Journal of Artificial Intelligence Practice (2023)
Clausius Scientific Press, Canada

DOI: 10.23977/jaip.2023.060508
ISSN 2371-8412 Vol. 6 Num. 5

Abnormal Event Detection and Localization Based on Crowd Analysis in Video Surveillance

Xu Xiao[1,a,*]

[1]*Department of Engineering Management, Jinan Engineering Vocational Technical College, Jinan, China*
[a]*xx19940811@126.com*
[*]*Corresponding author*

Keywords: Video surveillance, Abnormal event detection and localization, Collectiveness feature, Kinetic energy feature, Kanade-Lucas-Tomasi key point tracker

Abstract: With the rapid development of economy and urban construction, abnormal events detection has arouse spread attention in need of public security. When an abnormal event occurs, the pedestrians in crowd escape or run instinctively which will lead to sharp change in the collectiveness feature and kinetic energy of crowd. This paper proposes a method based on the Collectiveness Energy Index (CEI) which combines the two features mentioned above to detect the abnormal events because it is not unreliable to utilize either of the two features singly. Besides, this paper also presents a means to locate abnormal behaviours in the anomalous scenes. Firstly, we obtain spatial coordinate of particles existed on individuals in each frame using generalized Kanade-Lucas-Tomasi key point tracker (gKLT). Then, the Collectiveness Energy Index (CEI) of each frame is calculated and compared with an adaptive threshold for abnormal events identification. In order to locate abnormal behaviours, this paper splits each input frame of video sequences into blocks without overlapping and then calculates the velocity and individual collectiveness of each block for classifying it as anomalous or not. Experiments conducted on UMN dataset and UCSD dataset verify the effectiveness and superiority of our detection and localization method.

1. Introduction

With the rapid development of economy and urban construction, video surveillance technology has arouse spread attention in need of public security. As one of the main tasks of video surveillance, the detection and analysis of crowd abnormal events have become a hot research topic in recent years. The timely detection of abnormal events is important to protect the safety of people and public property.

Abnormal events are defined as those that deviate from regular pattern [1]. Generally speaking, abnormal events are caused by objects with unexpected appearance or motion patterns. Thus, abnormal event detection is equivalent to detecting the irregular motion or appearance of pedestrians or other non-human objects [2]. Recently, considering the outstanding performance of deep learning on various computer vision tasks [3, 4], a variety of deep neural networks such as deep convolutional autoencoder (CAE) [5, 6] and generative adversarial network (GAN) [7, 8] have been utilized for video anomaly detection. However, most of deep learning-based methods process the entire video sequences in an end-to-end manner, without modelling the local interaction relationship between objects in the scene, which causes the loss of spatial context in the video sequence.

To this end, we present a novel video anomaly detection strategy based on the detailed analysis of object interaction. All of objects under a surveillance scene can be regarded as a crowd and the collective motion is the most significant feature of the crowd systems. We utilize the collectiveness [9] for modelling the spatially coherent structure of collective motion, which illustrates one important structural property of collective motion: behaviour consistency remains high among individuals in local neighbourhood, while low among those that are far apart, despite in the same collective manifold. Therefore, the collectiveness measurement is effective for monitoring the transition of a crowd system from ordered to disordered states, which further is capable of detecting crowd anomaly such as escaping. As for the crowd anomaly without apparent order transition such as running suddenly, we integrate the collectiveness with kinetic energy (reflecting the motion state of crowd) for accurate detection. Besides, this paper investigate the individual collectiveness and motion velocity for local anomalous object identification.

The primary contributions of this paper are listed as below: 1) An algorithm that fuses the collectiveness measurement and motion state is proposed to detect various types of anomalies. 2) The proposed method is capable of detecting crowd abnormal events and locating individual anomalous action simultaneously; 3) Experimental results on two public benchmark datasets validate the effectiveness and robustness of the proposed method. The remainder of this paper is organized as follows: In section 2, we describe the detection method of abnormal events and present the localization scheme of abnormal behaviours in detail. In section 3, the experiment and evaluation results of two different datasets are provided to demonstrate the advantages of the proposed method. The final section gives a conclusion of the study.

2. Methodology

The system of this paper is shown in Fig. 1 and Fig. 2. We firstly use Kanade-Lucas-Tomasi key point tracker (gKLT) to extract individuals trajectories that are expressed by different spatial coordinates of particles existed on individuals at different time. After all required features are obtained, the abnormal events detection and anomaly localization are conducted.

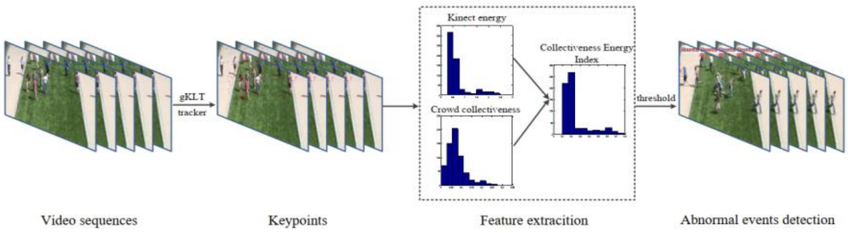

Figure 1: The framework of abnormal events detection.

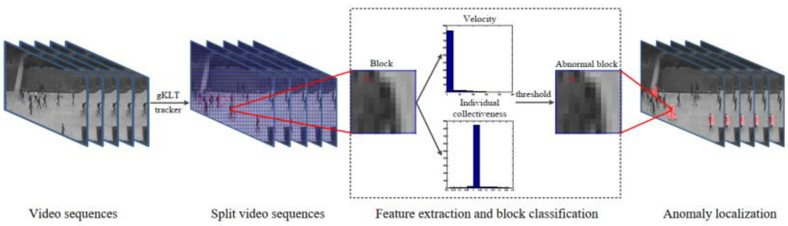

Figure 2: The framework of abnormal behaviours localization.

2.1. Individuals Tracking

All particles' spatial coordinates in each frame got by gKLT algorithm are expressed as follows:

$$H_t = \{h_{t1}, h_{t2}, \ldots h_{tI}\}, h_t = (X_{ti}, Y_{ti}, t), V_{tix} = X_{(t+1)i} - X_{ti}, \; V_{tiy} = Y_{(t+1)i} - Y_{ti}, \; V_{ti} = \sqrt{V_{tix}^2 + V_{tiy}^2}, \quad (1)$$

where H_t is the particles set at time *t*, (X_{ti}, Y_{ti}) is the coordinate of particle *i* at time *t*, *I* is the number of particles, V_{tix} and V_{tiy} are horizontal and vertical velocity of particle *i* respectively.

2.2. The Collectiveness Feature

A crowd is more than a gathering of individuals. Under certain circumstances, individuals in a crowd are organized into a unity with different levels of collective motions [9]. Collectiveness measures the behaviour similarity of an individual with others in its neighbour. It is calculated by the cosin of an angle between two adjacent particles' direction of velocity. Assuming *i* and *j* are two adjacent particles, the behaviour similarity (X_{ti}, Y_{ti}) between *i* and *j* is defined as:

$$w_t(i,j) = \begin{cases} C_t(i,j), \; if \; C_t(i,j) > 0 \\ 0, \; if \; C_t(i,j) < 0 \end{cases}, \; C_t(i,j) = \frac{V_{ti} \bullet V_{tj}}{\sqrt{|V_{ti}|^2 + |V_{tj}|^2}}. \quad (2)$$

Existing empirical studies of collective motion show that animals maintain local interaction among neighbours with a fixed number of neighbours on topological distance, rather than with all neighbours within a fixed spatial distance [10]. Therefore, the $w_t(i,j) \in [0,1]$ is only applicable when *i* and *j* are in a neighbourhood, it cannot accurately estimate the behaviour similarity when *i* and *j* are at a distance. We should calculate the behaviour similarity based on topological structure: path.

Suppose that $\gamma_l = \{p_0 \to p_1 \ldots \to p_l\}$ indicates a path of length *l* through $p_0, p_1, \ldots p_l$ between *i* and *j*. The path behaviour similarity based on γ_l is expressed as:

$$s_{\gamma_l} = \prod_{k=0}^{l} w_t(p_k, p_{k+1}). \quad (3)$$

We define a set P_l that contains all paths of length *l* between *i* and *j*, so the behaviour similarity of *l*-path is represented as:

$$s_l(i,j) = \sum_{\gamma_l \in P_l} s_{\gamma_l}(i,j). \quad (4)$$

After utilizing the generating function to make the sum of $s_l(i,j)$ converge, the individual collectiveness of particle *i* is denoted as:

$$CollectivenessSet(i) = \sum_{j \in C} \sum_{l}^{\infty} z^l s_l(i,j), \quad (5)$$

where *z* is a real-valued regularization factor, and z^l can be interpreted as the weight for *l*-path similarity. The individual collectiveness would not increase with *l* when *z* < *l*.

2.3. Abnormal Events Detection

2.3.1. Crowd Feature Extraction

The Kinetic Energy: The kinetic energy of crowd represents the intensity of the crowd's movement. We utilize the particles obtained by the gKLT algorithm to denote the pedestrians. Therefore, the kinetic energy of crowd equals to the summation of all particles' kinetic energy. According to the definition of kinetic energy in physics, the kinetic energy of crowd is expressed as:

$$E_k(t) = \frac{\sum_{i=1}^{I} m_i |V_{ti}|}{I}, \quad (6)$$

where m_i is the velocity weight of particle i, and m_i usually adopts the same constant. V_{ti} is the velocity of particle i at time t.

Crowd Collectiveness: The crowd collectiveness is defined as the mean of all particles' individual collectiveness at time t:

$$CrowdCollectiveness(t) = \frac{1}{I}\sum_{i=1}^{I} CollectivenessSet(i), \quad (7)$$

where $CollectivenessSet(i)$ is the individual collectiveness of particle i.

2.3.2. Collectiveness Energy Index

In some scenes, pedestrians walk randomly all the time and begin to escape suddenly such as the UMN dataset [11], and the crowd collectiveness maintain a low value from start to end. Thus, using the crowd collectiveness singly is unreliable to detect the abnormal events. While in some other circumstances, pedestrians in crowd move quickly and towards the same direction at the initial time. Then pedestrians begin to escape. It is apparent that the kinetic energy is high all the time so that utilizing the kinetic energy singly to detect the abnormal events is unsuitable.

In conclusion, we should utilize feature multiplication method to combine crowd collectiveness with kinetic energy of crowd, which is expressed by Collectiveness Energy Index (CEI). The value of the crowd collectiveness ranges from 0 to 1, which is far lower than that of kinetic energy. So it is necessary to normalize the kinetic energy before combination. The normalized kinetic energy is represented as:

$$E_k^n(t) = \frac{E_k(t) - \min_{1 \le t \le N} E_k(t)}{\max_{1 \le t \le N} E_k(t) - \min_{1 \le t \le N} E_k(t)}, \quad (8)$$

where N is the number of frames, $E_k^n(t)$ is the normalized kinetic energy

The Collectiveness Energy Index is defined as:

$$CEI(t) = E_k^n(t) \times |1 - CrowdCollectiveness(t)|, \quad (9)$$

where CEI(t) denotes the Collectiveness Energy Index at time t.

2.4. Abnormal Behaviours Localization

Abnormal behaviours includes crowd abnormal behaviours and individual abnormal behaviours.

In some cases, all of pedestrians would escape or run from dangerous region because of instinctive response when an abnormal event happens, which are considered as crowd abnormal behaviours. Therefore the localization of crowd abnormal behaviour equates to locate each escaping or running pedestrian. The main task of individual abnormal behaviour localization is to locate pedestrians or objects whose velocity and direction are obviously different.

We propose a novel method to locate the abnormal behaviours. Each frame of the video sequences is split into many blocks of size $M*M$, with block whose coordinate is (i, j) at frame t is represented by $B_t(i, j)$. The x-coordinate of blocks range from 1 to w/M, and the y-ordinates of blocks range from 1 to h/M, where w and h is the width and height of video frame.

2.4.1. Block Feature Extraction

Velocity: We estimate the velocity feature of block $B_t(i, j)$ by computing the optical flow between two consecutive frames using Lucas-Kanade algorithm, which is represented as:

$$V_t(i, j) = \sum_{n=1}^{N} \left\| \left(v_x^n, v_y^n \right) \right\|_2 , \tag{10}$$

where v_x^n and v_y^n are the optical flows of pixel n in the vertical and horizontal directions respectively, N is the total number of pixels within block $B_t(i, j)$.

Individual Collectiveness: Individual collectiveness indicates the consistency in movement direction of an individual with others. The individual collectiveness of block $B_t(i, j)$ is defined as:

$$I_t(i, j) = \sum_{n=1}^{M} CollectivenessSet_n(i, j), \tag{11}$$

where $CollectivenessSet_n(i, j)$ is the individual collectiveness of particle n with coordinate (i, j), M is the number of particles whose value of individual collectiveness are not equal to 0 within block $B_t(i, j)$.

For some big objects such as carts, the number of particles existed on them obtained by gKLT tracker is more than that of small objects and these particles have the same direction of motion. So the value of $I_t(i, j)$ belonging to block $B_t(i, j)$ that covers the big objects is larger than that of normal block

2.4.2. Block Classification

In order to locate the abnormal behaviours, we just need to determine that which block $B_t(i, j)$ is labelled as anomalous. Each block is examined by two classifiers. If the block is judged as anomaly by the first classifier, the second classifier is not carried out. Given certain velocity thresholds V, individual collectiveness threshold I. The block $B_t(i, j)$ is considered as anomalous if either of the following formulas is workable:

$$V_t(i, j) > V , \quad I_t(i, j) > I . \tag{12}$$

The first equation is a motion check, where objects with irregular speed are detected as anomalous such as the riding bike in the crowd. The second equation is an appearance check, where objects with irregular appearance are detected as anomalous such as the driving car in the crowd.

3. Experimental Results

Quantitative evaluation of our proposed method is conducted utilizing the accuracy index, which is calculated as below:

$$accuracy\left(ACC\right)=\frac{TP+TN}{N}, \tag{13}$$

where true positive (TP) is abnormal sample that is correctly detected, true negative (TN) is normal sample that is correctly identified, N is the total number of frames for the test video.

3.1. Crowd Abnormal Events Detection and Localization

The UMN dataset are shot in 3 different scenes, including lawn, indoor and plaza. In each scene, pedestrians walking randomly at the initial time is considered to be normal events and pedestrians fleeing suddenly is regarded as abnormal events. There are 11 abnormal events in total corresponding to 11 videos that are partitioned from the whole video set according to the 3 different scenarios. We select two videos from two typical scenes (lawn and plaza) in UMN dataset for experiments. Fig. 3 shows the CEI curves of the two videos.

From Fig. 3(a), the threshold is 0.3930, the CEI curve exceeds the threshold at 496th frame when the pedestrians start to escape. So the abnormal event occurs at the 496th frame. After 574th frame, the curve begins to below the threshold with pedestrians moving away from the field of view. Fig. 3(b) shows 185 the same tendency like Fig. 3(a). The threshold is 0.2838 and the abnormal event is detected at the 554th frame.

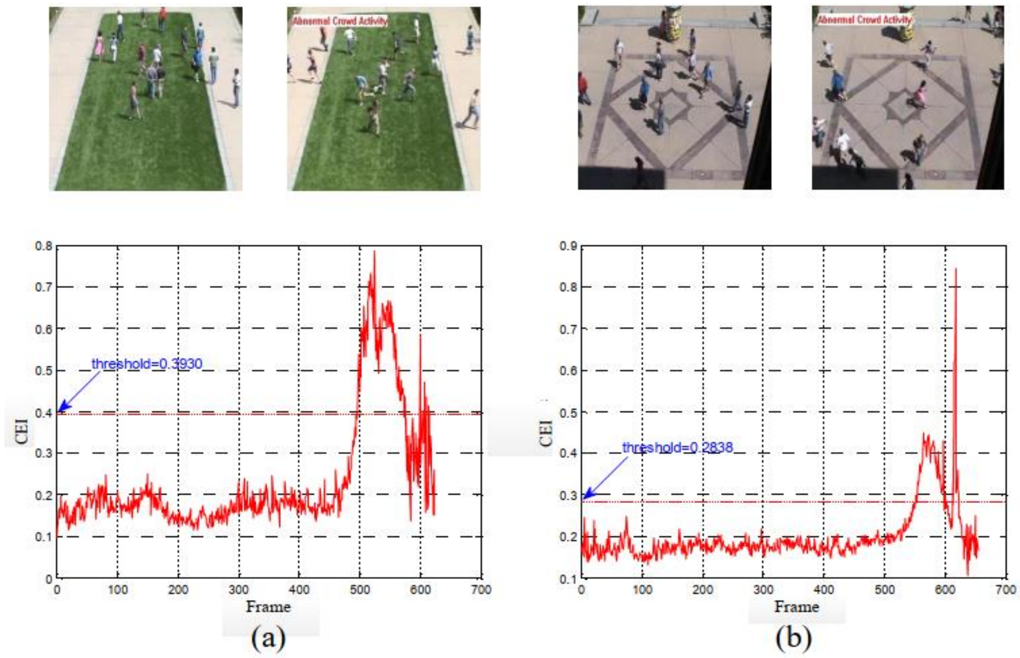

Figure 3: The CEI curve of two videos in UNM dataset.

We compare the proposed method with Collectiveness Model [9] (denoted as CM), Social Force Model [11] (denoted as SFM) and Energy Model [12] (denoted as EM) in identifying the crowd escape events. Table 1 shows the accuracy of the four methods for crowd escape event detection. Our method achieves an accuracy of 93.26%, which is higher than that of CM (81.60%), EM (89.66%) and SFM (87.62%).

Table 1: Accuracy (%) comparison of proposed method with CM, SFM and EM on UMN dataset.

Algorithm	CM [9]	SFM [11]	EM [12]	Proposed method
Video 1	79.44%	84.41%	90.79%	92.32%
Video 2	83.76%	90.83%	88.53%	94.2%
Average	81.60%	87.62%	89.66%	93.26%

3.2. Individual Abnormal Behaviours Localization

We carried out experiments on UCSD Anomaly Detection dataset [13] by the proposed localization algorithm. The dataset is acquired with a stationary camera and includes two subsets (Ped1 and Ped2) with diverse crowd density. The image size in Ped1 is 238×159 pixels, while on Ped2 it is 360×240. Anomalies commonly occurring in the dataset include: bikers, skaters, cars and other objects, while the videos that contain only pedestrians are regarded as normal. Some qualitative localization results of UCSD dataset are shown in Fig. 4, which confirms that the proposed method is capable to locate individual abnormal behaviours efficiently.

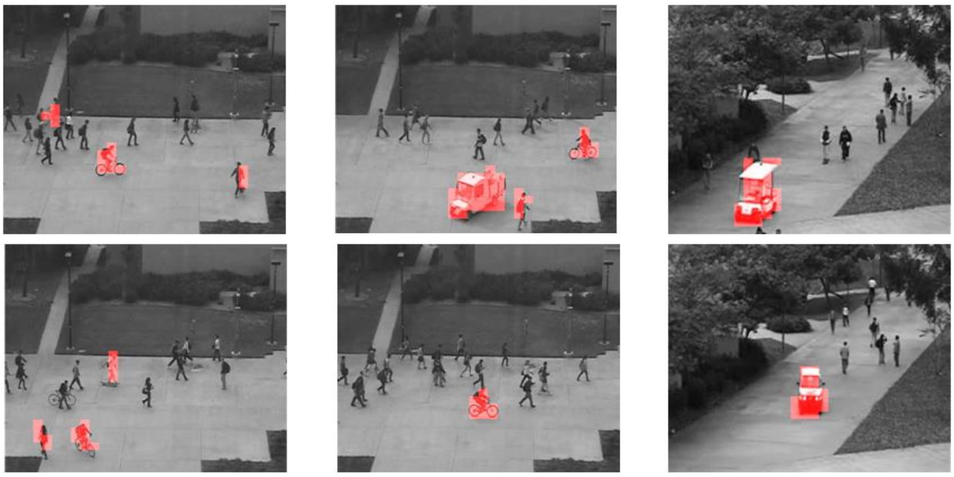

Figure 4: The localization results of UCSD dataset.

4. Conclusion

In this paper, we propose a brand new scheme for video anomaly detection based on the crowd interaction analysis. The proposed method employs the collectiveness measurement for modelling the spatially coherent structures of crowd, which is combined with the kinetic energy for multiple types of crowd anomalies detection. In addition, the motion velocity and individual collectiveness are jointly applied for local anomalies identification. Experiments are conducted on the UMN and UCSD datasets, and the testing results demonstrate that our method is effective for anomaly detection and localization simultaneously. Besides, the detection performance is superior to several compared methods in terms of accuracy. However, there are two factors that deteriorate the proposed method: 1) The gKLT tracker often tracks some particles that do not exist on objects, which may lead to producing wrong detection results; 2) The objects' size of the same kind vary greatly owing to the scene's perspective, and the equal-sized block structure in our method is not appropriate to deal with this issue.

References

[1] W. Luo, W. Liu, D. Lian, et al., Video anomaly detection with sparse coding inspired deep neural networks, IEEE

Transactions on Pattern Analysis and Machine Intelligence 43 (3) (2019) pp.1070-1084.
[2] Y. Zhang, H. Lu, L. Zhang, et al., Combining motion and appearance cues for anomaly detection, Pattern Recognition 51 (2016) pp. 443-452.
[3] W. Yu, H. Xu, Co-attentive multi-task convolutional neural network for facial expression recognition, Pattern Recognition 123 (2022) 108401.
[4] J. Li, X. Liu, W. Zhang, et al., Spatio-temporal attention networks for action recognition and detection, IEEE Transactions on Multimedia 22 (11) (2020) pp. 2990-3001.
[5] M. Hasan, J. Choi, J. Neumann, et al., Learning temporal regularity in video sequences, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 733-742.
[6] D. Xu, Y. Yan, E. Ricci, et al., Detecting anomalous events in videos by learning deep representations of appearance and motion, Computer Vision and Image Understanding 156 (2017) pp.117-127.
[7] W. Liu, W. Luo, D. Lian, et al., Future frame prediction for anomaly detection-a new baseline, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 6536-6545.
[8] M. Sabokrou, M. Khalooei, M. Fathy, et al., Adversarially learned one-class classifier for novelty detection, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 3379-3388.
[9] B. Zhou, X. Tang, X. Wang, Measuring crowd collectiveness, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3049-3056.
[10] M. Ballerini, N. Cabibbo, R. Candelier, et al., Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study, in: Proc. National Academy of Sciences 105 (4) (2008) pp. 1232-1237.
[11] R. Mehran, A. Oyama, M. Shah, Abnormal crowd behavior detection using social force model, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 935-942.
[12] G. Xiong, J. Cheng, X. Wu, et al., An energy model approach to people counting for abnormal crowd behavior detection, Neurocomputing 83 (2012) pp. 121-135.
[13] V. Mahadevan, W. L. V. Bhalodia, et al., Anomaly detection in crowded scenes, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2010, pp. 13-18.