CrossMark

A thermodynamics-inspired feature for anomaly detection on crowd motions in surveillance videos

Xinfeng Zhang [1] · Su Yang [1] · Yuan Yan Tang [2] ·
Weishan Zhang [3]

Abstract Identification of abnormal behaviors in surveillance videos of crowds plays an important role in public security monitoring. However, detecting abnormal crowd behaviors is challenging in that movements of individuals are usually random and unpredictable, and the occlusions caused by over-crowding make the task more difficult. In this paper, we introduce thermodynamic micro-statistics theory to detect and localize abnormal behaviors in crowded scenes based on Boltzmann Entropy. For this purpose, the scene of interest is modeled as moving particles turned out from a general optical flow algorithm. The particles are grouped into a set of prototypes according to their speeds and directions of moving, and a histogram is established to figure out how the particles distribute over the prototypes. Here, Boltzmann Entropy is computed from the histogram for each video clip to characterize the chaos degree of crowd motion. By means of such feature extraction, the crowd motion patterns can be represented as a time series. We find that when most people behave anomaly in an area under surveillance, the corresponding entropy value will increase remarkably in comparison with those of normal cases. This motives us to make use of Boltzmann Entropy to distinguish the collective behaviors of people under emergent circumstances from their normal behaviors by

✉ Su Yang
suyang@fudan.edu.cn

Xinfeng Zhang
10110240026@fudan.edu.cn

Yuan Yan Tang
yytang@umac.mo

Weishan Zhang
zhangws@upc.edu.cn

[1] Shanghai Key Laboratory of Intelligent Information Processing, College of Computer Science, Fudan University, Shanghai 201203, China

[2] Department of Computer and Information Science, University of Macau, Macau, China

[3] Department of Software Engineering, China University of Petroleum, Qingdao 266580, China

🄳 Springer

evaluating how significantly the current feature value fits into the Gaussian model of normal cases. We validate our method extensively for anomaly detection and localization. The experimental results show promising performance compared with the state of the art methods.

Keywords Boltzmann Entropy · Crowd · Collective behavior · Abnormal event detection · Anomaly detection

1 Introduction

With worldwide urbanization and population growth, public places such as railway stations, subways, and airports become highly crowded [16]. This brings in the risk to incur catastrophic events. For example, in Fig. 1, a surveillance video from Tencent News shows the anti-government activity in Bangkok in January 2014. The video faithfully records an emergency of bomb explosion among the crowd and the consequent abnormal collective behaviors. The frequently occurred safety problems have attracted much attention to crowd surveillance in terms of both sociology [10] and technological issues [7]. Automatic detection of abnormal events has the potential to provide alarm or assistance to human observers, which can greatly save operators' valuable attention when they face a large number of monitors for long hours [7].

The abnormal behaviors in crowded scenes usually appear as crowd commotion, for example, the crowd commotion following the occurrence of the explosion as shown in Fig. 1, so that anomaly detection is in general a problem of detection of crowd commotion [14, 30]. The traditional object-based approaches treat crowd as a collection of individuals. As this kind of methods describe group behaviors through individuals, its performance is subject to the precision of object segmentation [25] and tracking [2, 19, 27]. In low-density scenes, object-based approaches may perform well. In crowded scenes, such as the case shown in Fig. 1, since a large number of people perform random moving in a small space, the high density of objects and the occlusions in both static and dynamic senses will degrade the accuracy of object segmentation and tracking severely, which makes it impossible to capture the behavior of a single individual [11].

To avoid the difficulty of segmenting individuals, a possible solution is to model the crowd commotion of large population as the spatial-temporal movement of a large number of moving particles. In such a context, the well-known social force model [9] is proposed to simulate the movement of a large number of persons as moving particles by taking into account the constraints arising from the mutual interactions among them [14, 17], where the particles are computed from the optical flows. Although it aims to model the physical nature of crowd

Fig. 1 An example of the surveillance video showing the anti-government event at Bangkok with grenade attacks. The explosion position is marked by the *red circles*. The displayed frames successively show the process of the attack: Normal state of people's moving, explosion of grenade, people escaping in panic, and the final crowd evacuation

commotion, the social force model is still far from revealing the reality of people's responses under emergent circumstances due to the inaccurate representation of individual objects in terms of computer vision, which deviates from the physical nature of the original social force model [9]. Aside from that, the social force model has the emphasis on the local physical phenomenon such as the interaction between neighboring elements but fails to figure out the evolution of crowd commotion from a global point of view, which accounts for why its performance is not satisfactory.

The main contribution of this study is: By employing optical flows as low-level representation, we characterize complex crowd commotions from a global point of in terms of thermodynamics. Intuitively, crowd motions in surveillance videos are comparable to heat motions of basic particles such as molecules in terms of thermodynamics. This motivates us to investigate into thermodynamics to seek effective measures for characterizing crowd motion. Boltzmann Entropy is originally proposed in thermodynamics to investigate the statistical regularity of heat motion of molecules [8]. It captures the relation between the disorder of the motion of a mass of microscopic molecules and the chaotic degree of the corresponding macrostate, namely, the motion pattern of the group of interest as a whole. Inspired by Boltzmann Entropy, we introduce a computer vision method to detect and localize abnormal crowd commotion. The Boltzmann Entropy-based method applied in this study can figure out the spatiotemporal motion pattern of a set of particles in the sense of statistics so that it avoids segmenting or tracking of objects in high-density crowd videos. Figure 2 summarizes the main steps of the approach. As observed experimentally in this study, when abnormal crowd movement occurs, Boltzmann Entropy will increase correspondingly, so abnormal events can be detected by applying a threshold to the time series of Boltzmann Entropy without complex machine learning or large volume of training data. For decision making, here, we need only a simple Gaussian model established from the mean and standard deviation of the time series of the Boltzmann Entropy in correspondence with normal cases. High detection rate and low false alarm rate are achieved on the University of Minnesota (UMN) and web datasets by using the proposed method. The experiments demonstrate not only the effectiveness of the proposed method in detecting and locating abnormal behaviors in crowded scenes but also the efficiency for online monitoring.

The remainder of the paper is organized as follows. In Section 2, we review the related works in the literature. In Section 3, we introduce Boltzmann Entropy to describe the chaotic degree of the crowded scenes in terms of statistics, and then the anomaly detection based on Boltzmann Entropy. In Section 4, we evaluate the performance of the proposed method in detecting and locating abnormal behaviors. In Section 5, we draw conclusions.
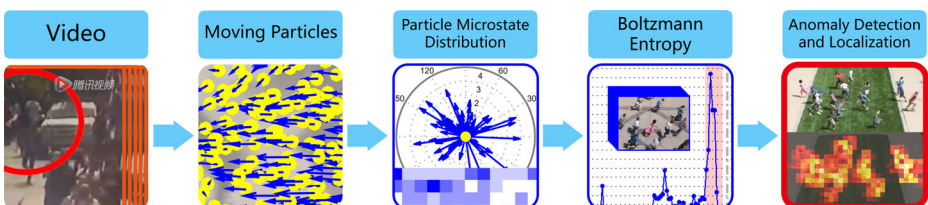

Fig. 2 The pipeline of the proposed method

2 Related work

The traditional object-based approaches treat crowd as a collection of individuals and describe the collective behaviors of a group through individuals, so their performance depends highly on object segmentation [25] or trajectory reconstruction [2, 19, 27]. In low-density scenes, object-based approaches can achieve good performance. However, in crowded scenes such as the video frames shown in Fig. 1, when a large number of persons perform random actions in a small space, the high densities of objects and severe occlusions make object segmentation and tracking impossible in terms of capturing the behavior of every individual [11]. Besides, the computational cost should be considerably high.

To tackle the problem of unreliable object segmentation or tracking in crowded scenes, an emerging trend for abnormal event detection is to model global motion patterns from local motions of small cells, namely, pixels, local image blocks/patches, and local 3D cuboids/ bricks. According to different features applied, these approaches can broadly be classified into three categories: Cell interaction-based approaches [12, 14, 17, 29], cell trajectory-based approaches [20], and cell property-based approaches [1, 4, 13, 18, 22, 28, 32].

For cell interaction-based approaches, a well-known work is the so-called social force model (SFM) proposed by Mehran et al. [14], where group actions are modeled as interaction forces of subjects computed from optical flows. However, SFM is easy to incur false alarms, which are usually caused by the error-prone computation of local interaction forces. The particle swarm optimization (PSO) method introduced by Raghavendra et al. [17] is a robust algorithm to optimize SFM via selective particles but it is difficult to apply such method to online surveillance due to the time-consuming optimization. Cui et al. define interaction energy potential to represent the current state of a subject based on the positions and velocities of itself and its neighbors [29]. The energy potential among neighbors and the action of the subject of interest in the sense of velocity are used to model abnormal group behaviors. However, the weakness of SFM, say, the unstable relationships in terms of local interactions remains unsolved in [29]. Kaltsa et al. deploy compact swarms over optical flows, whose positions are continuously updated following the rules of interaction with each other as well as relationship with regard to the optical flow magnitude [12]. Then, the histograms of oriented swarms constructed from the swarms' positions and the histograms of oriented gradients [5] are combined to detect anomalous regions via support vector machine (SVM). The fundamental principle of [12] is similar to SFM as well as PSO optimized SFM such that it inherits the same limits, that is, error-prone local interaction modeling and computationally high cost.

For cell trajectory-based approaches, Wu et al. [20] make use of chaotic invariants of particle trajectories as feature, which is known as maximal Lyapunov exponent and correlation dimension, to detect and localize anomalies in crowded scenes. When people's movements are spatially constrained such as in corridor and underpass, however, discriminating normal and abnormal events according to trajectories is extremely difficult due to the violation of the assumption in chaos theory.

For cell property-based approaches, the fundamental is to characterize motion patterns of crowded scenes based on optical flow computing [1]. Mahadevan et al. employ a mixture of dynamic textures to represent jointly the appearance and the dynamics of local portions of videos in crowded scenes [13, 28]. This approach has an outstanding performance in detecting temporal and spatial anomalies at the cost of heavy load of computation. Reddy et al. propose multiple features including motion, size, and texture from each foreground cell and the decision resulting from each feature is fused to make the final decision [18]. Here, to guarantee

computational efficiency, only foreground pixels are considered in modeling crowd motions. Yang et al. detect abnormal events by computing the sparse reconstruction cost based on the dictionary consisting of multi-scale histograms of optical flows in normal cases [4, 32]. Wu et al. construct the class-conditional probability density functions of optical flows in terms of position, magnitude and direction with regard to non-escape and escape cases, respectively [22]. Then, they detect crowd-escaping behaviors by using the class-conditional probability density functions in the framework of Bayes decision.

The features describing crowd actions used in most of the aforementioned methods are high-dimensional features. This means that a complex learning strategy and a large number of training samples are needed due to the sparseness caused by the high dimensionality of the data. In general, the learning process is time-consuming and the classification performance is subject to what training samples are available. Provided the training samples are inadequate or not representative, the detection precision will become worse remarkably.

In contrast to the high-dimensional features that require complex machine learning for pattern classification, an alternative solution is to develop low-dimensional features [23, 24, 30] that do not need an explicit learning procedure. For such features, the decision-making is based on watching whether the target feature value exceeds a predefined threshold. Cao et al. combine crowd kinetic energy and direction variation computed from optical flows to detect abnormal events in crowded scenes [24]. Xiong et al. propose an energy model to detect two types of abnormal crowd behaviors, people gathering and running [30]. Here, people counting is necessary for detecting the two types of abnormal behaviors but the number of people in crowded scenes is difficult to be estimated. In [15], grid-based motion features addressing the speed and direction of moving objects are utilized to detect abnormal events in crowded scenes. Since grid-based computation of motion features is so coarse and thus error-prone, the crowd flux estimation subject to it should be correspondingly unstable. Gu et al. use the product of Shannon Entropies of particle distributions in the vertical and horizontal directions to estimate the parameters of Gaussian Mixture Model (GMM) over the normal crowd behaviors and apply the GMM to detect anomalies [6]. The limitation of Shannon Entropy is that it concerns only the probability of moving particles' distribution in the whole plane but misses to count the total amount of the subjects involved in crowd motion. Unfortunately, the chaos degree of crowd motion is subject to the number of the subjects involved. Susan et al. [23] adopt the non-extensive entropy, which is a variant of Shannon Entropy [21], to detect anomaly in motion vector fields over three consecutive frames. Since non-extensive entropy is approximately the exponent of Shannon Entropy, it cannot overcome the limitation of Shannon Entropy, that is, it concerns only the probability of particle distribution but misses the total amount of the subjects involved in crowd motions, which is easy to cause false alarms.

In sum, the problems for the state of art methods are: Some features are not simple enough and complex machine learning on high-dimensional data is needed. This makes such methods not only time-consuming but also not applicable as a general solution due to the example-specific machine learning. Some other features are simple and efficient enough to ensure fast decision making such as the Shannon Entropy feature. However, such features do not promise robust performance under various circumstances. The major reason is that all these features do not reflect the physical nature of crowd motions. In view of the limits of the aforementioned methods, we propose to make use of Boltzmann Entropy in terms of thermodynamics as a feature for anomaly detection on collective human mobility. The initial work of this idea is presented in [33]. Here, we extend the work to include more details with rigorous and extensive validation of the method.

The proposed feature gains advantage over the aforementioned methods in the following aspects: Due to the physical nature of Boltzmann Entropy, the proposed feature describes the chaos degree of crowd motions with explicit physical meanning. This accounts for why it leads to better performance in the experiments. In the meantime, the efficient computation of the feature promises real-time on-line monitoring. Besides, the proposed method works no matter how many persons are in the present of the crowded scenes because it roots in the statistics of optical flows rather than object detection or tracking. In this sense, the proposed feature overcomes the limit of Shannon Entropy and its variants, which miss to distinguish the volume of the crowd involved in abnormal behaviors. To the best of our knowledge, the proposed feature is specific in that it can promise both accuracy in anomaly detection and efficient computation. In contrast to [33], this study includes more details as follows: An extensive survey of the state of the art, a more detailed description of the feature extraction method with formal mathematical definition and the motivation rooting in physics, the decision making method to identify anomaly patterns, evaluation of the discriminant power in terms of outlier degree and feature ranking measure, and more experimental comparisons with more baseline methods based on both the UMN benchmark and some additional real-world videos. To highlight the background as well as the motivation of this study, the state of the art works are summarized in Table 1 for comparison.

3 Boltzmann Entropy to feature crowd motion

Boltzmann Entropy is a quantitative measure of disorder degree of thermal motion of a large number of basic particles such as molecules and atoms. Motivated by the physical nature of Boltzmann Entropy in terms of thermodynamics, when monitoring collective crowd motion in video surveillance, the optical flows computed from continuous frames are regarded as moving particles and Boltzmann Entropy is employed to figure out quantitatively how chaotic or disordered the collective motion of a group of people of interest is. As shown in Fig. 3, the particles computed from the optical flow cover most of the motion field, which characterize the motion pattern of people in a scene. Since the Boltzmann Entropy computed from continuous video frames forms a time series, the anomaly detection problem can thus be solved with ease by applying simply a threshold to the time-varying Boltzmann Entropy measure that characterize the group motion, and capturing the remarkably sudden changes as indication of possible abnormal events.

As outlined in Fig. 2, the computation of the feature includes three steps: (1) Obtaining particles as all the moving regions and pixels computed from optical flows in continuous video frames; (2) Building a histogram to figure out the distribution of particles' moving directions and speeds; (3) Calculating the Boltzmann Entropy by taking into account every possible configuration of particles' motion states that can result in such a histogram. The scheme is detailed as follows:

3.1 Motion signature of crowd as a histogram of particles' moving direction and speed

A video is comprised of a series of clips and each clip consists of a couple of frames streaming over time. Here, each clip is represented by a matrix of $w \times h \times t$ size, where $w \times h$ denotes the frame resolution (width by height) and t is the number of sequential frames. We see each clip

Table 1 Summary of the state of the art

Categories		Features	Characteristics
Traditional object-based approaches		Treat crowd as a collection of individuals	**Weakness:** Unreliable object segmentation or tracking in crowded scenes; Impossible to capture the behavior of every individual.
Cell-based approaches	Cell trajectory-based approaches	Obtain trajectories as evolutionary cell positions	**Weakness:** Difficult to capture the trajectories; Precondition of chaos theory destroyed in case of spatially constrained motions.
	Cell interaction-based approaches	Calculate mutual interactions among neighboring elements	**Weakness:** The relationships of local interactions are unstable and error-prone; The models of interactions are complex and deviate from the original physical nature.
	Cell property-based approaches High-dimension	Texture; Histogram of optical flows; Dynamic textures.	**Weakness:** Complex learning strategy and a large number of training samples needed; Performance depends heavily on example-specific machine learning, not generic solution.
	Low-dimension	Kinetic energy; Potential energy; Entropy.	**Strength:** Complex machine learning free and large volume of training data not necessary; **Weakness:** Discriminant power is a challenging issue for originally low-dimensional representations such that it is critical to figure out the actual physical nature of crowd commotion.
		This work	**Physically sound; Only 1-dimensional measure; Less learning; High precision; Real time; Generic solution across scenes.**

Fig. 3 The up-left images are two continuous frames used to calculate the optical flow. The bottom-left image shows that particles computed from the optical flow cover the motion field. The right shows the enlarged view of the corresponding region

as a dynamic system, which captures the motion pattern of the crowd of interest in a short time window. In general, an abnormal event will last for a time over a couple of frames at least, so the collective behavior of crowd should be described in the context of a clip. This is why we let t frames constitute a clip and let clip be the basic unit to perform pattern analysis.

In each frame, an isolated pixel or a couple of pixels connected to form a region with nonzero optical flow values are referred to as moving particles or particles for brief. In contrast, the pixels with the optical flows of zero values can be considered as the background. On a micro level, the mobility of a particle is defined as the velocity $\mathbf{V}_i = (v_i^x, v_i^y)$, where

$$|\mathbf{V}_i| = \sqrt[2]{(v_i^x)^2 + (v_i^y)^2} \tag{1}$$

and

$$\theta_i = \arctan \frac{v_i^y}{v_i^x} \ (\theta_i \in [0, 2\pi)) \tag{2}$$

represent the magnitude and direction of the velocity of a moving particle, respectively.

To figure out the mobility of the particles in each clip in a statistical sense, we establish a histogram with $m_M \times m_A$ bins of equal intervals in both the magnitude and the angle of the vector of velocity. Here, m_M represents the number of the magnitude intervals and m_A that of the angle intervals. We assign the particles into such $m_M \times m_A$ bins according to their velocity magnitudes and directions. When computing the histogram, we fix the upper bound of the magnitude of the velocity in accordance with the limited physical capability of human beings' mobility. If the speed of a moving particle is greater than this value, we will confine it under the upper bound. After accumulating the particles of every frame in a given clip into one histogram, we obtain a histogram with N bins, where we let k_n denote the number of the particles falling in the nth bin for $1 \leq n \leq N$. Figure 4(a) shows an example, where a large number of particles are projected into 3×8 bins. To allow a more straightforward view of the

Fig. 4 **a** Partition of bins with
regard to magnitude and direction
of velocity of particles. **b**
Histogram of particle distribution

(a)   (b)
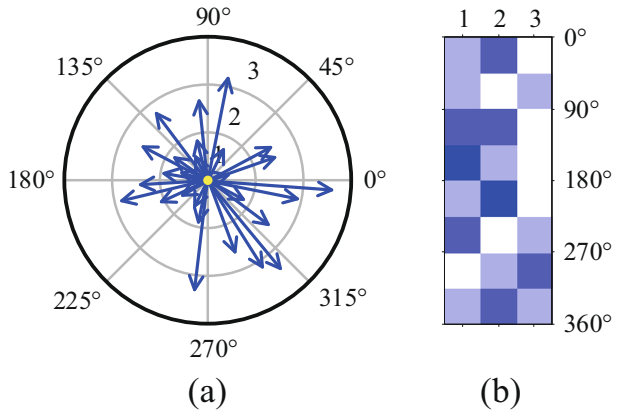
histogram, we illustrate a block with deeper blue color in Fig. 4(b) if the corresponding bin
contains more particles in Fig. 4(a). Every bin in the histogram as shown in Fig. 4 contains the
particles with similar moving speeds and directions, which is one prototype of the particle
motion pattern. Intuitively, the histogram can reveal the collective motion pattern of all
particles from a global point of view by means of counting in a statistical sense how the
moving particles of interest distribute over a couple of motion prototypes. Note that the
histogram utilized in this study is different from what is applied in [30], where motion vectors
of foreground pixels are decomposed into vertical and horizontal directions such that two
histograms are constructed, respectively, for projection along either direction.

3.2 Boltzmann Entropy

Although histogram is a widely used feature extraction means in terms of computer vision, the
spatial variation and high dimensionality make the aforementioned histogram-based pattern
description not suitable for discerning the distinction between normal and abnormal collective
motion patterns. According to our observation, in response to emergency events, most people
will behave in an abnormal manner with a highly disordered mobility pattern from a global
point of view. This motives us to make use of Boltzmann Entropy [8] in distinguishing the
collective behavior of people under emergent circumstances from what they behave in normal
cases since Boltzmann Entropy is a measure of the disorder degree of a whole system from the
perspective of molecule activities in terms of thermodynamics. In the following, we present
how to compute Boltzmann Entropy from a given histogram with N bins, namely $\{k_n : 1 \leq n \leq N\}$, where k_n denotes the number of particles falling into the nth bin. Note that for the total K
particles, $K = \sum_{n=1}^{N} k_n$, there are many possible arrangements of them that can lead to the same
histogram. For the example shown in Fig. 5, we illustrate the moving directions and speeds of
3 particles as well as the corresponding histogram, where we use red, green, and blue to mark
the 3 particles, respectively. It is apparent that the same histogram could result from a couple of
different configurations of the particles of interest in terms of moving direction and speed.
Counting all the cases to result in the given histogram, we obtain

$$W = C_K^{k_1} \cdot C_{K-k_1}^{k_2} \cdots C_{K-k_1-\cdots k_{N-1}}^{k_N} \tag{3}$$

Fig. 5 In the left part, the arrow in each color represents the moving direction and speed of one particle and there are 3 particles arranged in 6 manners in this example. The histogram in the right part figures out the particles' moving directions and speeds
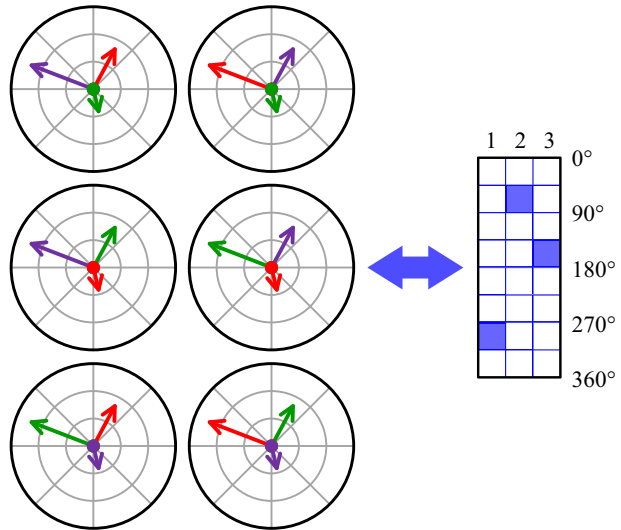
where C_P^Q represents the combination of Q elements taken from the P elements in total. As an example illustrated in Fig. 5, in total, six arrangements of the three particles are able to produce the same one histogram, so the total number W of all the possible particle arrangements computed from the histogram is 6. By multiplying the natural logarithm of W with the so-called Boltzmann constant k_B, Boltzmann Entropy is then defined as:

$$BE = k_B \cdot \ln W \qquad (4)$$

It is obvious that the Boltzmann Entropy BE is totally determined by W. W is computed in a statistical sense and it reflects the disorder degree of the mobility of all the particles under surveillance. A larger W indicates high disorder degree of the movements of the particles. In contrast, a small W means highly ordered mobility pattern of the particles. An extreme case is: When all particles move in the same direction with the same speed, the value of W is 1, that is, only one prototype of particles' motion exists. Based on the definition of Boltzmann Entropy in the context of thermodynamics, we reformulate it for video processing as follows:

$$BE = k \cdot \ln \left(\prod_{n=1}^{N} C_{K-\sum_{j=1}^{n-1} k_j}^{k_n} \right) \qquad (5)$$

where we let $k = 1.38 \times 10^{-5}$.

3.3 Boltzmann Entropy against Shannon Entropy

It is known that there are many definitions regarding entropy. In the following, we will compare Boltzmann Entropy with Shannon Entropy to justify that the pattern descriptor rooting in thermodynamics simulates and characterizes crowd mobility better due to the similar physical mechanism.

Except for Boltzmann Entropy in the literature of thermodynamics, more widely used entropy in the context of information sciences is Shannon Entropy, also referred to as Information Entropy [21]. Shannon Entropy is a measure of uncertainty regarding a random variable. Consider a given histogram $\{k_n : 1 \leq n \leq N\}$ with the associated probabilities $\{p_n = \frac{k_n}{K} : 1 \leq n \leq N\}$, where k_n denotes the number of particles falling into the nth bin, N is the number of histogram bins, $K = \sum_{n=1}^{N} k_n$ denotes the total number of particles, $p_n \in [0,1]$, and $\sum_{n=1}^{N} p_n = 1$. The expression of Shannon Entropy is the expected value of $-\log_2 p_n$. The formal definition is

$$SE = \mathrm{E}(\mathrm{I}(p_n)) = \sum_{n=1}^{N} p_n \cdot \mathrm{I}(p_n) = -\sum_{n=1}^{N} p_n \cdot \log_2 p_n \tag{6}$$

In the case of $p_n = 0$, the value of the corresponding item $0\log_2 0$ is set to be 0. If Shannon Entropy is zero ($SE = 0$), it means that there is no uncertainty. Because $p_n \in [0,1]$, Shannon Entropy is always non-negative ($SE \geq 0$).

Almost all of the previous entropy-related approaches for anomaly detection [6, 23] are based on Shannon Entropy and its variant. An example is as follows: A non-extensive entropy is used in [23] for increasing the nonlinearity of the exponential terms, which is defined in a quadratic exponent form.

$$NE = \mathrm{E}(\mathrm{I}(p_n)) = \sum_{n=1}^{N} p_n \cdot \mathrm{I}(p_n) = \sum_{n=1}^{N} p_n \cdot e^{-p_n^2} \tag{7}$$

We compare Shannon Entropy with Boltzmann Entropy in the following two tests. First, as shown in Fig. 6(a) and (b), we form two histograms with respect to the magnitude and direction of the velocity of particles and then compute the entropy values from each histogram using Shannon Entropy and Boltzmann Entropy, respectively. Based on this example, we

Fig. 6 Shannon Entropy and Boltzmann Entropy for two histograms

(a)   (b)
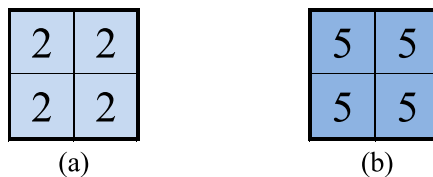
$$SE(a) = SE(b)$$

$$= -\frac{1}{4} \cdot \log_2 \frac{1}{4} - \frac{1}{4} \cdot \log_2 \frac{1}{4} - \frac{1}{4} \cdot \log_2 \frac{1}{4} - \frac{1}{4} \cdot \log_2 \frac{1}{4}$$

$$= \log_2 4$$

$$BE(a) = k \cdot \ln(C_8^2 \cdot C_6^2 \cdot C_4^2 \cdot C_2^2) = k \cdot \ln(2520)$$

$$BE(b) = k \cdot \ln(C_{20}^5 \cdot C_{15}^5 \cdot C_{10}^5 \cdot C_5^5) = k \cdot \ln(1.1733 \times 10^{10})$$

$$BE(a) < BE(b)$$

demonstrate the primary distinction between Boltzmann Entropy and Shannon Entropy. Second, we calculate the entropies of every clip in the 11 video sequences of three scenes of the UMN dataset. The range of each abnormal event is labeled as red shadow as shown in Fig. 7. The waveforms of the calculated entropy values are also shown in Fig. 7.

According to the results of the above two tests, we find that Shannon Entropy and its variants are limited in the following aspects: (1) Shannon Entropy is not subject to the total number of particles but fully determined by the ratio of the particle number in every bin to the total number of particles. As shown in Fig. 6(a) and (b), the Shannon Entropy value remains the same as $SE(a)=SE(b)$ for the two cases, where the particles are uniformly distributed in every bin but the total number of particles is different. (2) As shown in Fig. 7, sometimes Shannon Entropy reaches maxima when the pedestrian evacuation is going to finish, which can be observed in the video examples attached to the corresponding green maxima points. The reason is: Due to the limit as shown in Fig. 6, Shannon Entropy and its variants cannot distinguish the histograms resulting from the motion of a small group of people from those corresponding with the motion of a large group of people. That is, even when only a few people are in fast moving state, the corresponding Shannon Entropy could be unreasonably high. This does not meet the requirement of video surveillance since the abnormal activities of a large number of persons should attract much more attention in practice. Therefore, Shannon Entropy and its variants are not suitable for monitoring the crowd movement in a whole scene.

In contrast, Boltzmann Entropy does not suffer from such shortcomings as Shannon Entropy. Boltzmann Entropy is sensitive to the particle number as $BE(a)<BE(b)$ in Fig. 6. Even when the distribution of the particles follows the same probability density in the histograms, more particles taking part in the crowd motion will lead to higher value of Boltzmann Entropy. In Fig. 7, the waveform of Boltzmann Entropy exhibits a Gaussian function like shape when abnormal events occur. At the beginning of the crowd commotion, when some people are trying to leave their positions with atypical motions, correspondingly, the calculated Boltzmann Entropy begins to increase rapidly. When almost all the people under surveillance are taking part in the fast moving, the value of Boltzmann Entropy reaches the peak accordingly. When the pedestrian evacuation is finished, the Boltzmann Entropy value becomes extremely small and tends to be zero. This shows that Boltzmann Entropy can reflect the disorder degree of crowd movement in a whole scene, and the co-occurrence between crowd commotion and peak value of Boltzmann Entropy is obvious. This enables a simple but effective way to detect abnormal crowd commotion by applying just a threshold to the waveform of Boltzmann Entropy, which avoids error-prone and time-consuming object detection and tracking, or modeling complex interactions among objects.

3.4 Anomaly detection

Suppose that BE_i is the Boltzmann Entropy value for the ith clip. It is obvious that $\{BE_1, BE_2, \cdots, BE_T\}$ forms a time series provided the video sequence has T clips in total. As a result, the difficult problem of abnormal event detection in video surveillance is converted into a very simple problem of identifying outlying values in a time series. As mentioned previously, for a video clip, a greater W or BE corresponds with a higher disorder degree of the group motion, which could indicate the panic behavior of people in emergency cases. The evolution of crowd motion is usually as follows: As normal cases dominate the major portion of surveillance video
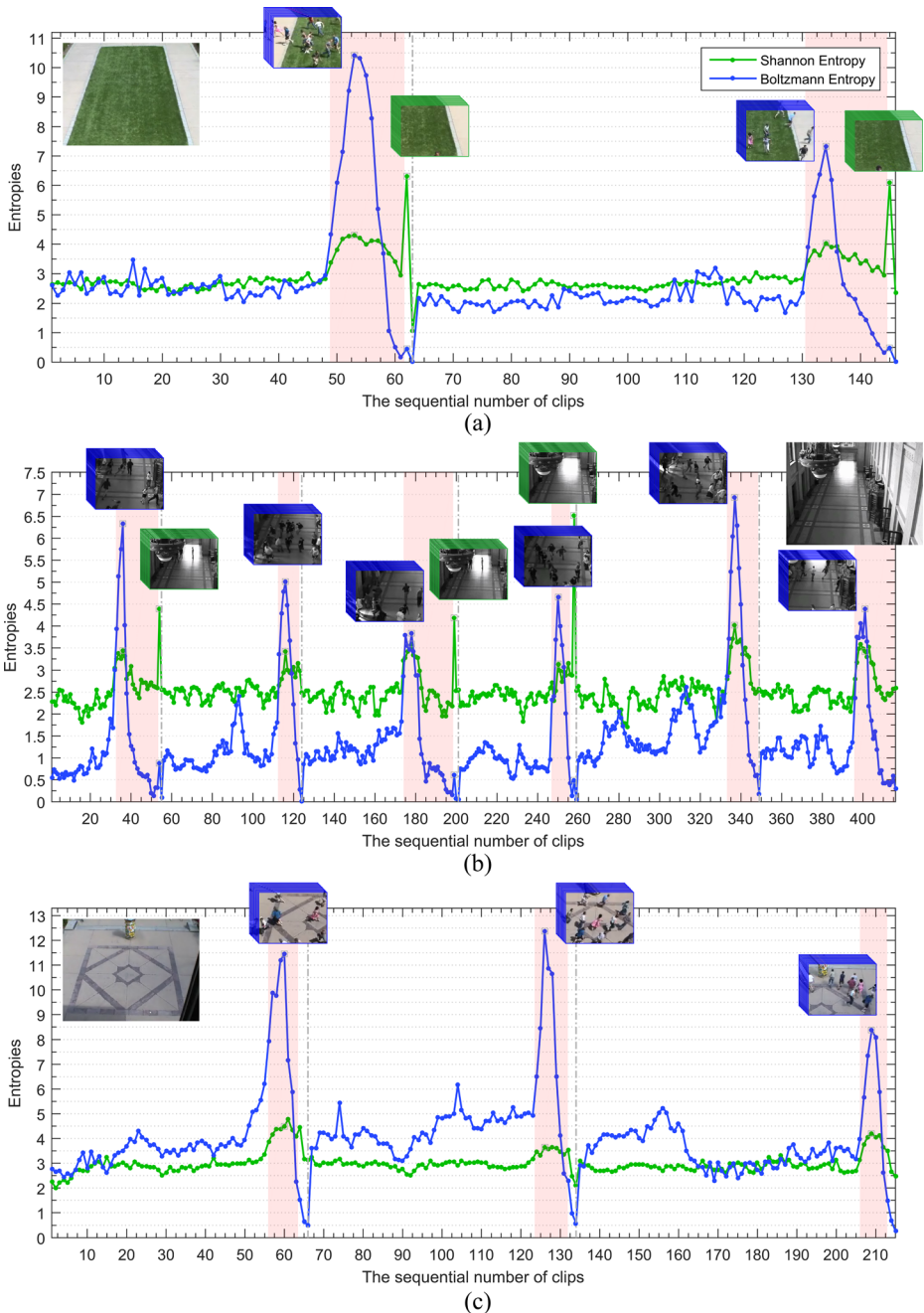
Fig. 7 Curves of entropy values against clips for the three scenes of the UMN dataset. The *blue curve* represents Boltzmann Entropy and the *green* one Shannon Entropy. The shadow in *red color* means anomaly. The online monitoring images are attached to the peaks of the curves

sequences, people should behave normally with relative low W or BE. When an emergent event takes place, people will react to it and their collective behavior will turn from normal pattern to abnormal pattern in a very short time. This causes the feature value W or BE to increase suddenly and drastically. Correspondingly, there should appear a turning point in the time series $\{BE_1, BE_2, \cdots, BE_T\}$. To capture such turning points, we further process the original time series into a difference sequence

$$\{0, (BE_2-BE_1), \cdots, (BE_T-BE_{T-1})\} = \{d_1, d_2, \cdots, d_T\} \tag{8}$$

where the difference between the entropy of the present clip and that of the previous clip is computed. In normal cases, the difference value between two consecutive clips should be small. In abnormal cases, such difference value will become remarkably higher than that in normal cases. It is known that abnormal events occur rarely. Thus, we need to establish a model corresponding with normal cases at first, and then detect abnormal event as those of low probability under the model. Here, we establish a Gaussian model $N(\mu, \sigma^2)$ from $\{d_1, d_2, \ldots, d_M\}$ of normal cases with the mean and standard deviation to be

$$\mu = \frac{1}{M} \cdot \sum_{i=1}^{M} d_i \tag{9}$$

and

$$\sigma^2 = \frac{1}{M-1} \cdot \sum_{i=1}^{M} (d_i-\mu)^2 \tag{10}$$

The Gaussian model $N(\mu, \sigma^2)$ is applied to calculate the probability that every entropy difference value may appear under the constraint of such model. Since the decrease of Boltzmann Entropy means that group motion becomes more ordered, we exclude the clips with negative entropy difference values directly without any further decision making on them. To enable reliable decision making, we perform smoothing as follows:

$$\text{If } 5d_i < BE_i-BE_{i-2}, (d_i + BE_i-BE_{i-2})\Big/2 \rightarrow d_i \tag{11}$$

4 Experiments

The method is tested on the publicly available dataset of normal and abnormal videos from the University of Minnesota (UMN) [26], and the crowd motion videos collected from Internet by us. To quantitatively evaluate different methods in anomaly detection, we manually mark the ground truth for all the videos following the guidelines that abnormal crowd motion is a small-probability event. We take 10 frames as a clip and there is no overlap between two successive clips. Hence, a clip is a spatial-temporal matrix of $w \times h \times 10$, where $w \times h$ represent the resolution of each image in the video. As observed in the experiments, the proposed method yields similar results when adopting different optical flow algorithms such as Brox, Farneback, PyrLK, and Horn-Schunck. Here, we use the Brox algorithm [3]. In the following, we will evaluate the discriminant power of Boltzmann Entropy [8] as a feature for anomaly detection, and then compare the proposed method with the methods based on pure optical flow, social force model (SFM) [14], particle swarm optimized-social force model (PSO-SFM) [17], chaotic invariants

[20], sparse reconstruction cost [32], Shannon Entropy [21], and non-extensive entropy [23] on the UMN dataset. Furthermore, we examine the performances of various low-complexity real-time methods using the more complex web dataset in terms of anomaly detection and localization.

4.1 The power of Boltzmann Entropy as a feature

We evaluate the discriminant power of Boltzmann Entropy as a feature for anomaly detection on the UMN dataset. The dataset comprises 11 video sequences of 3 different indoor and outdoor scenes and each sequence begins with normal behavior, followed by a portion of abnormal panic such as running and dispersion. The duration of the videos is 4.17 min (7739 frames), and the resolution of each frame is 320×240 pixels. Given a video sequence of U clips, we denote the Boltzmann Entropy features of all the clips as $\{F_1, \cdots, F_U\}$. We define the outlier degree of F_t as

$$OD_t = \sum_{u=1}^{U} \| F_t - F_u \|_1 \tag{12}$$

which is the sum of the L1 norm distance between F_t and all the other features. Then, we sort the outlier degrees $\{OD_1, OD_2, \cdots, OD_U\}$ in descending order to seek anomalies. We perform the above computations for each of the 11 video sequences in the UMN dataset. As a result, 11 sorted lists are obtained in total. *TotalAnomaly* represents the total number of frames involving abnormal behaviors in the 11 video sequences of the UMN dataset. Suppose that in the top-N ranked OD values of the ith video sequence of the UMN dataset, there are $True_i(N)$ frames of true anomalies. Then, the precision and recall regarding the 11 video sequences can be computed as follows:

$$P(N) = \frac{\sum_{i=1}^{11} True_i(N)}{11 \times 10 \times N} \tag{13}$$

and

$$R(N) = \frac{\sum_{i=1}^{11} True_i(N)}{Total\ Anomaly} \tag{14}$$

Precision concerns how many frames in the detected list are true anomalies while recall figures out how many true anomalies have been identified from all the true anomalies. The precision and recall rates against N under different parameters are shown in Fig. 8. As can be seen in Fig. 8, the precision and recall rates vary little with different parameter setting of m_M and m_A. This shows that Boltzmann Entropy is not sensitive to partition of histogram bins. In this sense, Boltzmann Entropy is a robust feature. Figure 8 also shows that Boltzmann Entropy possesses the discrimination power to distinguish anomalies to some extent. Yet, this should be further examined since the time series of Boltzmann Entropy should provide more clues in the sequential arrangement of the values. As shown in Fig. 7, in normal cases, the Boltzmann Entropy will remain at low level continuously until an abnormal event takes place. Then, the Boltzmann Entropy will go up suddenly and keep
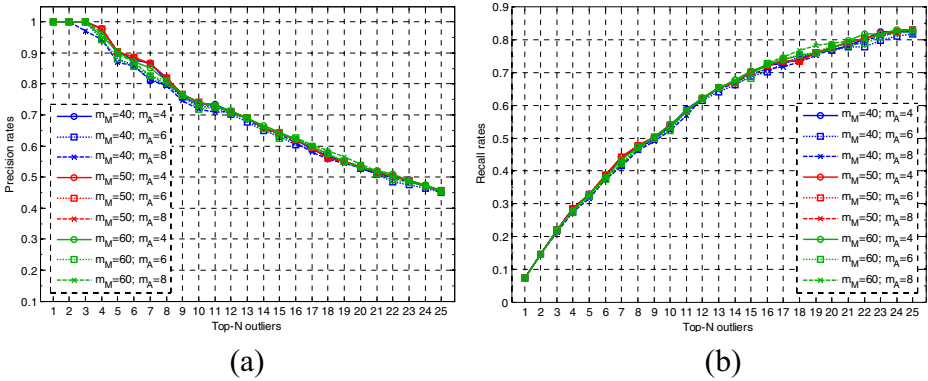
Fig. 8 The precision and recall rates under different parameters on the UMN dataset

on high values in a couple of consecutive clips. The consistency of the sequence lasting for a time and the turning point can provide extra information in judging anomalies. Accordingly, we can determine the starting time of an abnormal event via a threshold. The threshold can be applied to detect both suddenly appearing abnormal events and slowly appearing abnormal events.

As mentioned previously, the abnormal crowd movements are usually visible at the turning points of the time series of feature values. Therefore, we figure out the difference sequences of both Boltzmann Entropy and Shannon Entropy in Fig. 9(a) instead of the original time series to allow an intuitive insight into their pattern classification power. Moreover, in Fig. 9(b), (c), (d), and (e), we illustrate the histograms of the time sequences shown in Fig. 9(a) to reveal the distribution of the differential feature values belonging to normal and abnormal cases in terms of Boltzmann Entropy and Shannon Entropy, respectively. It is observed that the pattern classification power of Boltzmann Entropy is remarkably better than that of Shannon Entropy due to the much less overlap between the two classes in correspondence with normal and abnormal moments.

Furthermore, to examine quantitatively the discriminant power of the proposed feature in comparison with that of the Shannon Entropy feature, here, we employ the measure referred to as R-Metric [31] to evaluate the ability of either feature in distinguishing abnormal patterns from normal ones. Suppose that the two classes of the differential values of any entropy feature corresponding with normal and abnormal events are denoted as $\left\{d_1^x, d_2^x, \cdots, d_X^x\right\}$ and $\left\{d_1^y, d_2^y, \cdots, d_Y^y\right\}$, respectively, where X and Y represent the numbers of the samples belonging to the two classes, respectively. Prior to defining R-Metric, the relationship between d_i^x and d_j^y is quantifying through

$$H\left(d_i^x, d_j^y\right) = \begin{cases} 1 & \left|d_i^x\right| < \left|d_j^y\right| \\ 0 & \left|d_i^x\right| \geq \left|d_j^y\right| \end{cases} \tag{15}$$

where $i \in [1, X]$ and $j \in [1, Y]$, and $|\ |$ means absolute value. Then, R-Metric is defined by taking into account the relationships between all the samples in the two classes, that is,

$$RM = \max\left\{\sum_{i=1}^{X}\sum_{j=1}^{Y} H\left(d_i^x, d_j^y\right), \sum_{j=1}^{Y}\sum_{i=1}^{X} H\left(d_j^y, d_i^x\right)\right\} \tag{16}$$
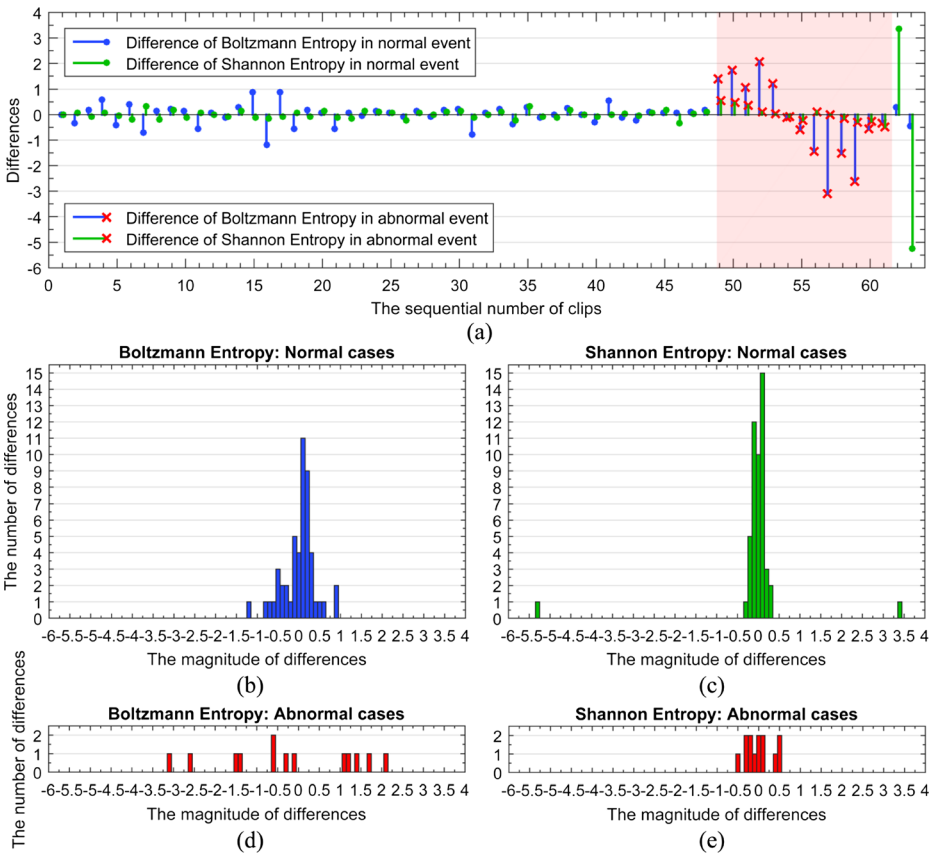
Fig. 9 Difference sequences of Boltzmann Entropy and Shannon Entropy and the corresponding histograms of a video sequence in the UMN dataset. **a** The differential sequences of 4 cases, where the shadow in *red color* corresponds with crowd commotion; **b, c, d, e** The histograms corresponding with the 4 cases, respectively

A bigger *RM* value corresponds with better separability between two classes since more samples belonging to the same class are consistently sorted into either of the two classes consisting of small values or big values and there is less overlap between the two classes. In Table 2, we list the comparison of the two features in terms of the *RM* value. It is notable that the *RM* values based on Boltzmann Entropy are obviously greater than those based on Shannon Entropy in almost all scenarios, except for the 11th video sequence. In an overall sense, Boltzmann Entropy is significantly superior to Shannon Entropy in discriminating anomalies.

Table 2 R-Metric of Boltzmann Entropy and Shannon Entropy features on each video sequence of the UMN dataset (SN: Serial number of video sequence; SE: Shannon Entropy)

SN	1	2	3	4	5	6	7	8	9	10	11	Total
Proposed	**588**	**826**	**848**	**677**	**890**	**610**	**945**	**1005**	**612**	**635**	418	**8054**
SE	442	792	730	518	799	497	711	830	483	508	**481**	6791

4.2 Evaluation on UMN dataset

Prior to anomaly detection, we establish a Gaussian model for each of the three scenes of the UMN dataset, where we select the first 30 clips from each video sequence to compute the difference sequence of Boltzmann Entropy as described previously. Then, we compute the mean and standard deviation of the difference values to render the corresponding Gaussian model. Once the Gaussian model corresponding with normal cases is established, then, we conduct the anomaly detection test as follows: If two consecutive clips appear with the probability lower than 0.1 under the corresponding Gaussian model, we label them as the start of an abnormal event. We choose the moment when Boltzmann Entropy reduces to 0.1 as the end of an abnormal event.

The detection results by using the proposed method on the 3 scenes composed of the 11 video sequences of the UMN dataset are shown in the 3 rows of Fig. 10 against the ground truth, where each red interval represents the lasting of an abnormal event in each video sequence while the green portions represent the course of normal situations. Note that there is no sole false alarm interval in the detection bars and every abnormal event in ground truth has been detected using the proposed method. In another words, there is no false alarm.

In Table 3, we compare the proposed method with optical flow [1], social force model [14], particle swarm optimized-social force model (PSO-SFM) [17], chaotic invariants [20], sparse reconstruction cost [32], Shannon Entropy [21], and non-extensive entropy [23] on the UMN

Fig. 10 Results of abnormal activity detection on the UMN dataset. Each row represents one scene. The *two bars* underlying each row are the detection result against the ground truth, where the *green color* represents the normal frames and the *red color* corresponds to the abnormal frames. Above the bar, the *left column* is an example of the normal frames and the *right column* is an example of the detected abnormal frames
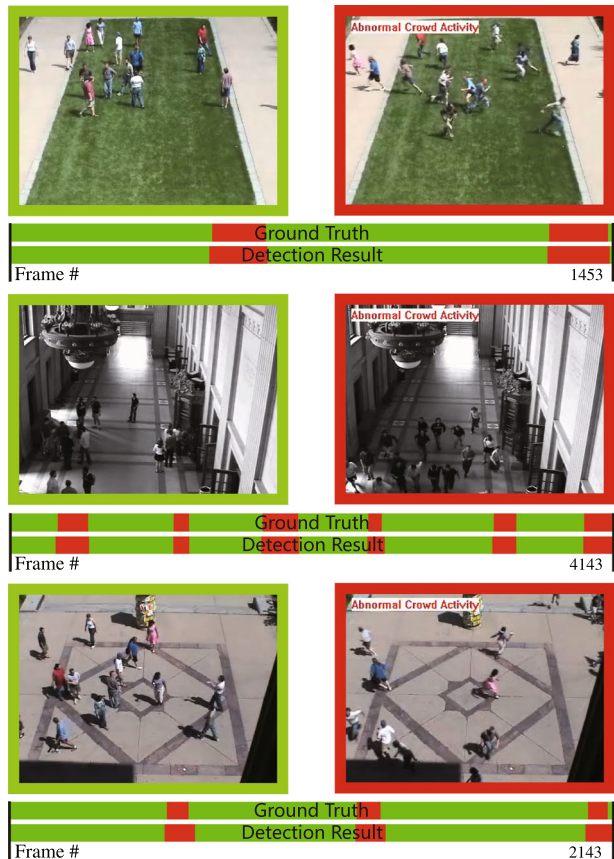
Table 3 Comparison of different methods for anomaly detection on the UMN dataset

Methods	AUC on UMN dataset	AUC on Scene 1/2/3	Average Speed (fps)
The Proposed	0.985	0.992/0.986/0.979	100
Shannon Entropy	0.943	0.987/0.913/0.976	103
Non-extensive Entropy	0.95	0.997/0.89/0.962	103
Sparse Reconstruction Cost	0.978	0.995/0.975/0.964	<20
SFM-PSO	0.996	0.996/0.993/0.999	<1
Chaotic Invariants	0.99	—	<5
Social Force Model	0.96	—	56
Optical Flow	0.84	—	88

dataset in terms of both accuracy and speed for abnormal crowd behavior detection. Here, we employ the area under ROC (receiver operating characteristic curve) curve (AUC) to measure the accuracy and the frame rate, say the number of frames processed per second (fps) to indicate the speed of each method in terms of computing the corresponding feature.

The baselines methods for comparison fall into 3 categories: (1) The optical flow [1] method detects anomalies from optical flows directly. Since no feature extraction is applied, it is obvious that the AUC of the optical flow method is much inferior to the performances of the other methods, which leads to the lowest AUC of 0.84 while those of the other ones are all above 0.94. The comparison shows that feature extraction is critical in differentiating abnormal and normal crowd motions. This shows why the effort to seek effective features is meaningful. The Shannon Entropy [21] and its variant, namely non-extensive entropy [23], are similar to the proposed method in that all the entropy-based features can be computed efficiently with the speed of roughly 100 fps. However, due to the different physical nature as demonstrated in Section 3.3, Shannon Entropy and non-extensive entropy lead to much poorer performance for scene 2, which are 0.913 and 0.89, respectively, in comparison with the 0.986 AUC of the proposed method. Note that scene 2 is an indoor scene with spatial constraint to force the evacuation behavior of people in panic homogenous, that is, the moving directions and speeds of particles are mostly the same. This is a difficult task for Shannon Entropy, which accounts for why the overall performances of Shannon Entropy and non-extensive entropy are inferior to that of the proposed method, that is, 0.943 and 0.95 against 0.985. As for sparse reconstruction cost [32], although the accuracy is 0.978 AUC, which is a bit lower than the 0.985 AUC of the proposed method, its speed is lower than 20 fps due to the time-consuming complex modeling.

(2) The performance of the social force model is 0.96, which is lower than the 0.985 AUC of the proposed model. Note that the original social force model proposed by Dirk Helbing is focused on modeling the interactions among individual persons [9]. In the computer vision context [14], however, due to the difficulty in segmenting every individual person in crowded scenes, the definition of particles is relaxed to be collections of optical flows instead of persons, which deviates from the original definition due to splitting one person into a couple of particles. This should be the major reason to degrade the performance of the social force model. To solve this problem, particle swarm optimization-based social force model is proposed to reorganize the computation with selective particles. It leads to the highest AUC of 0.996 among all the competitive methods. However, the performance improvement is obtained at the cost of decreasing the speed of the original social force model from 56 fps to the less than 1 fps, which is remarkably out of the scope for real-time video surveillance.

(3) As for the chaotic invariants, 0.99 AUC is achieved in [20] but the authors report only the performance on the 6 selective video sequences, not all the 11 video sequences of the UMN dataset. The chaotic invariants need to solve the evolution equations of the particle positions to obtain particle trajectories. Furthermore, it iteratively clusters the particle trajectories to extract representative trajectories for computing the largest Lyapunov exponent and the correlation dimension, which measure how much neighboring trajectories deviate from their original closeness to each other after a certain steps of evolution. Here, a huge number of particles are involved into the evolution and iteration processes, which makes the computational cost of the chaotic invariants high. So it an impractical solution on account of the heavy computation caused less than 5 fps low speed, which is far from real-time realization. Besides, for the scenes that the people' movements are constrained in a narrow space such as in corridors and underpasses, the evolution of the trajectories might not follow the precondition of chaos, which assumes that neighboring trajectories will increasingly fall apart accompanying elapse of time. In such a case, Lyapunov exponent may not reflect exactly the chaotic degree of the collective human mobility corresponding with anomaly due to the constrained evolution of the trajectories.

On account of the above comparison, we see that the proposed method promises the highest accuracy among all the real-time workable methods for abnormal crowd behavior detection.

4.3 Anomaly detection and localization on web data

There are many real-world crowd motion videos. We collect 32 challenging video sequences of different scenarios from Internet as additional benchmark to further evaluate the various methods that promise real-time surveillance. Figure 11 shows the snapshots of 6 of the 32 video sequences. The first video provides a top view of the rapid evacuation of the crowd shocked by high-altitude falling. The second one records the terroristic attack in Kunming train station in March 2014, where people are running away during the suddenly happened terroristic attack. The third video depicts the anti-government activities in Bangkok 2014 during the bomb attack. The fourth video records how policemen disperse the crowd using water gun. The fifth one provides a high altitude view of the scenario during the gunfight between arrest resistant and police. The sixth one is a surveillance video showing the fighting in a supermarket. The other videos of the web dataset record more crowd events, which are available online.[1]

We locate the positions corresponding with abnormal behaviors in crowded scenes through Boltzmann Entropy and Shannon Entropy for the image patches of uniform size. Some results for the 6 scenes from the web dataset are shown in Fig. 11, where the regions of anomalies that possess higher entropy values in contrast to the surroundings are highlighted with red squares in the frames. It can be observed that the precision to localize anomaly regions based on Boltzmann Entropy is much higher than that based on Shannon Entropy as Shannon Entropy-based localization is sometimes over-segmented. The results demonstrate Boltzmann Entropy is a general model that can be used to detect global abnormal events as well as localize abnormal activities.

In Fig. 12, we show how Boltzmann Entropy changes with time. In the 3 cases shown in Fig. 12, originally, the Boltzmann Entropy remains at a relatively low level but when an abnormal event happens to invoke people's chaotic movements for escaping, the Boltzmann Entropy goes up suddenly and violently. Such examples show that the proposed feature is effective in dealing with real-world data.

---

[1] http://pan.baidu.com/s/1qWwTY9m

Fig. 11 Localization of abnormal behaviors. The *left column* is the original frames. The *middle column* is the localized abnormal behaviors using Boltzmann Entropy. The *right column* shows those obtained using Shannon Entropy. Highlighted *squares* correspond to the highly possible abnormal regions

Moreover, we compare the proposed method with the optical flow method [1], social force model [14], Shannon Entropy [21], and non-extensive entropy [23] on the total 32 web videos since these methods work in real time on the UMN dataset. In the experimental setting, the same number of normal video frames are used to calculate the features for each method so as to establish the Gaussian model corresponding with the normal cases, where a multivariate Gaussian model is used for the optical flow histogram [1]. When 3 consecutive clips corresponding with a probability lower than 0.1 appear using such Gaussian model, an abnormal event is alarmed. The results are illustrated in Fig. 13, where we see that the proposed feature

(a)

(b)

(c)
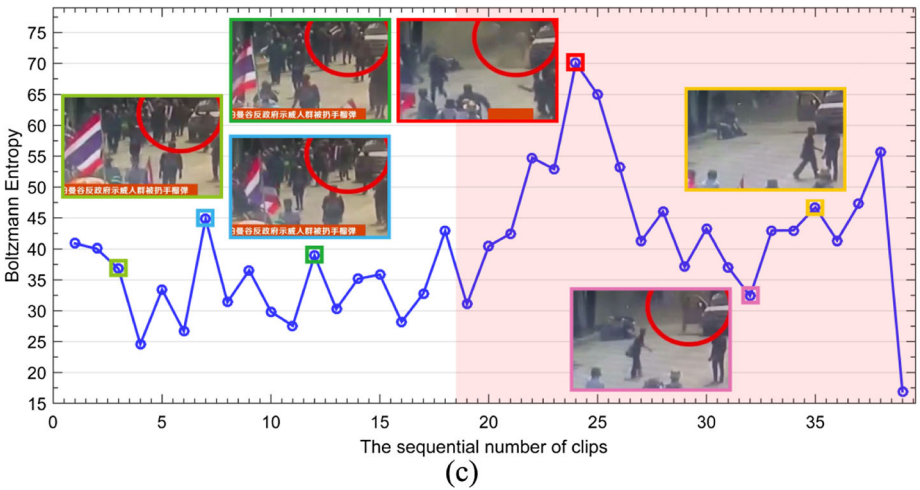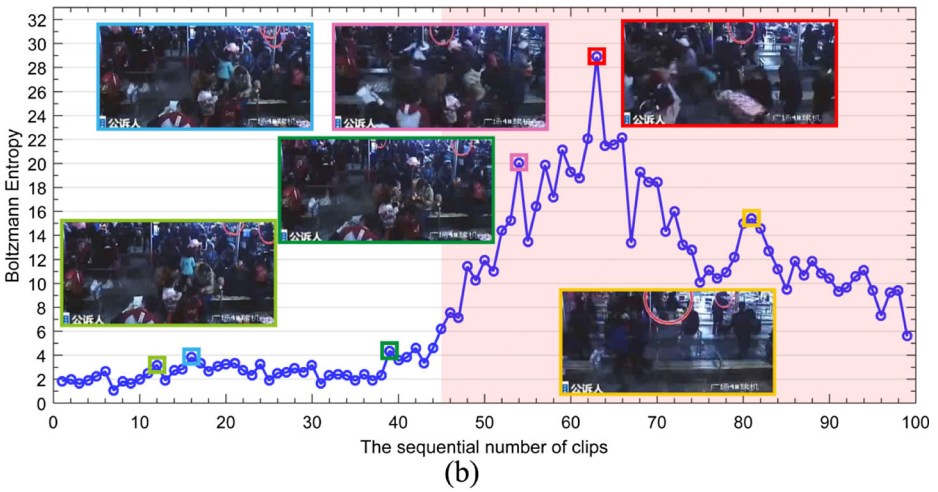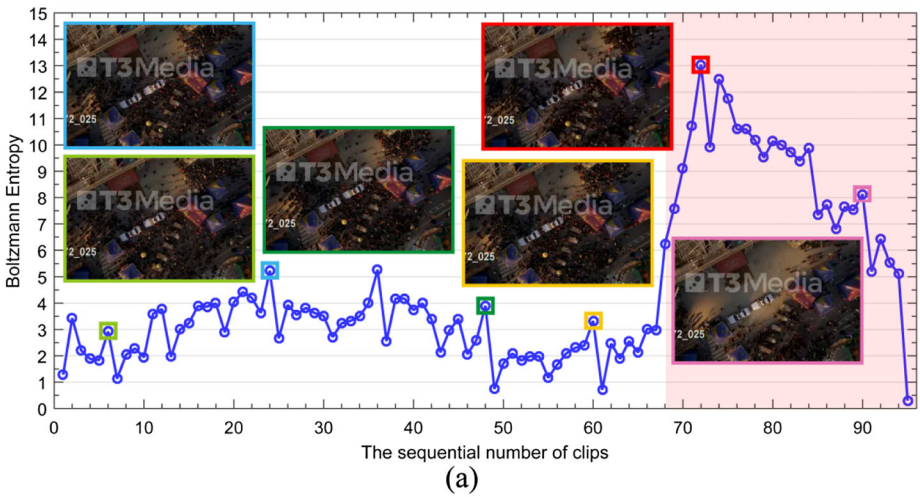
◀ **Fig. 12** Curves of Boltzmann Entropy values against clips for 3 scenes of the web dataset. The shadow in *red color* corresponds with the lasting time of crowd commotion. The online monitoring images are also attached to the corresponding clips

consistently leads to relatively accurate anomaly detection close to the ground truth in almost all such 32 scenarios while the other methods do not work satisfactorily for some of the 32 cases. The precise performance data are reported in Table 4, where we see the same results as shown in Fig. 13. The AUCs (area under ROC curve) with regard to the 32 video sequences listed in Table 4 show that the proposed method yields the best performance in 14 cases while in the other cases, its performance is very close to the best one in comparison with the optical flow [1], social force model [14], Shannon Entropy [21], and non-extensive entropy [23].

The performance curves of those methods are compared in Fig. 14. The AUC of the proposed method is the highest one and the only one over 0.9. According to Table 3, Table 4, and Fig. 14, the performance of social force model degrades dramatically on the real scenarios since it fails to detect 11 video sequences of the web dataset in Fig. 13 and only 0.736 AUC is achieved in contrast to the 0.96 AUC on the UMN benchmark. Its performance is only better than using optical flow feature directly as the AUC of the optical flow method drops from 0.84 on the UMN benchmark to 0.689 on the real-world data. This manifests that the social force model in the sense of computer vision deviates far from the nature of human dynamics in complex crowded scenarios. As shown in Table 3 and Fig. 14, the AUCs of the Shannon
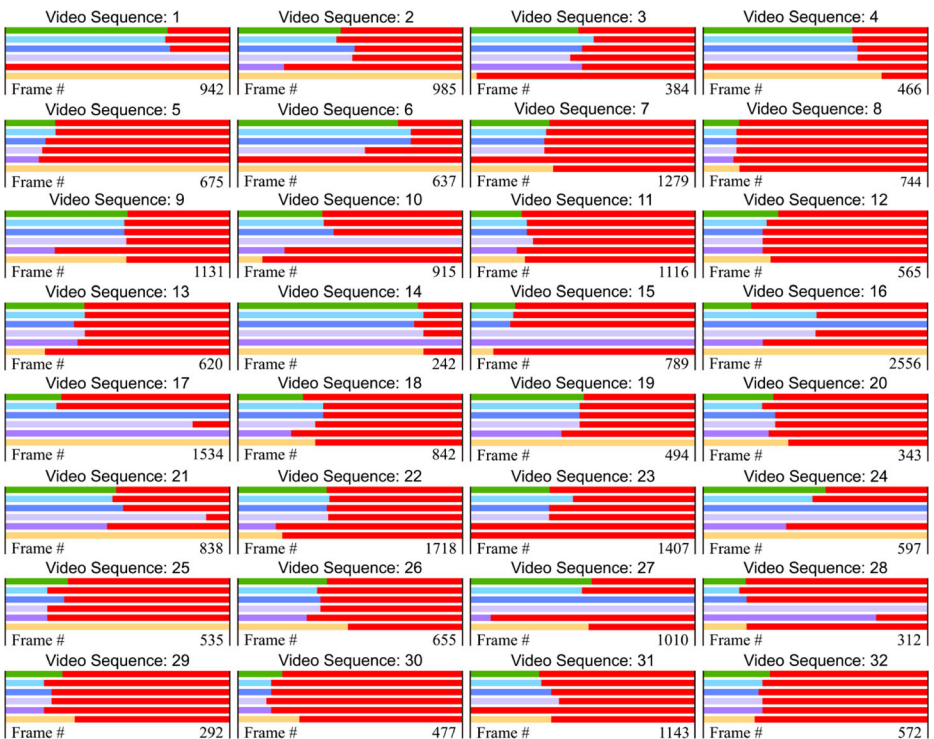

Fig. 13 Performance comparison regarding anomaly detection on the web dataset. For each video sequence, the *color bars* from top to bottom represent the ground truth, the results of the proposed method, Shannon Entropy, non-extensive entropy, optical flow, and social force model, respectively, where the *red color* corresponds with abnormal cases and the other colors the normal ones

Table 4 AUC by using different methods on the web dataset. SN: Serial number of each video sequence; BE: The proposed method; OF: Optical flow method; SE: Shannon Entropy; NE: Non-extensive entropy; SFM: Social force model

SN	BE	SE	NE	OF	SFM	SN	BE	SE	NE	OF	SFM
1	**0.993**	0.981	0.500	0.500	0.500	17	**0.956**	0.500	0.610	0.500	0.500
2	**0.979**	0.943	0.952	0.723	0.500	18	0.937	0.937	**0.962**	0.910	**0.962**
3	0.935	**0.985**	0.962	**0.985**	0.527	19	**0.982**	**0.982**	**0.982**	0.902	0.500
4	**0.997**	0.965	0.965	0.500	0.806	20	0.921	**0.994**	**0.994**	0.967	0.951
5	**0.999**	0.903	0.869	0.836	0.500	21	**0.983**	0.969	0.604	0.959	0.500
6	**0.902**	**0.902**	0.897	0.500	0.500	22	0.990	**0.999**	0.995	0.714	0.750
7	0.979	0.968	0.968	0.500	**0.987**	23	0.920	**0.996**	**0.996**	0.500	0.500
8	0.962	0.962	0.962	0.920	**0.999**	24	**0.946**	0.500	0.500	0.839	0.500
9	0.986	0.986	**0.994**	0.703	**0.994**	25	0.836	**0.970**	0.836	0.836	0.500
10	**0.995**	0.960	0.500	0.776	0.645	26	0.944	**0.963**	**0.963**	0.886	0.923
11	0.985	0.985	0.968	0.953	**0.991**	27	0.960	0.500	0.500	0.583	**0.987**
12	0.923	0.897	0.897	0.897	**0.950**	28	0.924	**0.998**	0.500	0.642	**0.998**
13	**0.999**	0.934	**0.999**	0.957	0.751	29	0.838	0.905	0.905	0.838	**0.963**
14	0.938	**0.990**	0.938	0.500	0.938	30	0.872	0.872	0.819	0.872	**0.953**
15	**0.978**	0.946	0.500	0.500	0.755	31	**0.993**	0.962	0.936	0.500	0.962
16	0.815	0.500	0.818	**0.967**	0.500	32	**0.944**	0.914	**0.944**	**0.944**	0.885

Entropy and its variant, non-extensive entropy decrease from 0.943 and 0.95 to 0.860 and 0.842, respectively, when the benchmark changes from the UMN dataset to the web dataset. Note that in Fig. 13, Shannon Entropy and non-extensive entropy do not work on 4 and 6 video sequences, respectively. Shannon Entropy and non-extensive entropy, which is a variant of it, can only reflect the probability of the distribution of particles but miss the total amount of the particles involved in crowd motion but the chaos degree of crowd motion is in fact subject to the number of active particles. This explains why the performances of the two entropy-based methods are similar and worse than that of the proposed method. The best and stable detection performance on the

Fig. 14 Comparison of the detection results using different methods on the web dataset. OF: Optical flow; SE: Shannon Entropy; NE: Non-extensive entropy; SFM: Social force model
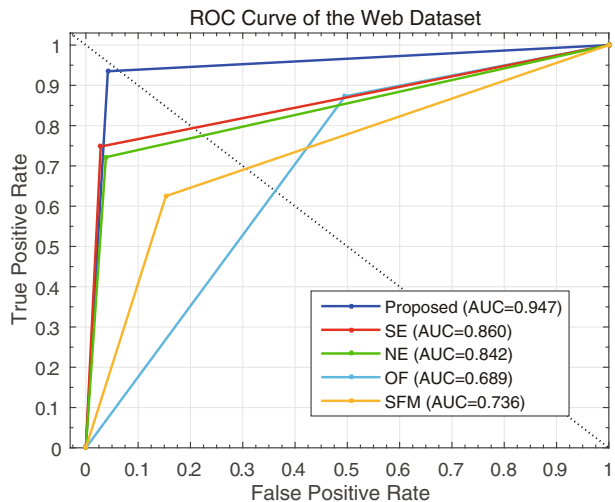
Table 5 Conclusive comparison

Method	Speed	Learning	Dimensionality	Performance stability (UMN➔web dataset)	Effectiveness (#Workable/#Total)
Proposed	Real Time	Simple	Low	Strong (0.985➔0.947)	Strong (32/32)
Shannon Entropy	Real Time	Simple	Low	Middle (0.943➔0.86)	Middle (28/32)
Non-extensive entropy	Real Time	Simple	Low	Middle (0.95➔0.842)	Middle (26/32)
Optical flow	Real Time	Simple	High	Weak (0.84➔0.689)	Weak (23/32)
Social force model	Real Time	Complex	High	Weak (0.96➔0.736)	Weak (21/32)

complex scenarios is obtained by using the proposed feature because Boltzmann Entropy directly characterizes the macroscopic chaos degree of crowd motion through microscopic particle movements in term of thermodynamics, where the AUC remains stable at 0.985 and 0.947 on both benchmarks. In accordance with Fig. 13, it is the only method that works for all the 32 challenging video sequences. It obtains the best performance on nearly half of the 32 video sequences according to Table 4 and the best overall performance in view of Fig. 14. These show that the proposed method is a universally promising feature.

5 Conclusion

Rooting in microscopic statistical theory, we introduce Boltzmann Entropy to capture the chaos degree of crowd behavior for video surveillance. The advantages lie in the following three aspects: (1) The Boltzmann Entropy characterizes the chaos heat motion of molecules, which is similar to the physical nature of crowd motion in surveillance videos. This accounts for why it performs with satisfactory accuracy on anomaly detection. (2) The Boltzmann Entropy feature is extracted from optical flow directly without the need to detect or track objects individually. (3) As the crowd motion is represented in a very simply and efficient way as a single time series, only a threshold is needed for decision making. This avoids complex and time-consuming machine learning on large volume training data and makes online monitoring a practical task. The proposed method promises competitive performance in detecting and localizing abnormal behaviors in crowded scenes while the computation is much more simple and efficient in comparison with the existing methods. A detailed comparison is summarized in Table 5, where we see that the proposed method is the unique one that promises high precision on different scenes and datasets, and can work on all the 32 challenging video sequences in the sense of anomaly detection.

References

1. Andrade EL, Blunsden S, Fisher RB Modelling Crowd Scenes for Event Detection. In: Pattern Recognition, 2006. ICPR 2006. 18th International Conference on, 0–0 0 2006. pp 175–178. doi:10.1109/ICPR.2006.806
2. Basharat A, Gritai A, Shah M Learning object motion patterns for anomaly detection and improved object detection. In: Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, 23–28 June 2008 2008. pp 1–8. doi:10.1109/CVPR.2008.4587510
3. Brox T, Bruhn A, Papenberg N, Weickert J (2004) High accuracy optical flow estimation based on a theory for warping. In: Pajdla T, Matas J (eds) Computer vision - ECCV 2004, Lecture

Notes in Computer Science, vol 3024. Springer, Berlin Heidelberg, pp. 25–36. doi:10.1007/978-3-540-24673-2_3

4. Cong Y, Yuan J, Liu J (2013) Abnormal event detection in crowded scenes using sparse representation. Pattern Recogn 46(7):1851–1864. doi:10.1016/j.patcog.2012.11.021

5. Dalal N, Triggs B Histograms of oriented gradients for human detection. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, 25–25 June 2005 2005. pp 886–893 vol. 881. doi:10.1109/CVPR.2005.177

6. Gu X, Cui J, Zhu Q (2014) Abnormal crowd behavior detection by using the particle entropy. Optik – International Journal for Light and Electron Optics 125(14):3428–3433. doi:10.1016/j.ijleo.2014.01.041

7. Haering N, Venetianer P, Lipton A (2008) The evolution of video surveillance: an overview. Mach Vis Appl 19(5–6):279–290. doi:10.1007/s00138-008-0152-0

8. Halliday D, Resnick R, Walker J (2010) Fundamentals of physics extended 9 edition. Wiley:550-560

9. Helbing D, Molnár P (1995) Social force model for pedestrian dynamics. Phys Rev E 51(5):4282–4286

10. Helbing D, Johansson A, Al-Abideen HZ (2007) Dynamics of crowd disasters: an empirical study. Phys Rev E 75(4). doi:10.1103/PhysRevE.75.046109

11. Jie F, Chao Z, Pengwei H Online anomaly detection in videos by clustering dynamic exemplars. In: Image Processing (ICIP), 2012 19th IEEE International Conference on, Sept. 30 2012-Oct. 3 2012 2012. pp 3097–3100. doi:10.1109/ICIP.2012.6467555

12. Kaltsa V, Briassouli A, Kompatsiaris I, Hadjileontiadis LJ, Strintzis MG (2015) Swarm intelligence for detecting interesting events in crowded environments. Image Processing, IEEE Transactions on 24(7):2153–2166. doi:10.1109/TIP.2015.2409559

13. Mahadevan V, Weixin L, Bhalodia V, Vasconcelos N Anomaly detection in crowded scenes. In: Computer Vision and Pattern Recognition (CVPR), 2010 I.E. Conference on, 13–18 June 2010 2010. pp 1975–1981. doi:10.1109/CVPR.2010.5539872

14. Mehran R, Oyama A, Shah M Abnormal crowd behavior detection using social force model. In: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, 20–25 June 2009 2009. pp 935–942. doi:10.1109/CVPR.2009.5206641

15. Nam Y (2014) Crowd flux analysis and abnormal event detection in unstructured and structured scenes. Multimed Tools Appl 72(3):3001–3029. doi:10.1007/s11042-013-1593-7

16. Pathan S, Al-Hamadi A, Michaelis B (2011) Using conditional random field for crowd behavior analysis. In: Koch R, Huang F (eds) Computer vision – ACCV 2010 workshops, Lecture Notes in Computer Science, vol 6468. Springer, Berlin Heidelberg, pp. 370–379. doi:10.1007/978-3-642-22822-3_37

17. Raghavendra R, Del Bue A, Cristani M, Murino V Optimizing interaction force for global anomaly detection in crowded scenes. In: Computer Vision Workshops (ICCV Workshops), 2011 I.E. International Conference on, 6–13 Nov. 2011 2011. pp 136–143. doi:10.1109/ICCVW.2011.6130235

18. Reddy V, Sanderson C, Lovell BC Improved anomaly detection in crowded scenes via cell-based analysis of foreground speed, size and texture. In: Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 I.E. Computer Society Conference on, 20–25 June 2011 2011. pp 55–61. doi:10.1109/CVPRW.2011.5981799

19. Sethi RJ, Roy-Chowdhury AK (2010) Modeling and recognition of complex multi-person interactions in video. Paper presented at the Proceedings of the 1st ACM international workshop on Multimodal pervasive video analysis, Firenze, Italy,

20. Shandong W, Moore BE, Shah M Chaotic invariants of Lagrangian particle trajectories for anomaly detection in crowded scenes. In: Computer Vision and Pattern Recognition (CVPR), 2010 I.E. Conference on, 13–18 June 2010 2010. pp 2054–2060. doi:10.1109/CVPR.2010.5539882

21. Shannon CE (1948) A mathematical theory of communication. Bell System Technical Journal 27:379–423

22. Si W, Hau-San W, Zhiwen Y (2014) A Bayesian model for crowd escape behavior detection. Circuits and Systems for Video Technology, IEEE Transactions on 24(1):85–98. doi:10.1109/TCSVT.2013.2276151

23. Susan S, Hanmandlu M (2013) Unsupervised detection of nonlinearity in motion using weighted average of non-extensive entropies. SIViP:1–15. doi:10.1007/s11760-013-0464-z

24. Tian C, Xinyu W, Jinnian G, Shiqi Y, Yangsheng X Abnormal crowd motion analysis. In: Robotics and Biomimetics (ROBIO), 2009 I.E. International Conference on, 19–23 Dec. 2009 2009. pp 1709–1714. doi:10.1109/ROBIO.2009.5420408

25. Tu P, Sebastian T, Doretto G, Krahnstoever N, Rittscher J, Yu T (2008) Unified crowd segmentation. In: Forsyth D, Torr P, Zisserman A (eds) Computer vision – ECCV 2008, Lecture Notes in Computer Science, vol 5305. Springer, Berlin Heidelberg, pp. 691–704. doi:10.1007/978-3-540-88693-8_51

26. UMN Unusual crowd activity dataset of University of Minnesota. available from http://mhacsumnedu/movies/crowd-activity-allavi

27. Wang X, Tieu K, Grimson E (2006) Learning semantic scene models by trajectory analysis. In: Leonardis A, Bischof H, Pinz A (eds) Computer vision – ECCV 2006, Lecture Notes in Computer Science, vol 3953. Springer, Berlin Heidelberg, pp. 110–123. doi:10.1007/11744078_9

28. Weixin L, Mahadevan V, Vasconcelos N (2014) Anomaly detection and localization in crowded scenes. Pattern Analysis and Machine Intelligence, IEEE Transactions on 36(1):18–32. doi:10.1109/TPAMI.2013.111
29. Xinyi C, Qingshan L, Mingchen G, Metaxas DN Abnormal detection using interaction energy potentials. In: Computer Vision and Pattern Recognition (CVPR), 2011 I.E. Conference on, 20–25 June 2011 2011. pp 3161–3167. doi:10.1109/CVPR.2011.5995558
30. Xiong G, Cheng J, Wu X, Chen Y-L, Ou Y, Xu Y (2012) An energy model approach to people counting for abnormal crowd behavior detection. Neurocomputing 83(0):121–135. doi:10.1016/j.neucom.2011.12.007
31. Yang S Gene Ranking Using R-Metric. In: Computational Intelligence and Security, 2006 International Conference on, Nov. 2006 2006. pp 465–469. doi:10.1109/ICCIAS.2006.294178
32. Yang C, Junsong Y, Ji L Sparse reconstruction cost for abnormal event detection. In: Computer Vision and Pattern Recognition (CVPR), 2011 I.E. Conference on, 20–25 June 2011 2011. pp 3449–3456. doi:10.1109/CVPR.2011.5995434
33. Zhang X, Yang S, Tang YY, Zhang W (2015) Crowd Motion Monitoring with Thermodynamics-Inspired Feature. North America, mar. 2015., AAAI Conference on Artificial Intelligence

Xinfeng Zhang received a BE degree in electronic and information engineering from Hebei University in 2006 and an ME degree in information and communication engineering from Shantou University in 2010. He is pursuing his PhD degree in computer science at Fudan University. His current research interest is in the area of computer vision, including image registration and video understanding.

Su Yang is a full professor in College of Computer Science at Fudan University. His main research interest is pattern recognition and its applications in media processing and smart cities. His work in symbol recognition was widely cited in the context of Graphics Recognition. He received the best paper award from CPSCom 2010 and chaired the 7th SocialCom in Beijing, 2014. His recent effort is focused on symbol recognition in natural scenes, video surveillance for anomaly detection on crowd behaviors, and traffic flow prediction from big data point of view.

Yuan Yan Tang is a Chair Professor in Faculty of Science and Technology at University of Macau and Professor/Adjunct Professor/Honorary Professor at several institutes in China, USA, Canada, France, and Hong Kong. His current interests include wavelets, pattern recognition, image processing, and artificial intelligence. He has published more than 400 academic papers and is the author/coauthor of over 25 monographs/books/bookchapters. He is the Founder and Editor-in-Chief of International Journal on Wavelets, Multiresolution, and Information Processing (IJWMIP), and Associate Editors of several international journals. He is the Founder and Chair of pattern recognition committee in IEEE SMC. He has serviced as general chair, program chair, and committee member for many international conferences. Dr. Tang is the Founder and General Chair of the series International Conferences on Wavelets Analysis and Pattern Recognition (ICWAPRs). He is the Founder and Chair of the Macau Branch of International Associate of Pattern Recognition (IAPR). Dr. Y. Y. Tang is a Fellow of IEEE, and Fellow of IAPR.

Weishan Zhang is a full professor, deputy head for research of Department of Software Engineering, China University of Petroleum. His current research interests are big data processing, pervasive and service oriented computing. Weishan has published over 60 papers. According to google scholar, his current total citations is over 510, H-index is 13, and I10-index is 18.