

Fig. 1. Crowd scatters as abnormal events happen

In order to automatically detect crowd abnormal behaviors through surveillance videos, the primary task is to analyze the behavior of the crowd. In recent years, many computer vision methods have been proposed to deal with this problem for public safety. Mehran et al. [5] proposed a social force model to detect and localize abnormal crowd behaviors. The grid of moving particles is placed over the video frame and it is advected with the spatio-temporal average of optical flow. They regarded the moving particles as individuals and estimated their interaction forces to capture the dynamic behavior of the crowd. Recently, Zhou et al. [6] proposed a "Collectiveness" descriptor that integrated path similarities among crowd on collective manifold composed of self-driven particles and they applied it to crowd segmentation and collective merging. This "Collectiveness" descriptor can evaluate the level of individuals acting as a union in crowd scene.

Inspired by the above work, this paper focuses on the crowd abnormal activities and supposes the crowd (more than 50 people) have coherent and homogenous movement in a spatial region. A novel Direction-Collectiveness Model for crowd escape events detection is proposed. This model combines the randomness of direction with the Collectiveness descriptor and calculates a Direction-Collectiveness index which is a measure of the direction and movement randomness.

There are two assumptions for the effectiveness of our model. One assumption is that the cause which makes the crowd escape exists in the scene. The other assumption is that, in most cases of escape movement, pedestrians deviate from their intended trajectories quickly, which leads to the chaotic directions and then the crowds become unsystematic.

On the whole, the primary contributions of this paper can be summarized as:

- The “Collectiveness” descriptor is applied to distinguish crowd self-organized and unsystematic behaviors.
- A Direction-Collectiveness Model is proposed for detecting abnormal crowd escape event.
- Compared with traditional methods, the detection accuracy of our method has been significantly improved and it demonstrates the wide performance margin.

The remainder of this paper is organized below. Relevant work is reviewed in Section 2. The proposed Direction-Collectiveness Model for crowd escape detection is shown in Section 3. Detection mechanism and the threshold selection strategy are introduced in Section 4. Section 5 contains the experimental results, which demonstrates the feasibility and effectiveness of the proposed method using some challenging benchmark video sequences. Section 6 summarizes the merits and limitations of the proposed approach, as well as our future work.

2. Related Work

Zhan et al. [7] and Popoola et al. [8] gave thorough reviews of anomaly detection in crowd scene. But in this section, we mainly pay attention to the crowd motion pattern extraction and abnormal detection of escape events. Several important topics are listed and analyzed as below.

Crowd motion patterns or features extraction for escape behavior characterization have attracted considerable interest, which is very likely to improve the performance of anomaly detection. Most of the frequently-used patterns or features can be roughly divided into optical flow [5][9-11], trajectories [6][12-15], dynamic texture patterns [16-17] and spatio-temporal context [18-19]. In general, researchers need to select several of these patterns for feature fusion rather than just use only one. Zhang et al. [9] utilized a social attribute-aware force model (SAFM) based on optical flow by calculating the interaction force for abnormal crowd pattern detection. Zaidenberg et al. [10] proposed an unsupervised approach by combining dense trajectories based on optical flow with connected component analysis. This method can detect “sudden” movements in surveillance videos without requiring any training or camera calibration. Xu et al. [11] proposed an event recognition method based on the pyramid Lucas & Kanade method to calculate the optical flow for every point at every pyramid level. The crowd motion pattern was constructed by the orientation and magnitude of particles movement. Cosar et al [12] proposed an integrated pipeline model by combining object tracking with pixel based analysis. And they integrated objects’ velocity and direction into their model for abnormal behavior analysis, especially for complex and finer behavior. Piciarelli et al [13] proposed a method based on single-class Support Vector Machine (SVM) clustering to identify anomalous trajectories. They classified the normal and abnormal trajectories without priori knowledge on the distribution of outliers. Trojanova et al. [14] introduced a framework for motion pattern segmentation in crowd scene, this method used short tracklets detected by dense trajectories and revealed the collective motion of individuals independent of the crowd density. Fradi and Dugelay [15] proposed a method based on analyzing some attributes of feature tracks. The feature tracking allowed excluding feature points on the background and extracting long-term trajectories. The variation of these attributes (local density, speed, and flow direction) in time was employed to determine the ongoing crowd behaviors. Xiong et al. [16] proposed the energy model based on potential energy and kinetic energy to detect pedestrian gathering and running of crowd. They estimated the crowd density and Crowd Distribution Index (CDI) by extracting dynamic texture pattern to represent the potential energy and the dispersion respectively and combine CDI with optical flow to represent the kinetic energy. Cermeno et al. [17] proposed a holistic approach that they extracted global features like color, texture, and shape from each frame and put them together to form a frame feature vector. A Multi Layer Perceptron trained by back-propagation learning algorithm was used for feature vector classification. Wang et al. [18] employed high-frequency and spatio-temporal (HFST) features to characterize the crowd motion for both global and local abnormal crowd event detection. Kratz and Nishin [19] presented a statistical framework to model the local spatio-temporal crowd motion pattern in extremely crowded scenes. The temporal relationship between local spatio-temporal motion patterns is captured via a distribution-based Hidden Markov Model (HMM) and the spatial relationship by a coupled HMM.

For abnormal detection of escape events, it is necessary to choose a suitable detection mechanism, which will have a great impact on the computational complexity and accuracy of the algorithm. The detection mechanism of most abnormal event detection methods can be categorized into supervised, unsupervised and semi-supervised. Supervised methods [20] build motion pattern models relying on the labeled data, test videos which do not fit the models will be regarded as abnormal. Thus, they need a large number of training data which contains a comprehensive set of all possible scenarios [22] and these abnormal events should be well defined. Unsupervised methods [10][23-24] build motion pattern models directly without any labeled training data in advance. They learn the normal and abnormal behavior patterns from the statistical properties of the observed data. Both HMM and Bayesian model are common probability statistical tools for unsupervised methods. Semi-supervised methods [25] fall in-between the first two and they build motion pattern models using partially labeled data at either the features level or the clips level.

3. Direction-Collectiveness Model

In normal group movement, pedestrians would like to follow neighbors who are aiming for the same directions in self-organized motion. In most cases of escape movement, pedestrians deviate from their intended trajectories quickly, which causes the chaotic directions and unsystematic crowd as shown as in **Fig. 2**. We choose the “Collectiveness” feature and direction feature to represent the degree of consistency and the degree of randomness, respectively. **Fig. 3** shows the procedure of our model.

Fig. 2. Representative frames of normal and abnormal event: (a) normal walking, (b) abnormal escaping

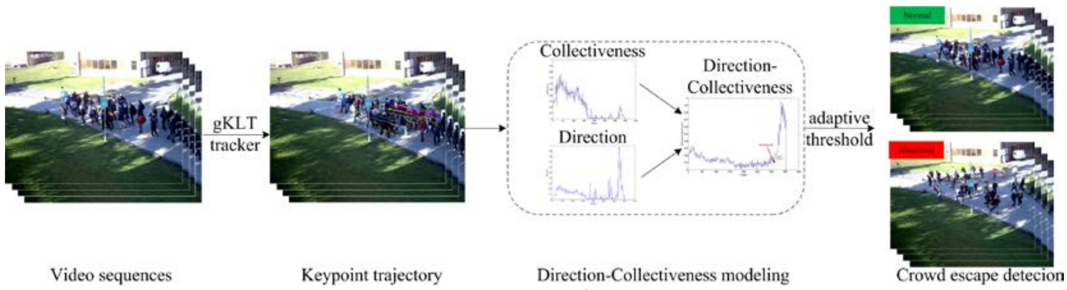

Fig. 3. The system framework

We first use the generalized Kanade-Lucas-Tomasi key point tracker (gKLT) [26] to extract motion particles' trajectories for each frame to get the trajectory coordinate matrix. Then, the velocity and direction of particles can be calculated by using the coordinate matrix. **Fig. 4** shows the mapping graph of particles which are extracted by gKLT and their parameters. Where k_{t_i} is i-th particle at time t, X_{t_i} and Y_{t_i} are the horizontal and vertical coordinates of particles respectively, N is the number of particles, and V_{t_i} and θ_{t_i} is the velocity and direction of particles respectively.

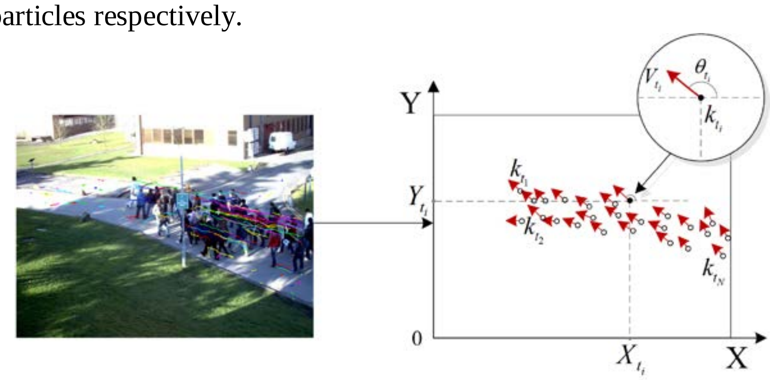

Fig. 4. Mapping graph of particles extracted by gKLT key point tracker

In order to facilitate the description of the subsequent anomaly detection algorithm, the key points set K_t at time t which obtained by the gKLT tracking algorithm is expressed as follows:

$$K_t = \{k_{t_1}, k_{t_2}, \cdots, k_{t_N}\} \ , \tag{1}$$

$$k_{t_i} = \{X_{t_i}, Y_{t_i}, V_{t_i}, \theta_{t_i}\} \ . \tag{2}$$

Then particles' position, direction and magnitude are calculated to obtain the degree of collectiveness and direction chaos. After that, the degree of direction chaos is combined with the collectiveness to get an index which is defined as Direction-Collectiveness. Finally, the Direction-Collectiveness index is used to detect the crowd abnormal event with an empirically defined threshold.

3.1 The Collectiveness Feature

In certain scenarios, individual movement influences neighbors' motion and eventually leads to the variation in the motion of the whole crowd. Hence, crowd collectiveness should be determined by the collectiveness of its constituent individuals. In this paper, collectiveness can be defined as the behavior similarity of neighbors in crowd motion, and behavior similarity can be characterized by the velocity correlation between adjacent particles. Suppose k_{t_i}, k_{t_j} are adjacent particles, φ is the angle between the velocity directions of the two particles, the behavior similarity $w_t(k_{t_i}, k_{t_j}) \in [0,1]$ is defined as:

$$w_t(k_{t_i}, k_{t_j}) = \begin{cases} C_t(i,j) \ , & 0 \le \varphi < \dfrac{\pi}{2} \\ 0 \ , & \varphi \ge \dfrac{\pi}{2} \end{cases} \ , \tag{3}$$

$$C_t(i,j) = \frac{\vec{\boldsymbol{V}}_{t_i} \bullet \vec{\boldsymbol{V}}_{t_j}}{\| \vec{\boldsymbol{V}}_{t_i} \| \times \| \vec{\boldsymbol{V}}_{t_j} \|} = \cos\varphi \quad . \tag{4}$$

The above method only considers the influence between adjacent particles. However, when two particles have a certain distance that they are separated by several other particles, the result would be unreliable to calculate the behavior similarity of these two particles using the above formula. In order to solve this problem, the behavior similarity should be estimated based on the path topological structure [5].

Assume that $\gamma_L = \{k_0, k_1, \cdots, k_L\}$ $(k_0 = k_{t_i}, k_L = k_{t_j})$ represents a path of length *L* from k_{t_i} to k_{t_j} passing through $k_1, k_2, \cdots, k_{L-1}$.Then the path behavior similarity of all *L*-length paths between k_{t_i} to k_{t_j} is defined as:

$$C_L(k_{t_i}, k_{t_j}) = \sum_{\gamma_L \in P} C_{\gamma_L}(k_{t_i}, k_{t_j}) = \sum_{\gamma_L \in P} \{\prod\nolimits_{n=0}^{L} w_t(k_{t_n}, k_{t_{n+1}})\} \ , \tag{5}$$

where *P* contains all *L*-length paths between k_{t_i} to k_{t_j}. Thus, the Collectiveness index of single particle can be denoted as:

$$Collectiveness\,(k_{t_i}) = \sum\nolimits_{L=1}^{\infty} \{\omega^L \times (\sum\nolimits_{k_{t_j} \in K} C_L(k_{t_i}, k_{t_j}))\} \ , \tag{6}$$

where ω is the real value regularization factor which contributes to reduce the effect of the Collectiveness index exponentially increasing with *L*. And the crowd Collectiveness index at time *t* which is bounded by [0,1] can be defined as:

$$Collectiveness(t) = \frac{1}{N} \sum\nolimits_{i=1}^{N} Collectiveness(k_{t_i}) \ . \tag{7}$$

3.2 The Direction-Collectiveness Descriptor

While the motion dynamics of the crowd escape events are extracted based on the

Collectiveness descriptor and it has superior performance for high density crowd, with the decrease in the number of crowds, the extracted trajectories will be reduced accordingly. It is a challenge for calculating the behavior similarity of the Collectiveness model, and the measuring performance will decline. In order to suppress the performance degradation, we propose a Global-Direction descriptor to extract the motion patterns at the global level while the Collectiveness descriptor calculates the behavior similarity of the adjacent particles at the local level. The Direction-Collectiveness Model can be represented by combining these two salient descriptors mathematically.

When an abnormal event occurs, pedestrians instinctively spread out from the dangerous place and deviate from their intended trajectories, which may lead to an increased randomicity of the crowd motion direction. The directions of the mean velocity $\theta_{mean}(t)$ and the maximum velocity $\theta_{\max}(t)$ of all particles are two important indexes that characterize the overall motion pattern of the crowd, and different between the velocity direction of individual particles θ_{t_i} and the relevant $\theta_{mean}(t)$ and $\theta_{\max}(t)$ can reflect the deviation of individual particles from the crowd motion. Hence, the Global-Direction descriptor $Direction(i,t) \in [0,1]$ is the degree of directional randomness for particle *i* at time *t* can be defined as:

$$Direction(i,t) = \rho \cdot \sqrt{\frac{\Delta\theta_{mean}^2(i,t) + \Delta\theta_{\max}^2(i,t)}{2}} \,, \tag{8}$$

where $\Delta\theta_{mean}(i,t)$ is the difference between the velocity direction of the particle *i* and the mean velocity direction of all particles at time *t*, $\Delta\theta_{\max}(i,t)$ is the difference between the velocity direction of the particle *i* and the maximum velocity direction at time *t*. And $\sqrt{\frac{\Delta\theta_{mean}^2(i,t) + \Delta\theta_{\max}^2(i,t)}{2}}$ is the root mean square of $\Delta\theta_{mean}(i,t)$ and $\Delta\theta_{\max}(i,t)$ to characterize the average dispersion of directional randomness for particle *i*. And ρ is the normalization factor to maintain the range of $Direction(i,t)$ between 0 to 1, let $\rho = 1/\pi$ because the range of $\Delta\theta_{mean}(i,t)$ and $\Delta\theta_{\max}(i,t)$ are both between $-\pi$ to π. Therefore, the formula can be simplified as:

$$Direction(i,t) = \frac{\sqrt{\Delta\theta_{mean}^2(i,t) + \Delta\theta_{\max}^2(i,t)}}{\sqrt{2}\pi} \,. \tag{9}$$

The specific formulas of $\Delta\theta_{mean}(i,t)$ and $\Delta\theta_{\max}(i,t)$ are as follows:

$$\Delta\theta_{mean}(i,t) = \left|\theta_{t_i} - \theta_{mean}(t)\right| = \left|\theta_{t_i} - \frac{\sum_{i=1}^{N}\theta_{t_i}}{N}\right| \,, \tag{10}$$

$$\Delta\theta_{\max}(i,t) = \left|\theta_{t_i} - \theta_{\max}(t)\right| \,. \tag{11}$$

In the most case of escape behaviors, the Collectiveness index is considerably below normal behaviors' level. Meanwhile the directional distribution of crowd motion will become chaotic. Combining these two salient features makes it possible to distinguish abnormal behaviors from video sequences. Feature addition and feature multiplication are two commonly used methods of feature combining. **Fig. 5** shows the fused value curves by using these two methods. With the direction from orderly to randomly (from 0 to 1) and the Collectiveness index from large to small (from 1 to 0), the fused value of feature addition is showing parabolic trend growth followed by exponential growth. And the fused value of feature multiplication is showing exponential growth with the larger base number than addition method. Thus, using feature multiplication rather than feature addition can be better to distinguish between normal and escape behavior and the Direction-Collectiveness descriptor at time t can be denoted as follow:

$$Direction\text{-}Collectiveness(t) = \frac{1}{N}\sum\nolimits_{i=1}^{N} Direction(i,t)\left|1 - Collectiveness(k_{t_i})\right|, \quad (12)$$

where N is the quantity of particles at time t.

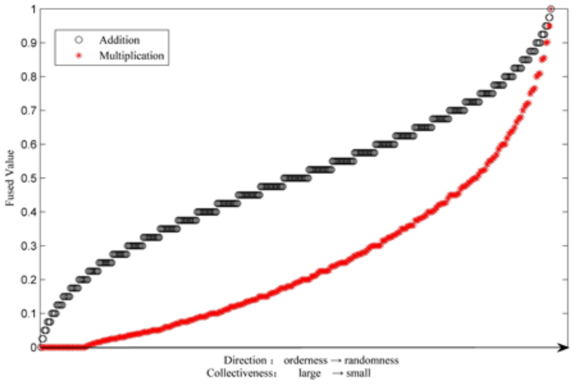

Fig. 5. The curve of features addition and features multiplication

After the above operation, the Direction-Collectiveness value for each frame is obtained, and the value is between 0 and 1. Large value indicates that the cluster degree is low and the motion direction is chaotic. **Fig. 6** shows the Collectiveness curve and the Direction-Collectiveness curve from normal walking to abnormal escaping.

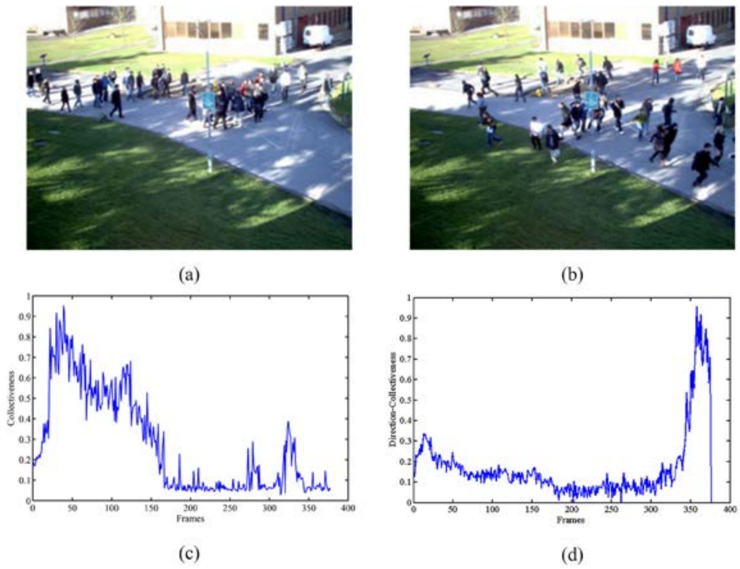

Fig. 6. Two curves from (a) normal walking to (b) abnormal escaping: (c) Collectiveness curve, (d) Direction-Collectiveness curve

From frame 1 to 190, people walked normally, and crowd gathered from frame 191 to 334, then crowd suddenly fled at the 335th frame. Through the varying curve we can see that, the Collectiveness curve begins to increase at about 14th frame when the pedestrian gradually appeared in the surveillance area, and after the 167th frame it falls into the trough. In addition, in the 271~286 frame and 317~335 frame, the curve has two steep rise with small amplitude. The Direction-Collectiveness curve begins to increase at about 340th frame which coincides with the time of the actual escape. And at about 358th frame it reaches the peak. After 372th frame, curve begins to fall with pedestrians leaving out of the surveillance area. Through the comparison of the two curves we can see that our model which added global direction component can better distinguish between normal and escape events than the original collectiveness model.

4. Abnormal Crowd Event Detection

In the previous section, we propose the Direction-Collectiveness descriptor and introduce how to apply this new descriptor for abnormal crowd event detection. The test video will be input into our model to calculate the Direction-Collectiveness value for each frame. Considering the continuity of the abnormal behavior, when a sudden change of the Direction-Collectiveness value arises, we regard it as the noise and the system will not alarm. And when this rise is maintained for over 10 consecutive frames, it is assumed that an abnormal behavior occurs. We summarize the algorithm in **Algorithm 1**.

Algorithm 1 Anomaly Detection

INPUT: $\{X_{t_i}, Y_{t_i}, V_{t_i} | i \in \mathbf{K}\}_t$
Calculate w_t using Eq.3
Calculate *Collectiveness* using Eq.7
Calculate *Direction-Collectiveness* using Eq.11
$T_{count} \leftarrow 0$
for *t*=1 to length(*Video*)
  if *Direction-Collectiveness(t)>threshold* **then**
    $T_{count} = T_{count} + 1$
  else $T_{count} = 0$
  end if
  if $T_{count} > 10$ **then**
  do *Alarm*!
  end if
end for

After obtaining the curve of Direction-Collectiveness, it is crucial to find an appropriate threshold to distinguish between normal and abnormal for detection performance and accuracy. Manually selecting the appropriate threshold will result in inefficient detecting and universality lacking. In this paper, we use iterative threshold method to train the threshold. Let Φ be the matrix which stores the Direction-Collectiveness value for each frame, and let the initial threshold *Th0* be the average of the maximum and minimum values in matrix Φ.The video frame is divided into abnormal frames and normal frames according to the obtained initial threshold. Then calculate the average Direction-Collectiveness values of the normal frames $\overline{\phi}_{nor}$ and abnormal frames $\overline{\phi}_{abn}$ respectively. Update the new threshold *Th* to be the average of $\overline{\phi}_{nor}$ and $\overline{\phi}_{abn}$ and constantly iterate until the threshold no longer changes, the final threshold is used for anomaly detection. We summarize the algorithm in **Algorithm 2**. The thresholds of our model are mainly affected by the angle between the shooting direction of the camera and the horizontal direction of the ground. Due to the tangential distortion, the smaller the angle is, the less significant the angle between the adjacent particles in the direction of velocity, so the threshold is relatively smaller.

Algorithm 2 Threshold Selection

INPUT: Direction-Collectiveness set $\mathbf{\Phi} = \{\phi(1), \phi(2), \cdots, \phi(\mathrm{n})\}$
1: $Th_0 = (\mathrm{MAX}(\mathbf{\Phi}), \mathrm{MIN}(\mathbf{\Phi}))/2$
2: $\phi_{nor} \leftarrow 0 \quad N_{nor} \leftarrow 0$
3: $\phi_{abn} \leftarrow 0 \quad N_{abn} \leftarrow 0$
4: **for** *t*=1 to length(*Video*)
5:     **if** $\phi(\mathrm{t}) \leq Th_0$ **then**
6:         $\phi_{nor} = \phi_{nor} + \phi(\mathrm{t})$
7:         $N_{nor} = N_{nor} + 1$
8:     **else** $\phi_{abn} = \phi_{abn} + \phi(\mathrm{t})$
9:         $N_{abn} = N_{abn} + 1$
10:     **end if**
11: **end for**
12: $\bar{\phi}_{\mathrm{nor}} = \phi_{nor} / N_{nor}$
13: $\bar{\phi}_{\mathrm{abn}} = \phi_{abn} / N_{abn}$
14: $Th = (\bar{\phi}_{\mathrm{nor}} + \bar{\phi}_{\mathrm{abn}})/2$
15: **if** $Th = Th_0$ **then**
16:     OUTPUT Th
17: **else** Update $Th_0 = Th$ **then**
18:       **back to** *STEP2*
19: **end if**

5. Experimental Results

Our model is evaluated on PETS2009 dataset [27] and UMN dataset [28] which have been widely used for performance evaluation of anomaly detection. In order to verify the flexibility and robustness of our model, we also evaluated the proposed method on WEB dataset [29] which have more complex and diverse escape event videos. We compare our Direction-Collectiveness model (D-CM) with the Collectiveness Model (CM) [6], the Social Force Model (SFM) [5], the Bayesian Model (BM) [30], and the Energy Model (EM) [31]. In order to assess our model more intuitively, we manually mark the ground truth of all the test video sequences in advance. The evaluation criterion is defined as follows:

$$\mathrm{Accuracy(ACC)} = \frac{TP + TN}{N}\ , \tag{13}$$

$$TPR = \frac{TP}{TP + FN}\ , \tag{14}$$

$$FPR = \frac{FP}{FP + TN} \ , \tag{15}$$

where N is the total number of frames for the test video. The Receiver Operating Characteristic (ROC) curve which consists of true positive rate (TPR) and false positive rate (FPR) is used to measure the accuracy for multiple threshold values. Where true positive (TP) is abnormal sample that is correctly detected, true negative (TN) is normal sample that is correctly detected, false positive (FP) is abnormal sample that is incorrectly detected and false negative (FN) is normal sample that is incorrectly detected [32]. Due to the limited number of samples, the cross validation method is applied for data training and testing. When a video sequence in a certain dataset is used for test sample, other remaining video sequences of the same dataset are regarded as training samples.

We implement the proposed model with MATLAB R2014a, and all the simulation experiments are performed on a 3.10GHz quad core with 4.00GB of RAM. For a given video with a resolution of 768×576, it takes on average 1.06 second to compute the Direction-Collectiveness value and takes on average 0.269 second to detect abnormal crowd escape event.

5.1 Evaluation on PETS2009 Dataset

The PETS2009 dataset provides video sequences from a multi-camera installation. In this paper, we use their event recognition set which have four different cameras in four different views recording at the same time [17].Since significant differences in location and illumination among four views, it is extremely challenging to recognize abnormal escape event. In tested scene, pedestrians walk orderly to the gathering point in the first 190 frames. Then they remain in a state of aggregation until the 334th frame, and they quickly escape in multiple directions until the 377th frame. Our model and the other four models are used to detect anomalies in four views video sequences and the accuracy of four models is obtained by comparing their results with ground truth. The detection results of five models are shown in **Figs. 7-10**, the green bar below the image indicates normal behavior and the red bar indicates escape behavior. **Table 1** illustrates the accuracy of these methods for abnormal escape detection. As shown in **Table 1**, the accuracy of our model is superior to the other four methods in four views, and our model also has more stable performance than the other four algorithms. In particular, our model has a qualitative improvement in the detection accuracy compared to the original Collectiveness model. On the other hand, in the forth view, the detection performance of our model is slightly declined by the illumination and the shooting angle.

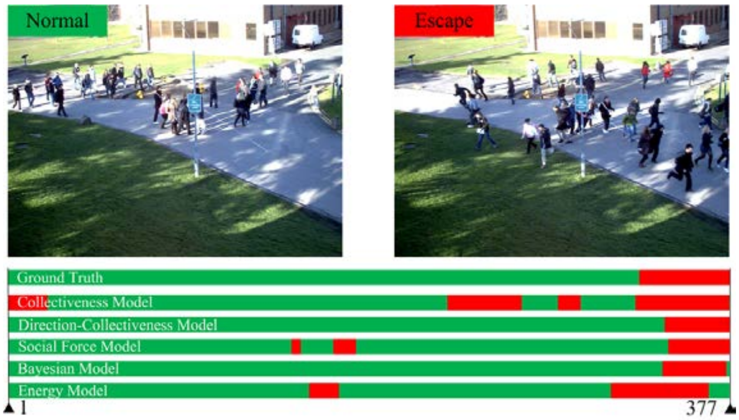

Fig. 7. Abnormal detection result of view 1 in PETS2009 dataset

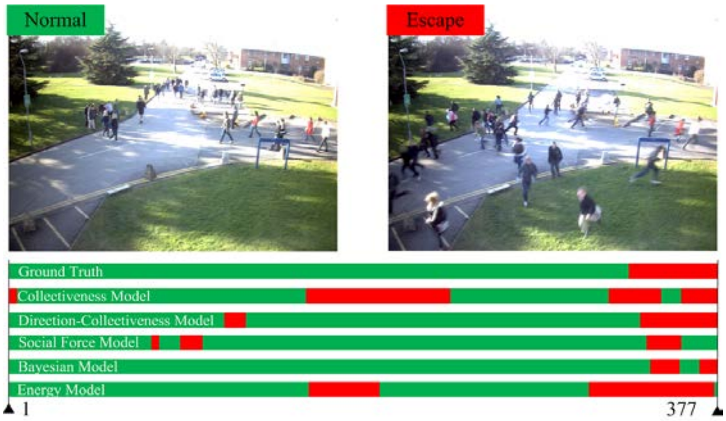

Fig. 8. Abnormal detection result of view 2 in PETS2009 dataset

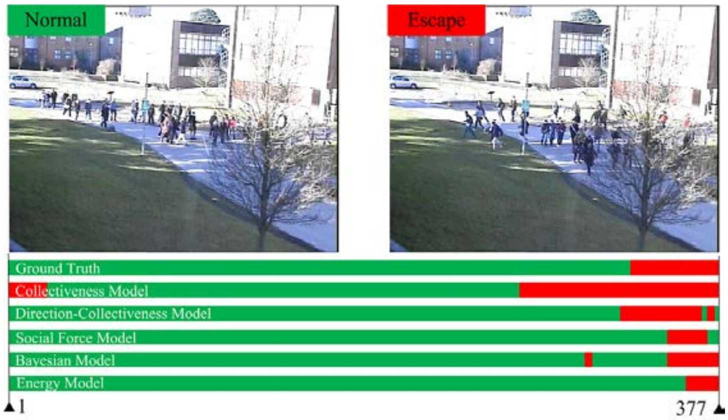

Fig. 9. Abnormal detection result of view 3 in PETS2009 dataset

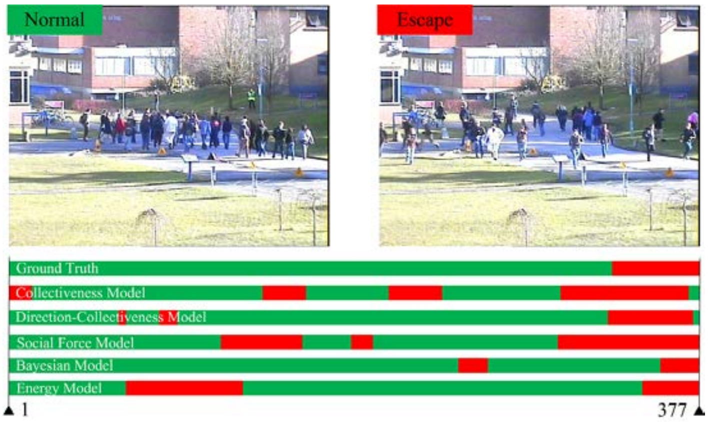

Fig. 10. Abnormal detection result of view 4 in PETS2009 dataset

Table 1. Accuracy (%) comparison of D-CM with CM, SFM, BM and EM in PETS2009 dataset

	CM [6]	**D-CM**	**SFM** [5]	**BM** [30]	**EM** [31]
View_001	81.43	**96.35**	91.22	96.01	93.15
View_002	72.50	**94.18**	89.36	94.15	84.17
View_003	79.15	**95.98**	94.68	95.21	92.12
View_004	73.89	**92.64**	72.14	91.49	86.25
Average Accuracy	76.74	**94.78**	86.85	94.22	88.92

In order to compare the sensitivity and specificity of these algorithms, we plot the ROC curve of the five models by frame-level measurement. We treat one frame as a test sample, the escape frames and normal frames in the ground truth are labeled as positive samples and negative samples respectively. The TPR and FPR are computed with the threshold increasing in turn, and they form the ROC curve which is shown in **Fig. 11**. Subsequently, the area under the ROC curve is calculated as AUC, and it is used to evaluate the detection performance. From the data in the **Table 2**, we can see that our model has higher AUC than the other four algorithms, which demonstrates that our model has better robustness.

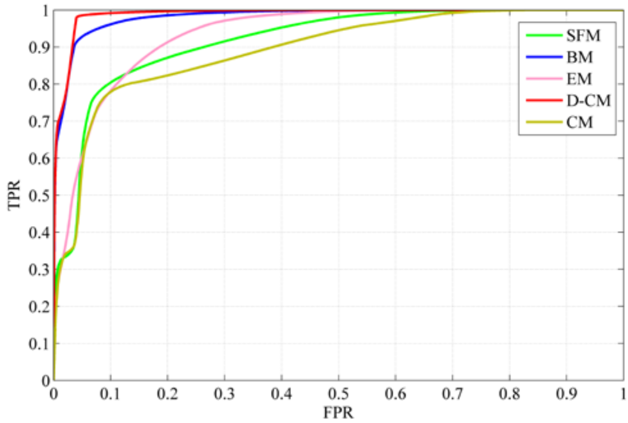

Fig. 11. The ROC curves of abnormal detection in PETS2009 dataset

Table 2. AUC comparison of D-CM with CM, SFM, BM and EM in PETS2009 dataset

Model	CM [6]	D-CM	SFM [5]	BM [30]	EM [31]
AUC	0.902	**0.992**	0.927	0.985	0.943

5.2 Evaluation on UMN Dataset

The UMN dataset provides video sequences from a single-camera installation with multi-scenario. The total length of the video is 7740 frames, in which 1431 frames involve escape. It contains 11 different escape scenarios which can be divided into three different shooting scenes (scene 1: lawn, scene 2: lobby and scene 3: square). According to the different scenes, the first two scenarios are divided into the scene 1, the following six scenarios are divided into the scene 2, and the last three scenarios are divided into the scene 3. The crowd movement in this dataset can be described as: the pedestrians walk around freely in outdoor or indoor open area, and then they suddenly flee away from the field of vision. **Table 3** shows the accuracy of these methods for abnormal escape detection in the three scenes and some representative results are shown in **Figs. 12-14**. The average accuracy of our methods achieves an accuracy of 91.89%, which is higher than that of CM (79.25%), SFM (85.86%), and EM (90.39%) and lower than that of BM (97.01%). Compared with the Collectiveness model, our method has been greatly improved in terms of accuracy. But compared with the previous results in PETS2009 dataset, the detection performance of our method in UMN dataset has declined. The detection results in **Fig. 12-14** show that our model will regard few normal frames which have extremely disordered directions as escape frames. From the following representative results, it can also demonstrate that compared with the other four models, our model can detect the escape frames more completely.

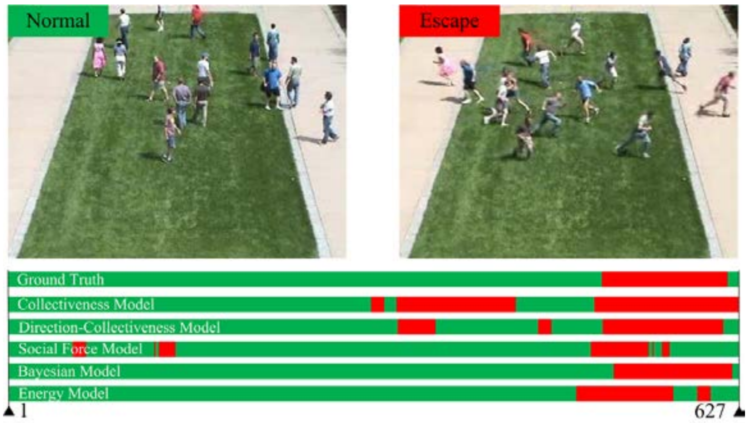

Fig. 12. Abnormal detection result of scenario 1 in UMN dataset

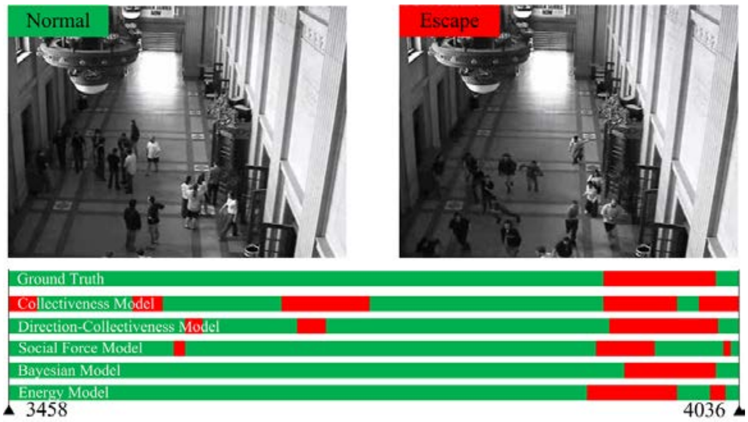

Fig. 13. Abnormal detection result of scenario 2 in UMN dataset

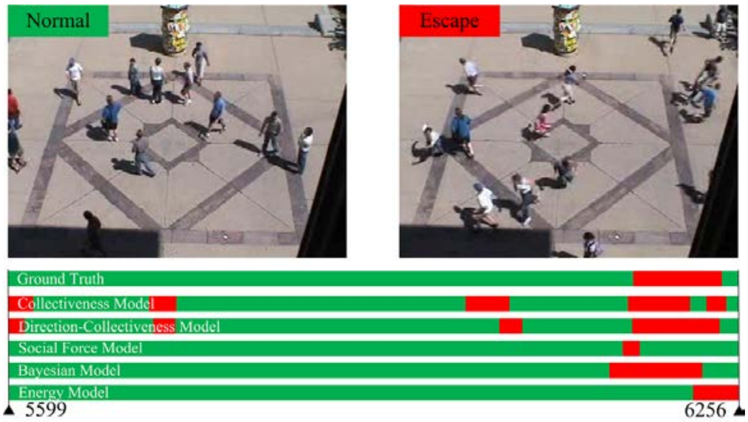

Fig. 14. Abnormal detection result of scenario 3 in UMN dataset

Table 3. Accuracy (%) comparison of D-CM with CM, SFM, BM and EM in UMN dataset

	CM [6]	**D-CM**	**SFM** [5]	**BM** [30]	**EM** [31]
Scene_001	79.44	92.45	84.41	**99.03**	90.79
Scene_002	74.56	91.37	82.35	**95.36**	91.84
Scene_003	83.76	91.86	90.83	**96.63**	88.53
Average Accuracy	79.25	91.89	85.86	**97.01**	90.39

The ROC curve of the UMN dataset is shown in **Fig. 15**. There are 1431 positive samples and 6309 negative samples in the dataset. The AUC of our method is 0.975, which is shown in **Table 4**. The result shows that our method outperforms these state of the art methods except BM (0.986). We attribute the performance to the fact that the robustness of our methods will reduce to some extent for detecting normal behaviors which have extremely disordered directions.

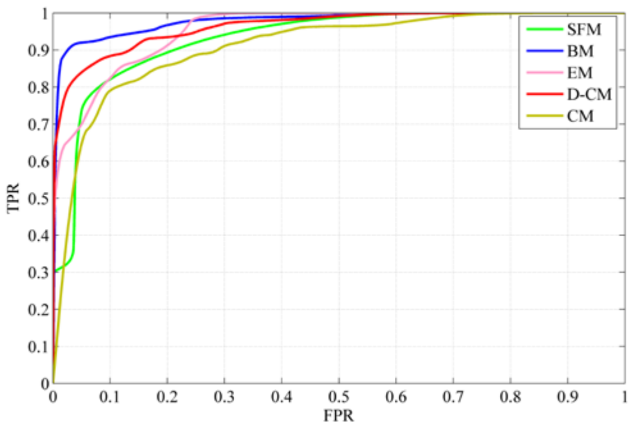

Fig. 15. The ROC curves of abnormal detection in UMN dataset

Table 4. AUC comparison of D-CM with CM, SFM, BM and EM in UMN dataset

Model	CM [6]	D-CM	SFM [5]	BM [30]	EM [31]
AUC	0.952	0.975	0.960	**0.986**	0.968

5.2 More Diverse Examples of Crowd Escape Detection

To make the evaluation more universal and find limitations of the proposed methods, we also test these five models in WEB dataset. The WEB dataset is a collection of abnormal escape events in real surveillance scenes or TV series. The videos in this dataset are collected from Thought Equity [29], and we have numbered the measured videos. Therefore, these videos can be found from Thought Equity website using their unique numbers. Compared with the man-made dataset, the scenes of WEB dataset are more complex and more difficult to detect due to the occlusion, illumination and obstacle. There are four different types of videos in WEB dataset selected for detection. In the first video, the crowd ran with an orderly manner at the first 200 frames, and then they began scattered escaping until the 409th frame. This sequence was taken in the dark environment with the low contrast. The second video contains 183 frames and it has the complex background, shooting incident occurred at the 68th frame, and then the crowd fled and evaded until the end of the video. The background of the third video is the football field, some people were still while few people walked in pairs. At the 450 frame, the crowd panicked and disorderly escaped until they disappeared in the screen. Different from other escape test videos, the crowd in this video was initially distributed around the edge of the screen. When a dangerous situation occurred, they fled into the frame where there are serious occlusions which make the identification of the crowd motion patterns more difficult. The fourth is a low-angle shot video, the crowd was still in the first 375 frames, and then they began to escape following the same path from the center to the right side of the screen.

Table 5 shows the accuracy of these methods for abnormal escape detection in WEB dataset and the final results are shown in **Figs. 16-19**. It shows that our method outperforms the other four models for the first three scenarios which illustrates that our model is robust to low-contrast, occlusion and complex background. However, for the last scenario, the accuracy of our method is only higher than CM and SFM, and has a certain gap with the EM and BM. It illustrates the limitation that the proposed method may incorrect recognize few escaping behaviors which have orderly motion patterns as the normal behaviors.

Fig. 16. Abnormal detection result of scenario 1 in WEB dataset

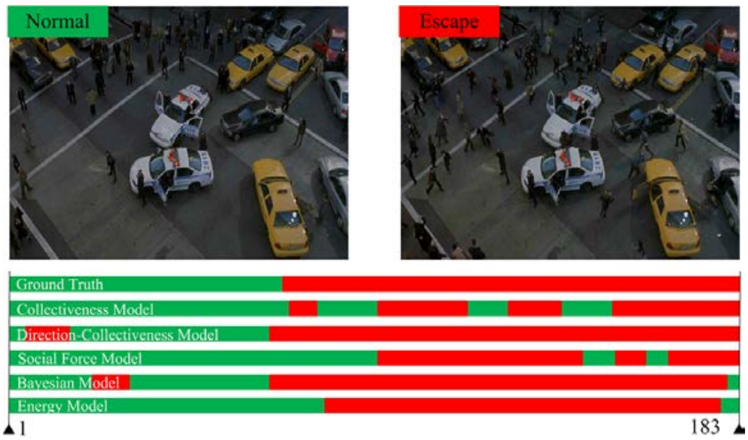

Fig. 17. Abnormal detection result of scenario 2 in WEB dataset

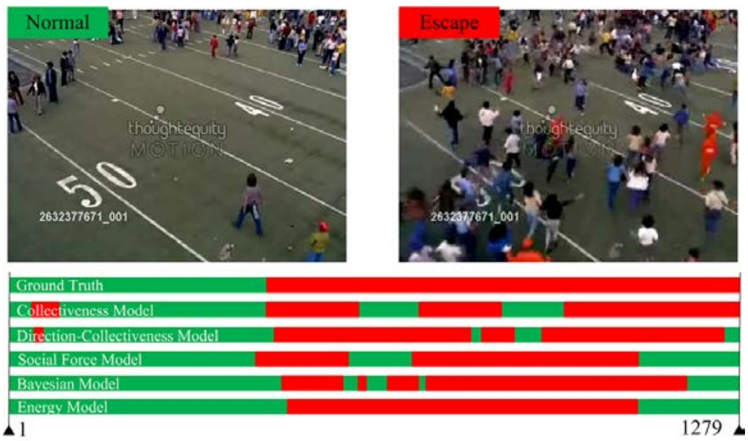

Fig. 18. Abnormal detection result of scenario 3 in WEB dataset

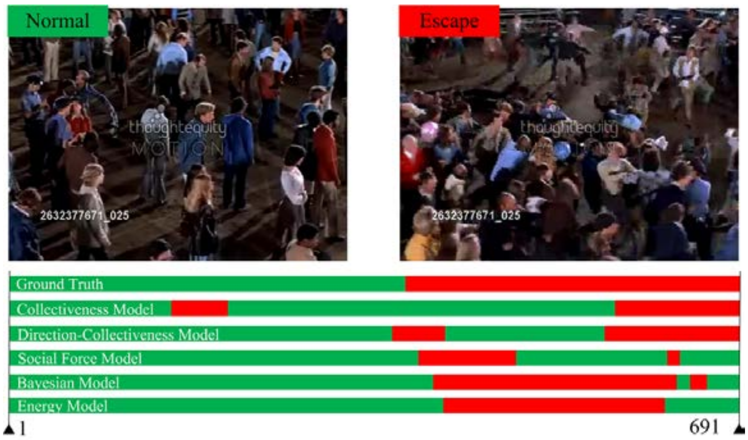

Fig. 19. Abnormal detection result of scenario 4 in WEB dataset

Table 5. Accuracy (%) comparison of D-CM with CM, SFM, BM and EM in WEB dataset

	CM [6]	**D-CM**	**SFM** [5]	**BM** [30]	**EM** [31]
Web_001 (No.345182_162741)	78.24	**93.94**	83.93	93.23	89.69
Web_002 (No.3452204_031)	85.80	**92.10**	79.68	91.84	90.22
Web_003 (No.2632377671_001)	79.56	**91.37**	76.10	84.87	83.17
Web_004 (No. 2632377671_025)	63.53	79.96	69.22	**87.12**	84.51
Average Accuracy	76.78	**89.34**	77.23	89.26	86.90

5.3 Time Complexity Analysis

In this paper, the proposed approach includes two main steps: one is extracting motion particles' trajectories and the other is calculating the Direction-Collectiveness index. So the cost of our method also is comprised of two parts. The cost of motion particles' trajectories extraction is mainly determined by the optically flow. And the cost of the Direction-Collectiveness index calculation is mainly determined by the number of particles. Therefore, the time complexity is $T(n) = O(n^2 \cdot \log n)$, where n is the number of the particles extracted by gKLT key point tracker.

6. Conclusions

In this paper, the Direction-Collectiveness Model is proposed for crowd escape event detection. When unusual event occurs, people will instinctively scattered run and escape away from dangers in most cases. Thus, we extract the Collectiveness index and direction randomness information to characterize the differences between normal events and escape events. Experimental results confirm that our model has a qualitative improvement in the detection performance compared to the Collectiveness model. Experiments conducted on PETS2009 dataset and WEB dataset demonstrate that the detection performance of our model is superior to several state-of-the-art models in terms of accuracy and AUC. And our model can reduce the impact of low-contrast, occlusion and complex background. However, experiments conducted on UMN dataset and scenario 4 of WEB dataset show that the detection performance of our model may be degraded when dealing with the normal behavior which has chaotic directions, and the escape behavior has orderly motion patterns. This is the major limitation of our method, and it should be investigated in the future work.

References

[1] Y. Benabbas, N. Ihaddadene, C. Djeraba, "Motion Pattern Extraction and Event Detection for Automatic Visual Surveillance," *Eurasip Journal on Image & Video Processing*, vol. 8, pp. 1-15, December, 2011. Article (CrossRef Link).

[2] C. Wang, X. Y. Wu, N. N. Li, et al., "Abnormal detection based on gait analysis," *10th World Congress on Intelligent Control and Automation*, pp. 4859-4864, IEEE, Beijing, China, July, 2012. Article (CrossRef Link).

[3] M. M. Najafabadi, M. Rahmati, "Anomaly detection in structured/unstructured crowd scenes," in *Proc. of 7th Int. Conf. on Digital Information Management*, pp. 79-83, IEEE, Macao, China, August, 2012. Article (CrossRef Link).

[4] D. G. Lee, H. I. Suk, S. W. Lee, "Crowd behavior representation using motion influence matrix for anomaly detection," in *Proc. of 2nd IAPR Asian Conf. on Pattern Recognition*, pp. 110-114, IEEE Computer Society, Naha, Japan, November, 2013. Article (CrossRef Link).

[5] R. Mehran, A. Oyama, M. Shah, "Abnormal crowd behavior detection using social force model," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 935-942, IEEE, Miami, Florida, June, 2009. Article (CrossRef Link).

[6] B. L. Zhou, X. O. Tang, H. P. Zhang, et al., "Measuring crowd collectiveness," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 36, no. 8, pp. 1586-1599, June, 2014. Article (CrossRef Link).

[7] B. B. Zhan, D. N. Monekosso, P. Remagnino, et al., "Crowd analysis: a survey," *Machine Vision and Applications*, vol. 19, no. 5, pp. 345-357, April, 2008. Article (CrossRef Link).

[8] O. P. Popoola, K. J. Wang, "Video-Based Abnormal Human Behavior Recognition—A Review," *IEEE Trans. Systems, Man, and Cybernetics, Part C*, vol. 42, no. 6, pp. 865 – 878, January, 2012. Article (CrossRef Link).

[9] Y. H. Zhang, L. Qin, H. X. Yao, et al., "Abnormal crowd behavior detection based on social attribute-aware force model," in *Proc. of 19th IEEE Int. Conf. Image Processing*, pp. 2689-2692, IEEE, Orland, Florida, September, 2012. Article (CrossRef Link).

[10] S. Zaidenberg, P. Bilinski, F. Bremond, "Towards Unsupervised Sudden Group Movement Discovery for Video Surveillance," in *Proc. of 9th Int. Joint Conf. on Computer Vision, Imaging and Computer Graphics Theory and Applications*, pp.388-395, IEEE, Lisbon, Portugal, January, 2014. Article (CrossRef Link).

[11] T. Xu, P. Peng, X. Fang, et al., "Single and multiple view detection, tracking and video analysis in crowded environments," in *Proc. of 10th IEEE Int. Conf. on Advanced Video and Signal Based Surveillance*, pp. 494-499, Beijing, China, September, 2012. Article (CrossRef Link).

[12] S. Cosar, G. Donatiello, V. Bogorny, et al., "Towards Abnormal Trajectory and Event Detection in Video Surveillance," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 27, no. 3, pp. 683-695, March, 2017. Article (CrossRef Link).

[13] C. Piciarelli, C. Micheloni, G. L. Foresti, "Trajectory-Based Anomalous Event Detection," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 18, no. 11, pp. 1544-1554, November, 2008. Article (CrossRef Link).

[14] J. Trojanová, K Křehnáč, F Brémond, "Data-Driven Motion Pattern Segmentation in a Crowded Environments," in *Proc. of Proceedings of ECCV 2016 Workshops*, pp. 760-774, Springer, Amsterdam, Netherlands, October, 2016. Article (CrossRef Link).

[15] H. Fradi, J. L. Dugelay, "Spatial and temporal variations of feature tracks for crowd behavior analysis," *Journal on Multimodal User Interfaces*, vol.10, no. 4, pp. 307-317, December, 2016. Article (CrossRef Link).

[16] G. G. Xiong, X. Y. Wu, Y. L. Chen, et al., "Abnormal crowd behavior detection based on the energy model," in *Proc. of IEEE Int. Conf. on Information and Automation*, pp. 495-500, IEEE, Shenzhen, China, June, 2011. Article (CrossRef Link).
[17] E. Cermeno, S. Mallor, J. A. Siguenza, "Learning crowd behavior for event recognition," in *Proc. of IEEE Int. Workshop on Performance Evaluation of Tracking and Surveillance*, pp. 1-5, IEEE, Clearwater Beach, Florida, January, 2013. Article (CrossRef Link).
[18] B. Wang, M. Ye, X. Li, et al., "Abnormal crowd behavior detection using high-frequency and spatio-temporal features," *Machine Vision and Applications*, vol. 23, no. 3, pp. 501-511, May, 2012. Article (CrossRef Link).
[19] L. Kratz, K. Nishino, "Anomaly detection in extremely crowd scenes using spatio-temporal motion pattern models," in *Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pp. 1446-1453, IEEE, Miami, Florida, June, 2009. Article (CrossRef Link).
[20] Y. Wang, K. Q. Huang, T. N. Tan, "Abnormal Activity Recognition in Office Based on R Transform," in *Proc. of 14th IEEE Int. Conf. on Image Processing*, pp. 341-344, IEEE, San Antonio, Texas, September, 2007. Article (CrossRef Link).
[21] F. Nater, H. Grabner, T. Jaeggli, et al., "Tracker trees for unusual event detection," in *Proc. of 12th IEEE Int. Conf. on Computer Vision Workshops*, pp. 1113-1120, IEEE, Kyoto, Japan, November, 2009. Article (CrossRef Link).
[22] Y. M. Zhang, F. L. Chang, M. D. Wang, et al., "Auxiliary learning for Crowd Counting via Count-net," *Neurocomputing*, vol. 273, no. 17, pp. 190-198, August, 2017. Article (CrossRef Link).
[23] C. S. Liu, F. L. Chang, Z. X. Chen, et al., "Fast Traffic Sign Recognition via High-Contrast Region Extraction and Extended Sparse Representation," *IEEE Trans. Intelligent Transportation Systems*, vol. 17, no.1, pp.79-92, August, 2015. Article (CrossRef Link).
[24] D. H. Hu, X. X. Zhang, J. Yin, et al., "Abnormal activity recognition based on HDP-HMM models," in *Proc. of Proceedings of the 21st Int. Joint Conf. on Artificial Intelligence*, pp. 1715-1720, Pasadena, California, July, 2009.
[25] D. Zhang, D. G. Perez, S. Bengio, et al., "Semi-Supervised Adapted HMMs for Unusual Event Detection," in *Proc. of IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, pp. 611-618, San Diego, California, June, 2005. Article (CrossRef Link).
[26] C. Tomasi, T. Kanade, "Shape and Motion from Image Streams under Orthography: a Factorization Method," *International Journal of Computer Vision*, vol. 9, no. 2, pp. 137-154, November, 1992. Article (CrossRef Link).
[27] J. Ferryman, A. Shahrokni, "PETS2009: Dataset and challenge," in *Proc. of 12th IEEE Int. Workshop on Performance Evaluation of Tracking and Surveillance*, pp. 355-360, IEEE, Snowbird, Utah, December, 2009. Article (CrossRef Link).
[28] UMN Dataset. [Online]. Available: http://mha.cs.umn.edu/movies/crowdactivity-all.avi.
[29] WEB Dataset. [Online]. Available: http://commerce.wazeedigital.com.
[30] S. Wu, H. S. Wong, Z. W. Yu, "A Bayesian Model for Crowd Escape Behavior Detection," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 24, no. 1, pp. 85-98, August, 2014. Article (CrossRef Link).
[31] G. G. Xiong, J. Cheng, X. Y. Wu, et al., "An energy model approach to people counting for abnormal crowd behavior detection," *Neurocomputing*, vol.83, no. 7, pp. 121-135, April, 2012. Article (CrossRef Link).
[32] Y. Cong, J. S. Yuan, J. Liu, "Abnormal event detection in crowded scenes using sparse representation," *Pattern Recognition*, vol. 46, no. 7, pp. 1851-1864, July, 2013. Article (CrossRef Link).

Mengdi Wang received her BS degree in internet of things engineering from Jiangnan University, China, in 2015. Now, she is currently pursuing her master degree in control science and engineering from Shandong University, China. She focuses on computer vision, pattern recognition, and abnormal event detection.

Dr. Faliang Chang received his BS, MS, and PhD degrees in automation from Shandong University, China, in 1986, 1989, and 2005, respectively. Now, he is a professor and PhD tutor in the School of Control Science and Engineering, Shandong University. His research interests are computer vision, pattern recognition, and image processing.

Youmei Zhang received her BS degrees in measurement and control technology and instrumentation from Shandong University, China, in 2013. Now, she is currently pursuing her PhD degree in pattern recognition from Shandong University, China. Her research interests are crowd behavior analysis and abnormal event detection.