



An entropy approach for abnormal activities detection in video streams

Md. Haidar Sharif^{a,*}, Chabane Djeraba^b

^a Gediz University, Izmir, Turkey

^b University of Lille 1 (USTL), France

ARTICLE INFO

Article history:

Received 29 June 2010

Received in revised form

6 August 2011

Accepted 26 November 2011

Available online 17 January 2012

Keywords:

Abnormality

Circular variance

Degree of randomness

Entropy

ABSTRACT

Detection of aberration in video surveillance is an important task for public safety. This paper puts forward a simple but effective framework to detect aberrations in video streams using *Entropy*, which is estimated on the statistical treatments of the spatiotemporal information of a set of interest points within a region of interest by measuring their *degree of randomness* of both *directions* and *displacements*. Entropy is a measure of the disorder/randomness in video frame. It has been showed that degree of randomness of the directions (*circular variance*) changes markedly in abnormal state of affairs and does change only direction variation but does not change with displacement variation of the interest point. Degree of randomness of the displacements has been put in for to counterbalance this deficiency. Simple simulations have been exercised to see the characteristics of these crude elements of entropy. Normalized entropy measure provides the knowledge of the state of anomalousness. Experiments have been conducted on various real world video datasets. Both simulation and experimental results report that entropy measures of the frames over time is an outstanding way to characterize anomalies in videos.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Abnormal activities detection in video surveillance is a necessary undertaking for public security and safety. As huge amount of video data makes it an exhausting work for people to monitor and find atypical events, an automatic system is badly needed for detecting suspicious events which would exist a potential threat. An abnormal video event, may vary greatly in duration, can be defined to be an observable action or change of state in a video stream that would be very important for security management. There are some works [11–13,25,18,15,27,28,1–3,5,19] which detect abnormalities in crowd flows. A system for automatically learning motion patterns for anomaly detection and behavior prediction based on a proposed algorithm for robustly tracking multiple objects can be seen in [11]. Authors in [25] also detected unusual events which have never occurred or occur so rarely that they are not represented in the clustered activities. The method includes robust tracking, based on probabilistic method for background subtraction. But the robust tracking method is not adapted to crowd scene, in which it is too complex to track objects. A spatial model to represent the routes in an image has been developed in [18]. But the system cannot differentiate between a person walking and a person lingering around, or

between a running and a walking person. A method for detecting nonconforming trajectories of objects has been proposed in [15]. A framework for automatic behavior profiling and abnormality sampling/detection without any manual labeling of the training dataset can be found in [27,28].

Authors in [2,3] combined HMM, spectral clustering and principal component for detecting crowd emergency scenarios. But the methods were experimented in simulated data. Using a supervised SVM method, authors in [13] proposed an approach which makes a step toward generic and automatic detection of unusual events in terms of velocity and acceleration. The problem of detecting irregularities in visual data has been addressed by [5]. The method would bring about attention, yet it needs some sort of learning process and/or training data. Authors in [1] proposed a holistic method for segmentation of high density crowds by introducing a method based on Coherent Structures from fluid dynamics and particle advection. Their framework is suitable to detect flow instabilities from the events, e.g., marathon, religious festival, etc. by identifying changes in the segmentation. For crowd segmentation, they have been taken into account the goal-directed nature of human crowds, where the members of the crowds have clear knowledge of what and where their goals rest, e.g., extremely large number of people at sporting events, religious festivals, etc. This goal-directed nature has been implemented on to the crowd segmentation framework, where segments are distinguished from each other on the basis of the fate of the particles belong to that segment. The particles with similar fate have similar goals, and, thus, characterize a distinct group of the

* Corresponding author. Tel.: +90 2323550000x2312; fax: +90 2324256813.
E-mail address: md-haidar.sharif@gediz.edu.tr (Md. Haidar Sharif).

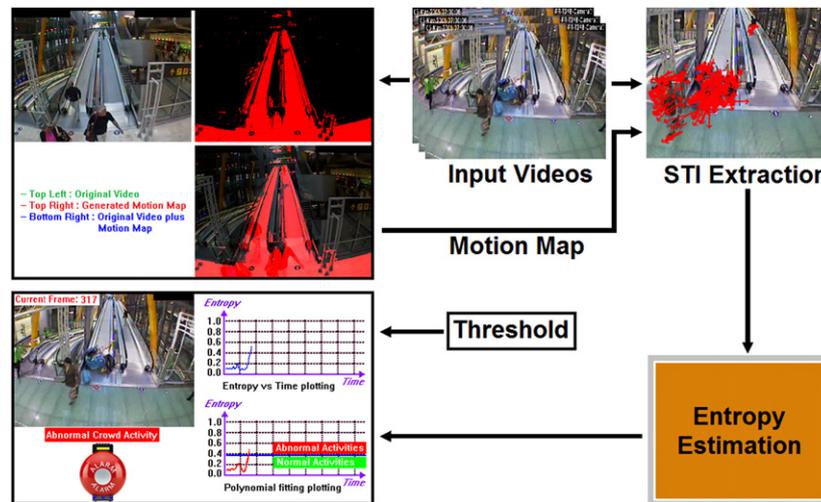


Fig. 1. The summary of the proposed framework.

crowd in a given scene. Results showed some satisfactory results for extremely high crowded scenes, yet medium or low crowded scenes such segmentation would be gone in vain. An approach to detect abnormal situations in crowded scenes by analyzing the motion aspect without tracking individual subjects can be viewed as [12]. The authors in [19] introduced a method by capturing the dynamics of the crowd behavior to detect and localize abnormal behaviors in crowd videos using Social Force model. A grid of particles was placed over the image and it was advected with the space–time average of optical flow. The experiments were conducted on different datasets and the algorithm captured the dynamics of the crowd behavior successfully.

However, most of the aforementioned methods require a learning period to estimate various parameters of the system, and hence reliable learning of unknown parameters is not always accurately possible which could potentially increase the rate of false alarms. For instance, the method of [19] detects and localizes abnormal behaviors in crowd videos using social force model. Their estimated social force model is capable of detecting the governing dynamics of the abnormal behavior, even in the scenes that it is not trained. But significant number of false positive detections in their model are the results of incorrect estimation of social forces. This is an unnecessarily extreme circumstances of their approach.

In this paper, we have introduced an approach which has no explicit learning period but expects a threshold, consequently, the false alarm rates are significantly low. The approach detects abnormalities in videos based on the statistical treatments of spatiotemporal information of a set of interest points within a region of interest on estimating their degrees of disorder/chaos over time entitled *Entropy*. Our primary goal is to introduce a holistic method, which is free from segmentation or individual subject tracking, to detect anomalies in videos by measuring the degrees of disorder/chaos present in them. Fig. 1 outlines the framework. A common aspect of our work and the works of [2,3,12] is that there is enough perturbation in the optical flow pattern in case of emergency. In the same vein, in the work of [13] authors used few nearby terms of us e.g., velocity and acceleration. Our approach would be deemed as a further enhancement of those works in some senses. Notwithstanding, we profited from a different course by using two statistical measures namely degree of randomness of the directions (circular variance) and degree of randomness of the displacements (coefficient of displacement variation) of interest points, which are the essential and sufficient crude elements of our defined entropy measure, which is a measure of the disorder or randomness in video frame. The more is the disorder/chaos presents in the video frame, the more is the

entropy, i.e., abnormal video frame has higher entropy than that of normal. We showed that circular variance, one of the two crude elements of the entropy, changes significantly in abnormal circumstances and it does not change with vector length variation of interest point but does change only direction variation of interest point. To compensate this shortcoming of circular variance we used a normalized and dimensionless quantity namely the degree of randomness of the displacements, another crude element of the entropy, which is a statistical measure of ratio between standard deviation and mean. Furthermore, we clarified that abnormality concerns displacements as well. Hence the degree of randomness of the displacements, the sufficient factor of the defined entropy measure, plays an important role to detect some kind of aberrations from videos. On estimating the entropy, we can detect anomalies directly without segmentation or tracking subject singly. In addition, the framework is effective for the high density mover scenes as well as low density scenes. It has other boons: (i) it detects all events in videos where entropy variations are important as compared to previous events; (ii) it works all directional flow of movers without imposing a restriction of their numbers in the videos; (iii) it does not expect efficient learning process and training data but would look for a prior cutoff.

In a nut shell, the strongest points of this paper are: (i) entropy based measurement for determining abnormal event is *novel*; (ii) proposed method works directly with the optical flows and therefore overcomes problems with detection and tracking of individual objects; (iii) extensive experiments have been performed on real world datasets as well as some simple simulated data.

The rest of the paper is organized as follows: Section 2 outlines the processing steps of the proposed framework; Section 3 reports a detailed detection abilities of the approach followed by few clues for further investigation; and Section 4 makes conclusion.

2. Processing steps

2.1. Region of interest (RoI) estimation

Irrespective of indoor and outdoor video surveillance, RoI makes the video processing faster. Based on applications and type of videos, RoI would extend from few parts of a video frame to the whole frame. In case of applications, e.g., to monitor escalators, linear passages, high-way, etc., video processing region can be fixed by using a mask instead of analyzing the whole video frame. We build a *motion map* (MM) for such applications. The MM,

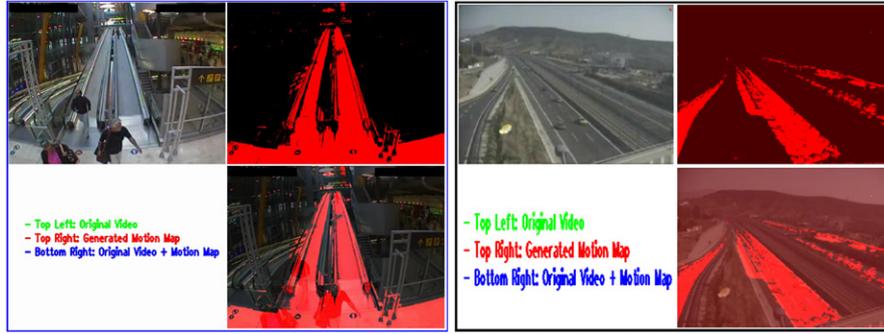


Fig. 2. Motion map (MM) generator's snapshot: escalator (left) and high-way (right) cases. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

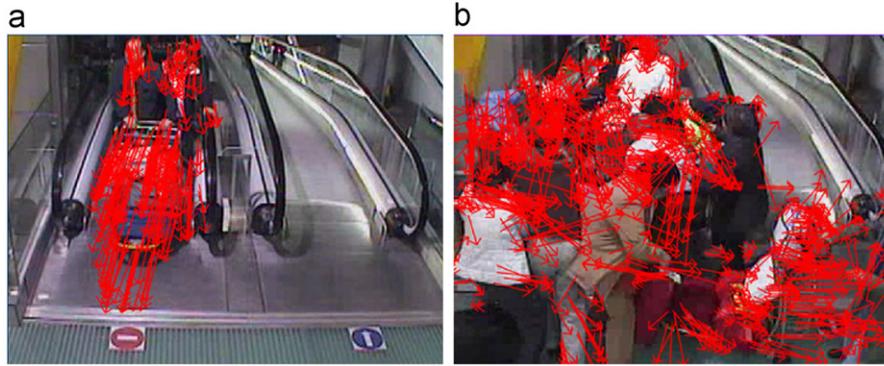


Fig. 3. Optical flow: (a) monomorphically directed vector flows *normal case*, (b) haphazardly directed vector flows *abnormal case*. The more is the disorder/chaos presents in the video frame, the more is the *Entropy*; e.g., entropy of (b) is greater than that of (a).

represented as color, is a 2D histogram expressing briefly the important (or all) regions of motion activity in videos. This histogram is built from the accumulation of real-time representation of object movements, so-called *motion history image* (MHI) as proposed by [4]. In an MHI pixel intensity is a function of the temporal history of position or motion at that point. The result of the function is a scalar-valued image where more recently moving pixels are brighter. If we wish to build a RoI, which would be either on-line or off-line, we need to store the information of pixels where motion happened far ago so that the accumulation of object silhouettes in the motion template can yield useful motion information along the contours of the silhouette. Fig. 2 depicts the occurrences of the obtained region of interests (red regions on the images) for escalator (frame size 640×480) and high-way (frame size 320×240) cases. Motion heat map [12] expects very long video to recommend the desirable RoI (hot areas), whereas the MM needs comparatively much less video duration and contrives the RoI very quickly. In general, RoI ameliorates the quality of results and makes the processing time fast a bit more.

2.2. Modeling of spatiotemporal information (STI)

To analyze the scene, we treat moving interest points as the main cue instead of tracking individual subjects. The RoI, ascertained by MM, is divided into small blocks. Once we define n points of interest in the RoI, we track those points over the small blocks of two successive region of interest images using optical flow techniques [17,24,6]. We take down the static and noise features. Static features are the features which moves less than two pixels. Noise features are the isolated features which have a big angle and distance difference with their near neighbors due to tracking calculation errors. Yet, one broad problem in some applications is that people near the camera are supposed to

produce ample optical flow vectors and people far from the camera cannot produce such fully sufficient flow vectors even if they would make very quick motion (e.g., running or falling). That might be right in many examples but it does not generalize. For example, a fronto parallel wall has the same depth everywhere in the RoI, same for a person close to the camera. In the direction of generalization one reasonable solution would be a vertical coordinate system in the image. Moreover, authors in [29] used vertical coordinate to model their motion vector. We can count vertical coordinate system of each block where a weighing coefficient ζ is calculated according to the vertical coordinate of the block. Vertical coordinate system is an implementation stage coordinate system, it depends on several factors of the context of application and implementation e.g., area of RoI, number of defined blocks within RoI, etc. A weighing coefficient $\zeta \leq 1$ is calculated according to the vertical coordinate of the block. A block far away from the camera has small vertical coordinate, as a result its ζ should be large. Equally, block with large vertical coordinate gets smaller ζ . If we see with attention the applications which are related to fronto parallel wall, then ζ is just 1. Finally, for each video frame (e.g., Fig. 3) irrespective of normal or abnormal events, we come into possession of a reliable and workable *spatiotemporal information* (STI), i.e., an $n \times 4$ matrix which is a function of time, broadly speaking a set of vectors \mathbf{V} of n elements variate in time, formulated as

$$\mathbf{V} = \begin{bmatrix} x_1 & y_1 & \delta_1 & \alpha_1 \\ \cdot & \cdot & \cdot & \cdot \\ x_i & y_i & \delta_i & \alpha_i \\ \cdot & \cdot & \cdot & \cdot \\ x_n & y_n & \delta_n & \alpha_n \end{bmatrix} \quad (1)$$

where $i \in n$, $x_i \mapsto x$ coordinate of any feature element i , $y_i \mapsto y$ coordinate of i , $\delta_i \mapsto$ some weighing factor ζ_i is multiplied with the displacement of i from one frame to the next, and $\alpha_i \mapsto$ moving direction of i . We will use *displacement* and *vector length* interchangeably. As simple trigonometric function atan comes into notice few potential problems e.g., infinite slope, false quadrant, etc., the trigonometric function atan 2 has been used to estimate the accurate moving direction α_i of the feature i . On the whole, the function atan 2 gracefully handles infinite slope and places the angle in the correct quadrant [e.g., $\text{atan}(\frac{-1}{-1}) = \pi/4$ differs from $\text{atan} 2(-1, -1) = -3\pi/4$, etc.].

2.3. Statistical treatments of the STI

In this subsection, we will formulate *Entropy*, which is a measure of the disorder or randomness of video sequence, from its two crude elements namely degree of randomness of the directions (*circular variance*) and the degree of randomness of the displacements (*coefficient of displacement variation*).

2.3.1. Degree of randomness of the directions

Consider two cars on the high-way have changed directions with respect to their original directions, i.e., one from 0° to 10° and other from 0° to 340° . The *arithmetic means* of these pairs of direction changes are 5° and 170° , respectively. The direction mean 5° seems intuitively reasonable, while the average of 170° is clearly in error. As the arithmetic mean is ineffective for angles, it is important to find a good method to obtain both the mean value and the measure for the variance of the angles. Assume that two interest points of a frame went somewhere in the next frame with a maneuver of unit vector lengths **A** and **B** having angles α_1 and α_2 , respectively. Their directional mean **R** can be found graphically as shown in Fig. 4. But the graphical solution becomes extremely inefficient when a large number of directions to be added and also often arises the problem of precision. Yet an elementary trigonometric analysis can solve the problem with high accuracies. If $\alpha_1, \dots, \alpha_i, \dots, \alpha_n$, where $i \in n$, be a set of directions of n interest points taken from a single origin, then the *tangent of R*, also called *mean of a series of angles* or *vector mean* and symbolized as θ_R , can be defined by

$$\theta_R = \text{atan} 2 \left(\frac{1}{n} \sum_{i=1}^n \sin(\alpha_i), \frac{1}{n} \sum_{i=1}^n \cos(\alpha_i) \right). \tag{2}$$

An interesting manner is that the sum of the *sines* of the angular deviations from each observation to the resultant is *zero*,

mathematically this property can be shown as: $\sum_{i=1}^n \sin(\alpha_i - \theta_R) = 0$. The variability of a sample of directional measurements is indicated by the length of **R**, which can be defined for n vectors using Pythagorean theorem as: $O_R = \sqrt{[\sum_{i=1}^n \sin(\alpha_i)]^2 + [\sum_{i=1}^n \cos(\alpha_i)]^2}$ which means the larger sample sizes can have longer resultant lengths than smaller samples without having less variability. A standardized measure of variability can solve this unacceptable property. To develop such a measure of variability it is necessary to account for differing sample sizes. Let α_i be a set of directional measurements with sample size n where $i \in n$, then the *degree of randomness of the directions* or *circular variance* C_v is defined as

$$C_v = 1 - \frac{\sqrt{[\sum_{i=1}^n \sin(\alpha_i)]^2 + [\sum_{i=1}^n \cos(\alpha_i)]^2}}{n} = 1 - \frac{O_R}{n}. \tag{3}$$

The O_R/n ranges from 0 to 1. Its extreme values have some agreeable properties. The case $O_R/n = 1$ implies that all the data points are coincident, whereas $O_R/n = 0$ does not imply uniform dispersion around the circle. Hence, O_R/n is not necessarily a useful indicator of dispersion or spread of the data unless they constitute a single group. The C_v provides a smooth (0,1) scale. The smaller is the value of C_v , the more is the concentration of distribution. It is worth mentioning that $0 \leq C_v \leq 1$, unlike an ordinary *linear variance*; and the interpretation of $O_R/n = 0$, the estimation of $C_v = 1$ does not necessarily imply a maximally dispersed distribution.

Forthwith, we wish to pay our attention on: *How differently does the circular variance behave in normal and abnormal situations?* Superposable to the observation of Fig. 3, where directions and displacements of interest points vary randomly in abnormal case and they are almost symmetrically directed in normal case, we have simulated the two cases in simpler way. Fig. 5 depicts the fate of 50 interest points for two cases. The directions of interest points have been simulated in between 0° and 30° with their vector lengths between 0.5 and 1 for normal case. While in abnormal case, directions vary in between 0° and 360° with vector lengths between 0 and 1. On account of simplicity outlier has not been taken into account. Both linear and circular measures have been estimated in each circumstances. In the symmetrically directed directions case, there is almost no pressure on the choice of which preferred direction, either linear or circular, is to be used because they both perform similarly. In other words, in normal case, either linear or circular direction can be the preferred direction painlessly. In spite of that, circular measure is preferable because of its more accuracy. The circular variance $C_v = 0.0123$ illustrates that the flow vectors of the interest points are well concentrated and the interest points are systematically directed. Emphatically, the movements of crowd in video for normal cases are hazard free. On the other hand, there is sufficient difference between linear and circular measures in abnormal case. Nevertheless, the linear mean goes wrong and only choice is the circular measure. The circular variance $C_v = 0.8950$ exemplifies that the flow vectors of the interest points are highly scattered around. Intuitively speaking, the movements of crowd in video for abnormal cases are full of hazard. Heretofore, we can conclude in a gross manner that the circular variance varies consequentially in abnormal circumstances.

How does the circular variance behave, if some (or all) points will move slower or quicker than those of previous frame without changing their directions in the next frame? Normally, vector lengths in running case are larger than that of walking. What does happen, in real world crowd video scenes, if some persons will stop or start running suddenly without changing their direction of movements? Based on the context both situations would be abnormal. For example, some persons stopped running while Marathon running or some persons started running while others walking. Do these cases concern with the circular variance, any way?

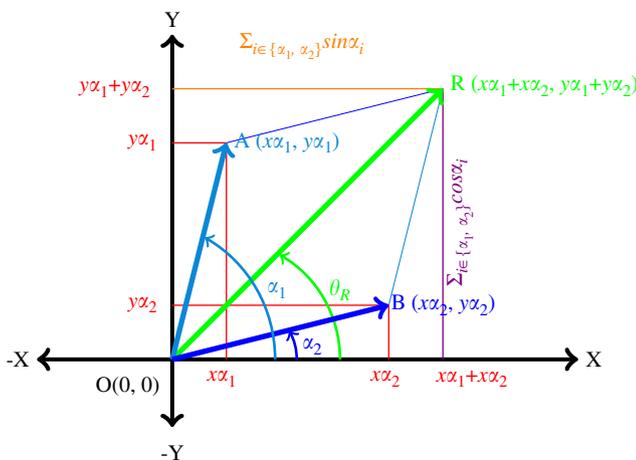


Fig. 4. Elementary vectors and trigonometric analysis.

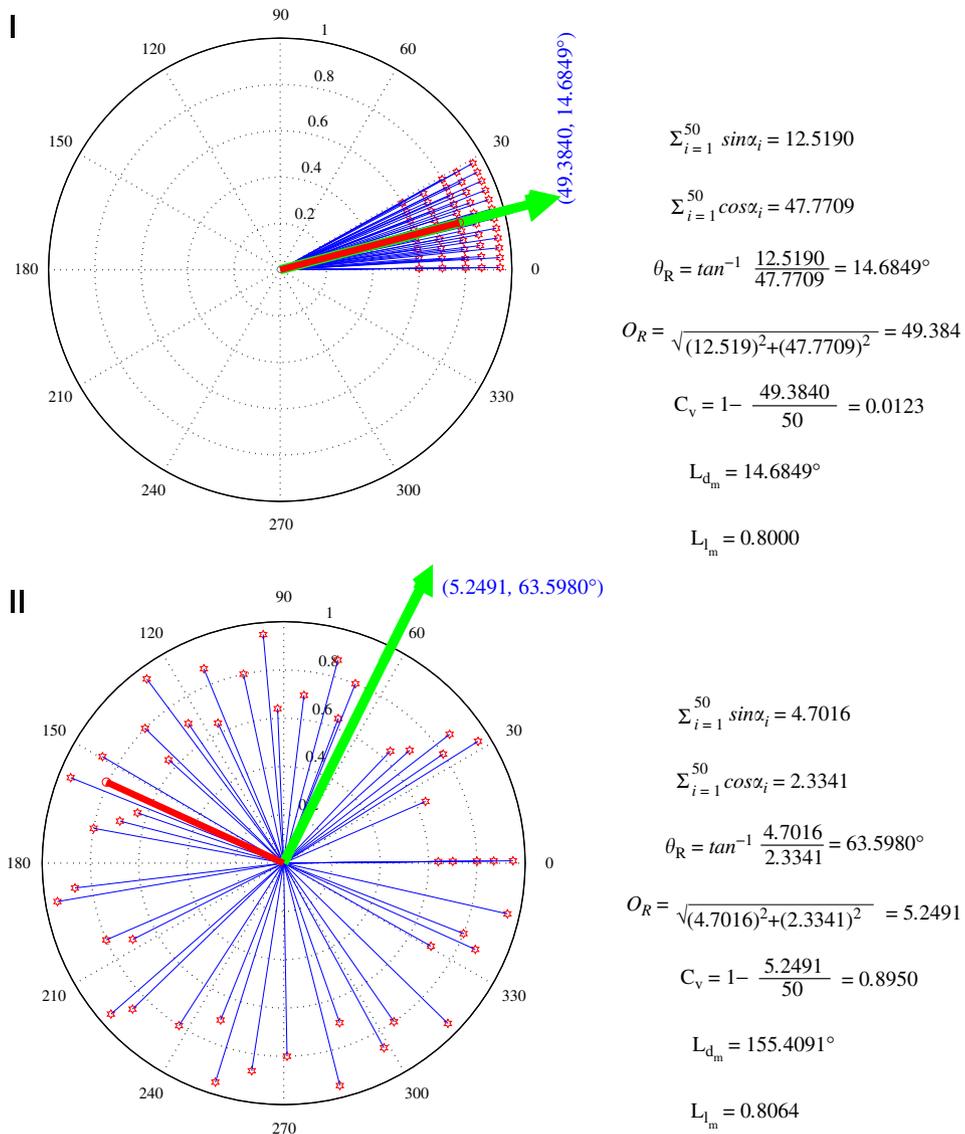


Fig. 5. A simple example of how the circular variances behave in normal and abnormal cases. *Linear mean of directions* (L_{d_m}) and *circular resultant vector lengths* (O_R) are shown using heavy red line and heavy green arrow, respectively. Unlike O_R , the *linear mean of vector lengths* (L_{l_m}) is normalized. There is a significant variation in C_v between normal and abnormal situations. (I) Vectors flow in normal case (e.g., Fig. 3(a)) and some statistical measures. (II) Vectors flow in abnormal case (e.g., Fig. 3(b)) and some statistical measures. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

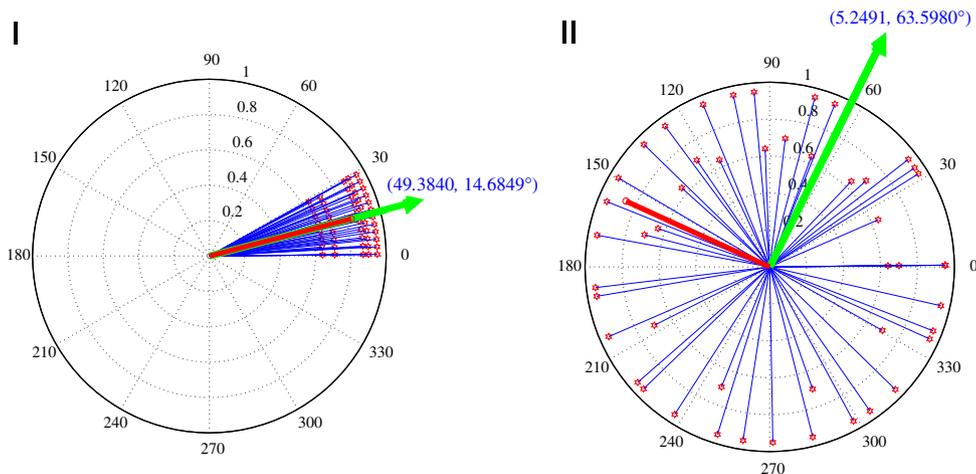


Fig. 6. Linear mean of vector length (L_{l_m}) varies with the length variation of interest point. Conversely, there is no effect on C_v , O_R , and L_{d_m} . Comprehensibly, circular variance does not change with vector lengths variation of interest points but does vary only their directions variation. (I) $C_v = 0.0123$, $L_{l_m} = 0.8302$. (II) $C_v = 0.8950$, $L_{l_m} = 0.8554$.

Let us take into account the vector length variation while direction remains unchanged in both cases of Fig. 5. Images in Fig. 6(I) and (II) depict circumstances where some interest points changed their vector lengths only. The estimated circular variances, circular resultant vector lengths, and linear mean of directions continue the same as estimated in Fig. 5, solely the linear mean of vector lengths has been changed from 0.8000 to 0.8302 and from 0.8064 to 0.8554 for normal and abnormal cases, respectively. From this estimation, it is easy to show that any change of the vector length without varying their directions, the circular variance remains unaffected. Without any shadow of doubt, we can conclude that the circular variance does not bear any information when some persons stopped running while Marathon running or

some persons started running while others walking, if and only if the direction of movement be the same. From this knowledge of observation, we can reach a conclusion that the circular variance is an extremely important factor for direction changing case but exclusively it is not always adequate to pick up abnormality from the real world video scenes. Henceforth, it needs its complement for detecting wide varieties of aberration.

2.3.2. Degree of randomness of the displacements

We have observed that circular variance is a necessary factor but not sufficient for detecting abnormality from the real world videos where both systematic and unsystematic movements

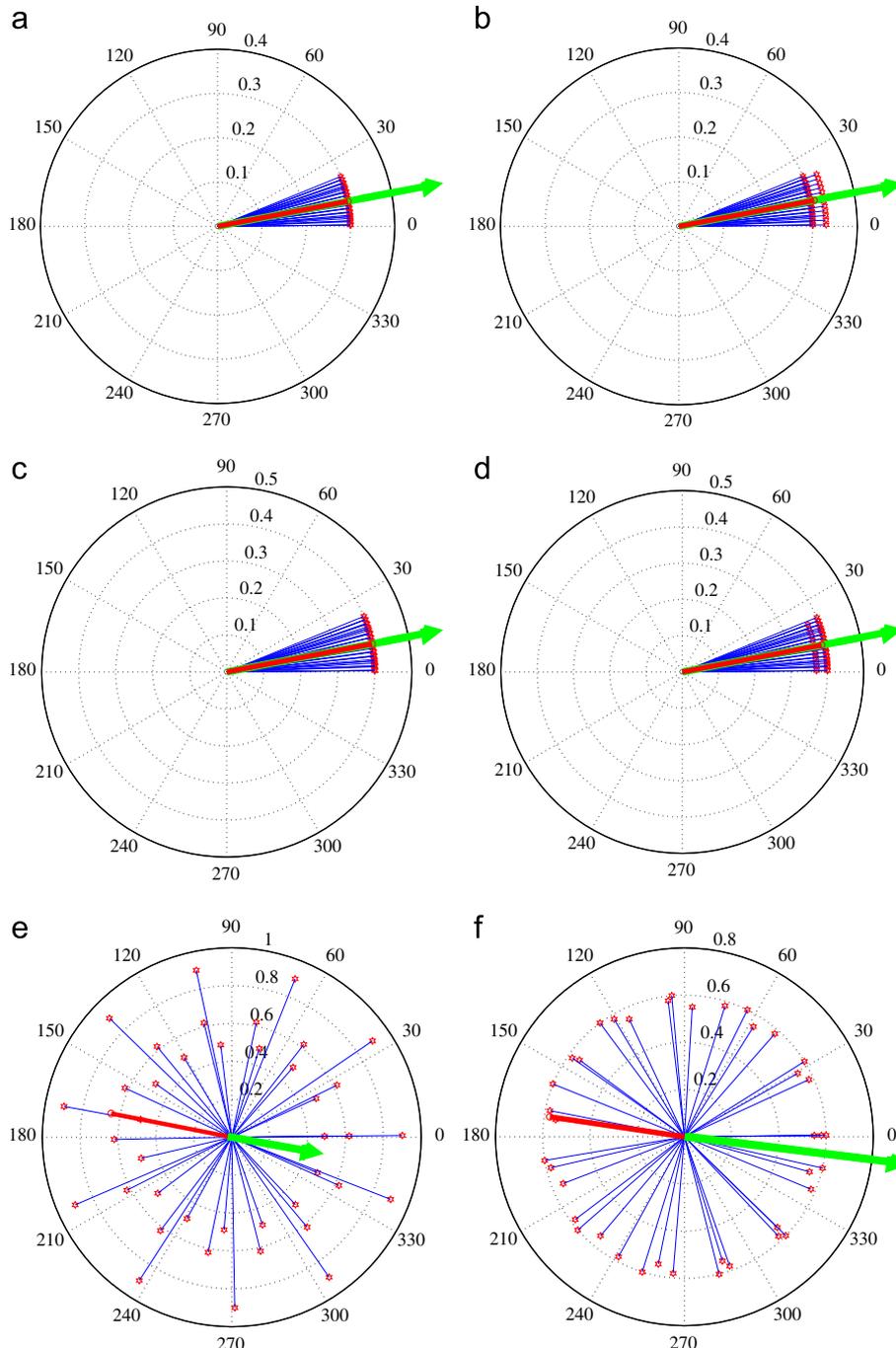


Fig. 7. Simulation of six different instances of the occurrence of an avenue race (e.g., Marathon). (a) All runners walking without changing direction and displacement. (b) Suddenly some runners start running without changing direction. (c) All runners running without changing direction and displacement. (d) Suddenly some runners running slowly without changing direction. (e) Suddenly all runners faced a panic. (f) All runners running without changing direction and displacement.

exist. Along with circular variance, it is important to take into consideration the vector lengths or displacements of the point of interests for exemplifying the aberration detection purposes.

One common query would be: *How does the displacement behave in normal and abnormal cases?* To accord the answer in a good way, let us simulate six different instances of the occurrence of a straight avenue race (e.g., Marathon) and the number of participating runners is 40. Beginning of the run all runners were walking with some 0.30 unit displacement per frame without changing their directions as simulated in Fig. 7(a). In real world scene, this type of event usually holds up no surprisal and thus it is normal. At certain frame, suddenly some runners started running with some 0.33 unit displacement per frame without changing their directions as simulated in Fig. 7(b). Such type of event poses some degree of visual attention for the primates and accordingly it would be abnormal. Afterwards, all runners were running with some 0.40 unit displacement per frame without changing their directions as simulated in Fig. 7(c). It is a usual event like Fig. 7(a) as very systematic run or walk does not sustain interesting facts. After a while, some runners grew fatigued and at certain frame suddenly they decided to run slowly at 0.37 unit displacement per frame without changing their directions as simulated in Fig. 7(d). Such type of change in the crowd has connection with some interesting information for the primates and in this way it would be an abnormal event. At certain frame, all runners faced a sudden panic situation (e.g., explosion, gun shot, fire) and accordingly they were randomly scattered, i.e., they changed their directions as well as displacements as simulated in Fig. 7(e). This type of variation in the crowd bears very high degree of interest for the primates and to this extent it is necessarily an abnormal event. After the explosion, all the scattered runners were running without changing their directions and displacements (maybe varying displacements with respect to others but fixed for each runner) over frames as simulated in Fig. 7(f). This event is similar to Fig. 7(a) and (c). It does not endorse interesting information, so it is normal.

From Fig. 7 it is noticeable that if an event where both direction and displacement vary, then there will be high possibility to become that event an abnormal. So it is important to consider carefully both direction and displacement simultaneously. The directional measure circular variance (C_v) is both dimensionless and normalized. On the other hand, displacement is neither a dimensionless nor a normalized quantity. Henceforth, we put forward a reasonable solution of these problems in a different way by taking ratio between two statistical measures of displacements. The displacement variance to mean ratio would be a good solution. Customarily, variance to mean ratio is a measure used to quantify whether a set of observed occurrences are clustered or dispersed compared to a standard statistical model. It provides a good measure of the degree of randomness of the displacements and may be dealt with normalization. But the variance of a variable has different units from the variable, for example square centimeters when the variable is in centimeters. As a result, the displacement variance to mean ratio has unit of centimeter. Since the displacement variance to mean ratio is dimensional, the unit does not cancel, the ratio is not scale invariant. Scale invariance is a feature of rules which do not change if length scales are multiplied by a mutual factor. One possible good solution of the scale invariance for this problem would be the standard deviation (the square root of variance) which is a widely used measure of the variability or dispersion. A useful property of standard deviation is that, unlike variance, it is expressed in the same units as the data (using the *mean* as a measure of scale). The unit of the displacement standard deviation to mean ratio is canceled out as they are measured in the same scale and is thus a pure number. Evidently, the obtained

ratio is now scale invariant. The displacement standard deviation to mean ratio (coefficient of displacement variation) is not only a dimensionless quantity but also can provide a good measure of the degree of randomness of the displacements. Having a complement factor of circular variance for a wide variety of aberration detections, the coefficient of displacement variation plays an important role to detect some kind of abnormalities from real videos. Deeming Eq. (1), the *mean* of displacements $\bar{\delta}$ is delimited by dint of:

$$\bar{\delta} = \frac{1}{n} \sum_{i=1}^n \delta_i \quad (4)$$

where n is the number of optical flow vectors in the frame. With this *mean* it is easy to ascertain displacement of *standard deviation* by means of

$$\delta_{std} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (\delta_i - \bar{\delta})^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n \delta_i^2 - \frac{n}{n-1} \bar{\delta}^2}. \quad (5)$$

The displacement standard deviation to mean ratio or *degree of randomness of the displacements* is formulated as the ratio of the standard deviation to the mean as

$$D_r = \frac{\delta_{std}}{\bar{\delta}} \quad (6)$$

where $\bar{\delta} > 0$. Accordingly, D_r is scale invariant and normalized particularly for positive distribution such as the exponential distribution and Poisson distribution.

2.4. Entropy estimation

Up until now, it is clear that *circular variance* and *coefficient of displacement variation* are necessary and sufficient factors to detect various aberrations in videos. How can the effective power of them get mixed together? One possible solution would be the usage of those statistical measures as the crude parameters of the *Entropy*. The more is the entropy, the more is the disorder/chaos in the system. For instance, to have order on the high-way means to have cars follow the order of lanes, speed limits, directions, etc. When these things get mixed in, entropy increases causing disorder/chaos on the high-way traffic system.

The term entropy was coined in 1865 by German physicist Rudolph Clausius [10]. Thermodynamic entropy indicates a measure of how organized or disorganized a system of atoms or molecules is. It has an enabling factor of energy. Information (also called Shannon) entropy with no inherent or integral energy factor, thus it is solely related in form and not in function. Shannon entropy, a measure of uncertainty, is the expectation value of $-\log_e(p)$, where p is the probability assigned to the measured value of a random variable. Shannon entropy is a broad and general concept which finds applications in information theory and thermodynamics. Shannon entropy and information uncertainty can be used interchangeably [14]. Definition of the Shannon entropy E_S is quite usual, and is expressed in terms of a discrete set of n probabilities p_i with $i \in n$ as

$$\begin{aligned} E_S &= p(x_1) \log_e \frac{1}{p(x_1)} + p(x_2) \log_e \frac{1}{p(x_2)} + \dots + p(x_n) \log_e \frac{1}{p(x_n)} \\ &= - \sum_{i=1}^n p(x_i) \log_e p(x_i) \end{aligned} \quad (7)$$

where $\sum_{i=1}^n p(x_i) = 1$. If $p(x_i) = 0$ for some i , the value of the corresponding summand $0 \log_e 0$ is taken to be 0. The entropy is zero signifies there is no uncertainty and hence there is no information. Consequently, entropy always follows the nonnegativity rule ($E_S \geq 0$).

To fit for the statistical measures of C_v and D_r in Eq. (7), the measures have been modeled with their respective probabilities as

$$p(c_v) = \frac{C_v}{C_v + D_r} \tag{8}$$

$$p(d_r) = \frac{D_r}{C_v + D_r} \tag{9}$$

Then the Shannon entropy at some frame f can be formulated as

$$E_f = p(c_v) \log_e \frac{1}{p(c_v)} + p(d_r) \log_e \frac{1}{p(d_r)} \tag{10}$$

where $p(c_v) + p(d_r) = 1$. The more is the E_f , the more is the disorder/chaos on the video frame. Higher value of E_f means the corresponding video frame has a high possibility to become a frame of abnormal case. To define and employ Shannon entropy by Eq. (10) is an agreeable way; where a change in the C_v or D_r (which directly reflects in E_f), separately or together, reports an abnormal activity.

Theoretically, in some situations the proposed entropy E_f can be both sensitive and incorrect. From Eq. (10) it is evident that either $C_v=0$ or $D_r=0$ can make the proposed entropy $E_f=0$. Therefore, if the values of C_v or D_r are the same (i.e., move towards the same direction or at the same velocity), then E_f will become zero. The estimation of $E_f=0$ would be very sensitive. For example, if $C_v=0$ and/or $D_r=0$, then minimum value of $E_f=0$ and a small disturbance can change the value of E_f from its minimum to maximum e.g., assume that $C_v = 6 \times e^{-6}$ and $D_r = 6 \times e^{-6}$, subsequently E_f will approach to its maximum value $\log_e 2$. When $C_v=D_r$ will occur in a frame, Eq. (10) will move aside $E_f = \log_e 2$. If there will exist an estimation of $E_f=0$ just before that frame, then the result of that frame will be counted as a finite impulse (an outlier) value. If there would result an impulse or close to an impulse value, then the decision might be either true positive or false alarm. If there would exist a ground truth, then that impulse value should be a true positive, otherwise a false alarm. Ground truth is the process of manually marking what an algorithm is expected to output. Theoretically, when the value of C_v and D_r are very small, a small disturbance will make a great change on E_f . However, in practice, due to optical flow estimation error (e.g., problems of illumination, motion discontinuities [16], etc.) and some other factors, the value of C_v and/or D_r usually is larger than the ground truth, henceforth, it is very unlikely to get a very small value and therefore becomes more stable than expected in theory.

2.5. Correct–incorrect estimation of E_f

In some cases the estimation of E_f could be incorrect. For example, when people are running towards the same direction at different speeds, this should be considered as abnormal. Nevertheless, the estimation of entropy in Eq. (10) will be both abnormal (correct) and normal (incorrect). Let us simulate this situation for 1024 people at six different races (contest of speed). Assume that 1024 people are running with various $D_r(i) | i = 1, 2, 3, \dots, 1024$ speeds at their 1st, 2nd, 3rd, 4th, 5th, and 6th races in $C_v(1) = 6 \times 10^{-1}$, $C_v(2) = 6 \times 10^{-2}$, $C_v(3) = 6 \times 10^{-3}$, $C_v(4) = 6 \times 10^{-4}$, $C_v(5) = 6 \times 10^{-5}$, and $C_v(6) = 6 \times 10^{-6}$ directions, respectively. During the simulation 1024 random values for $D_r(i)$ between 0 and 1 have been generated from gamma distributions with shape parameters 2.8.

Simulation results have been presented in both numerically and graphically. Due to the simplicity of presentation only first 10 values of $D_r(i)$ have been listed in Table 1. But Fig. 8 depicts all simulated results graphically. To avoid complexity data have been plotted after accomplishing an ascending order of $D_r(i)$ values. However, in the 1st case the outcomes of Eq. (10) make clear and visible of abnormal case as listed in the light gray colored column $E_f(i,1)$ in Table 1 as well as cyan colored curve in Fig. 8, which is a correct estimation. In the 2nd case the outcomes of Eq. (10) equally give evidence of abnormal case as listed in the light gray colored column $E_f(i,2)$ in Table 1 and also black colored curve in Fig. 8, which is again a correct estimation. In the 3rd case the outcomes of Eq. (10) may be deemed as an abnormal case as listed in the light gray colored column $E_f(i,3)$ in Table 1 as well as magenta colored curve in Fig. 8, which would nearly be a correct estimation. In the 4th case the outcomes of Eq. (10) represent normal case as listed in the dark red colored column $E_f(i,4)$ in Table 1 and also green colored curve in Fig. 8, which is an incorrect estimation. Similarly, in 5th and 6th cases the estimated outcomes of $E_f(i,5)$ and $E_f(i,6)$ are incorrect.

Explicitly, the definition of entropy in Eq. (10) can provide correct results up to certain range. Both Table 1 and Fig. 8 provide evident that this workable range could not go less than some 10^{-3} . Up to this point, we could apply a threshold on the obtained E_f measure to get a decision whether the frame belongs to normal or abnormal situations. But such kind of decision would potentially lead the method sensitive and approximately be equal to thresholding of $p(c_v)$ and/or $p(d_r)$. Consequently, false alarm rate would eventually be high. Nonetheless, to obtain better performance we would like to improve the workable range a bit more by performing an explicit normalization for entropy E_f .

Table 1
Simulation of people running towards the same $C_v(j)$ direction at varying speeds $D_r(i) | i = 1, 2, 3, \dots, 10$ using Eq. (10) where $j = 1, 2, \dots, 6$. Bold and italic columns are correct and incorrect estimations or workable and not-workable ranges, respectively.

$D_r(i) i = 1, 2, 3, \dots, 10$	$C_v(j) j = 1, 2, \dots, 6$					
	$C_v(1) = 6 \times 10^{-1}$	$C_v(2) = 6 \times 10^{-2}$	$C_v(3) = 6 \times 10^{-3}$	$C_v(4) = 6 \times 10^{-4}$	$C_v(5) = 6 \times 10^{-5}$	$C_v(6) = 6 \times 10^{-6}$
	$E_f(D_r(i), C_v(j))$					
$D_r(1) = 0.1685$	0.5260	0.5757	0.1496	0.0236	0.0032	0.0004
$D_r(2) = 0.0912$	0.3902	0.6716	0.2316	0.0394	0.0055	0.0007
$D_r(3) = 0.0164$	0.1229	0.5204	0.5808	0.1525	0.0241	0.0033
$D_r(4) = 0.1331$	0.4738	0.6196	0.1778	0.0287	0.0039	0.0005
$D_r(5) = 0.0743$	0.3469	0.6875	0.2657	0.0467	0.0066	0.0008
$D_r(6) = 0.0500$	0.2713	0.6890	0.3404	0.0643	0.0093	0.0012
$D_r(7) = 0.3855$	0.6693	0.3952	0.0792	0.0116	0.0015	0.0002
$D_r(8) = 0.0507$	0.2735	0.6896	0.3379	0.0637	0.0092	0.0012
$D_r(9) = 0.0038$	0.0383	0.2263	0.6681	0.3975	0.0799	0.0117
$D_r(10) = 0.0892$	0.3853	0.6739	0.2352	0.0401	0.0056	0.0007

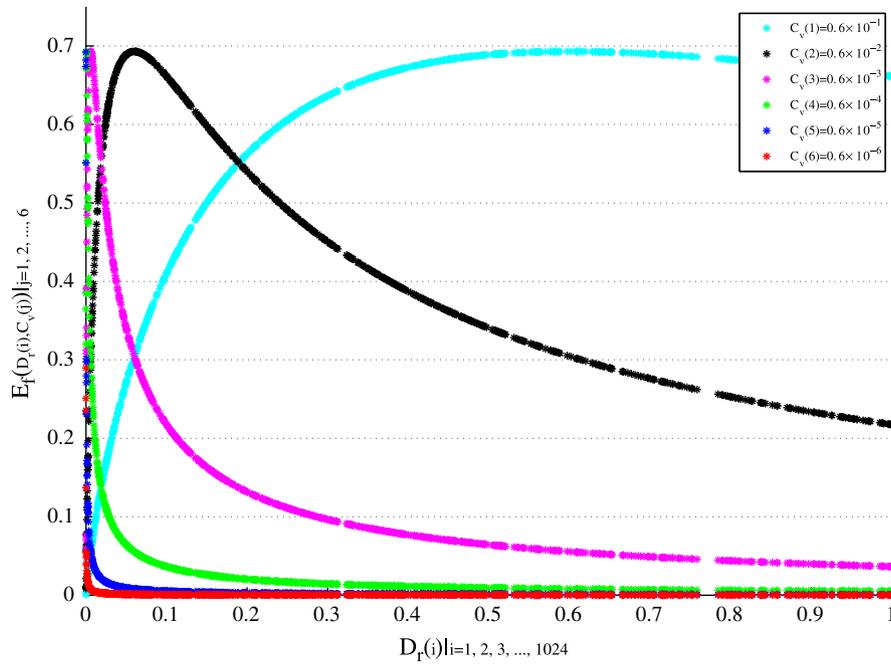


Fig. 8. Simulation result of 1024 people running towards each time one of $C_v(j) | j = 1, 2, \dots, 6$ directions at $D_r(i) | i = 1, 2, 3, \dots, 1024$ various speeds using Eq. (10). On doing an ascending order of $D_r(i)$ values, data have been plotted. Workable range does not pass less than some 10^{-3} , thus green, blue, and red colored curves fall on the not-workable range, i.e., incorrect estimation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.6. Explicit normalization

A normalized value is a value that has been processed in a way that makes it possible to be efficiently compared against other values. But the Shannon entropy is not normalized [23], i.e., Eq. (10) needs a little amelioration to have a normalized structure. For instant, consider a probability space where exists $p(x_1) = 0.51$, $p(x_2) = 0.26$, and $p(x_3) = 0.23$; then Eq. (7) estimates $E_S = 1.0317$, which is not normalized. For the sake of normalization, we may use the functions e.g., $1/\log_e E_f$, $1/(1 + \log_e E_f)$, etc. But the selection of the function $1/\log_e E_f$ for normalization is an unacceptable concept because to get something normalized from non-normalized state or condition, we need a normalizing function or factor; yet logarithm itself is not such a case. In addition, we wish to bring into existence the system parameters on the user choices. None of these functions do offer a friendly change option of E_f measure for the user, i.e., scaling problem. To solve these problems, we would like to take up a versatile distribution which has significant effect on its shape and scale parameters. In this respect, we take the advantage of the *cumulative distribution function of Weibull distribution* [26], which has strict lower and upper bounds between 0 and 1. Due to accurate model quality and performance characteristics of Weibull distribution and its flexibility that makes it ideal for analysis on a dataset with unknown distribution. It is worth mentioning that Weibull distribution can mimic the behavior of other statistical distributions such as the normal and the exponential [26]. At this time, we can essentially formulate the *normalized entropy* of some frame f by means of

$$[Entropy]_f = 1 - e^{-(E_f/\lambda)^\vartheta} \quad (11)$$

where $E_f \geq 0$ as well as $\lambda > 0$ and $\vartheta > 0$ respectively, denote scale and shape parameters of the distribution. Using Eq. (11) and being acquainted with values of λ , ϑ , and E_f we can desirably estimate the normalized entropy of any frame $[Entropy]_f$ between 0 and 1. Thus Weibull distribution not only provides a fair normalized measure for E_f between its strict lower and upper

bounds but also offers a friendly change option of that measure for the user by its λ and ϑ parameters. How does $[Entropy]_f$ vary in the effect of varying λ and ϑ with some fixed value of E_f ? To study such behavior, 50 pairs of λ and ϑ values have been generated on the 0.05 incremental basis. Figs. 9 and 10 depict the characteristics.

Using Eq. (11), we can widely and explicitly estimate *Entropy* of the simulated situations listed in Table 1 and Fig. 8. Numerical values in Table 2 and figurative Fig. 11 clarify further improvement of the workable range where user friendly parameters have been selected as $\lambda = 0.45$ and $\vartheta = 0.20$. Table 2 as well as Fig. 11 suggest that the new workable range would go less than some 10^{-6} . However, it has been already widened the workable range about a factor of $10^{-3}/10^{-6} = 1000$. It is noticeable that the final estimation depends not only on the estimated E_f but also on the significant effect of the *Entropy* definition in Eq. (11). The Eq. (11) can provide more stable estimation than that of Eq. (10). Nevertheless, there is a disbursement to be paid for these improvements. The selection of low or high values for λ and ϑ would give rise to problems for both false alarms and thresholding; and hence it is important to make a trade-off.

In an easily perceptible manner, we can estimate *Entropy* of the simulated situations as simulated in Figs. 5–7 with the help of Eq. (11). Table 3 shows their results for three different values of λ and ϑ . Based on λ and ϑ , the values of *Entropy* increase for sometimes and decrease for other times. This phenomena has circumscribed the freely selection options of λ and ϑ in some degree and thus it is related to the performance. If λ and ϑ will possess high and low values, respectively, then higher value of E_f will be compressed but lower value of E_f will be expanded exceedingly. All *Entropy* values will fall into somewhat very short range. If λ and ϑ will own low and high values, respectively, then higher value of E_f will be expanded highly and lower value of E_f will be compressed to a high degree. All *Entropy* values will fall into wide range. If both λ and ϑ will bear high values, then all values of E_f will be compressed. All *Entropy* values will fall into wide range. If both λ and ϑ will contain low values, then higher value of E_f will be expanded in lesser extent but lower value of E_f

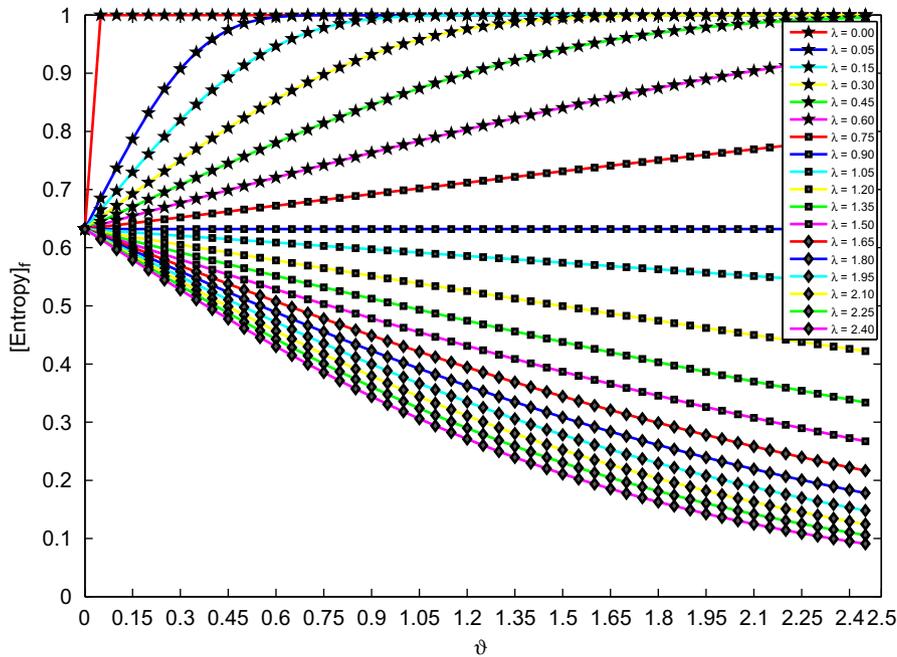


Fig. 9. Variation of $[Entropy]_f$ in Eq. (11) for $E_f=0.90$ with varying ϑ and each constant λ . If $\vartheta = 0$, then $[Entropy]_f = 0.6321$. $E_f = \lambda$ with varying ϑ results a ϑ -axis parallel line. All λ -fixed curves with $\lambda < 0.90$ and $\lambda > 0.90$ situate above and beneath of this line, respectively.

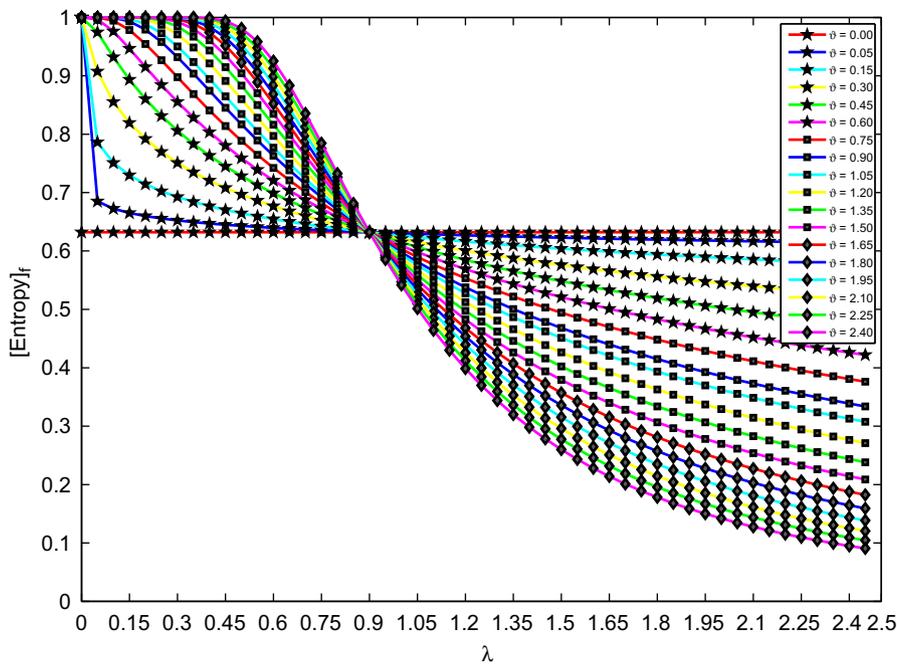


Fig. 10. Variation of $[Entropy]_f$ in Eq. (11) for $E_f=0.90$ with varying λ and each fixed ϑ . Exactly corresponding points are orchestrated by fixing $E_f = \lambda$ and varying λ . If $\lambda = 0$, then $[Entropy]_f = 1$. If $\vartheta = 0$, then $[Entropy]_f = 0.6321$ and thereof by varying λ scores a λ -axis parallel line. Before coincidence all ϑ -fixed curves were above the line, after they are underneath.

will be expanded to a high extent. All *Entropy* values will fall into somewhat short range. The phenomena of higher value compression and lower value expansion brings about high precision rate by minimizing false alarms but thresholding problem may occur as values of the *Entropy* will fall into a narrow range. But then again, the phenomena of higher value expansion and lower value compression puts through low precision rate by increasing false alarms but the thresholding problem can be solved in some good manner as values of the *Entropy* can fall into a wide range. Accordingly, it is important to select λ and ϑ values in a way so

that precision rate will be as high as possible and simultaneously thresholding problem can be minimized. Several simulation results on large sets of random data obtained from gamma distributions suggest that $0.35 \leq \lambda \leq 1.5$ and $0.56 \leq \vartheta \leq 2.2$ can be a good range of options for these user friendly parameters.

2.7. Supremacy of explicit normalization

From Fig. 8 at fixed direction e.g., $C_v(4) = 6 \times 10^{-4}$ (green curve) case, it can be noticed that in the distribution of E_f

Table 2

Simulation of people running towards the identical $C_v(j)$ direction at varying speeds $D_r(i)|i = 1, 2, 3, \dots, 10$ with Eq. (11) where $j = 1, 2, \dots, 6$ and user friendly parameters $\lambda = 0.45$ and $\vartheta = 0.20$. Bold columns are correct estimation or workable range.

$D_r(i) i = 1, 2, 3, \dots, 10$	$C_v(j) j = 1, 2, \dots, 6$					
	$C_v(1) = 6 \times 10^{-1}$	$C_v(2) = 6 \times 10^{-2}$	$C_v(3) = 6 \times 10^{-3}$	$C_v(4) = 6 \times 10^{-4}$	$C_v(5) = 6 \times 10^{-5}$	$C_v(6) = 6 \times 10^{-6}$
[Entropy] _f ($D_r(i), C_v(j)$) $\lambda = 0.45 \vartheta = 0.20$						
$D_r(1) = 0.1685$	0.6436	0.6502	0.5517	0.4256	0.3103	0.2176
$D_r(2) = 0.0912$	0.6216	0.6616	0.5834	0.4590	0.3390	0.2399
$D_r(3) = 0.0164$	0.5375	0.6428	0.6509	0.5532	0.4270	0.3115
$D_r(4) = 0.1331$	0.6359	0.6556	0.5642	0.4383	0.3211	0.2260
$D_r(5) = 0.0743$	0.6130	0.6633	0.5934	0.4703	0.3490	0.2478
$D_r(6) = 0.0500$	0.5949	0.6634	0.6116	0.4923	0.3687	0.2635
$D_r(7) = 0.3855$	0.6613	0.6226	0.5067	0.3819	0.2742	0.1902
$D_r(8) = 0.0507$	0.5956	0.6635	0.6110	0.4915	0.3680	0.2629
$D_r(9) = 0.0038$	0.4569	0.5815	0.6611	0.6231	0.5075	0.3827
$D_r(10) = 0.0892$	0.6207	0.6618	0.5845	0.4602	0.3401	0.2408

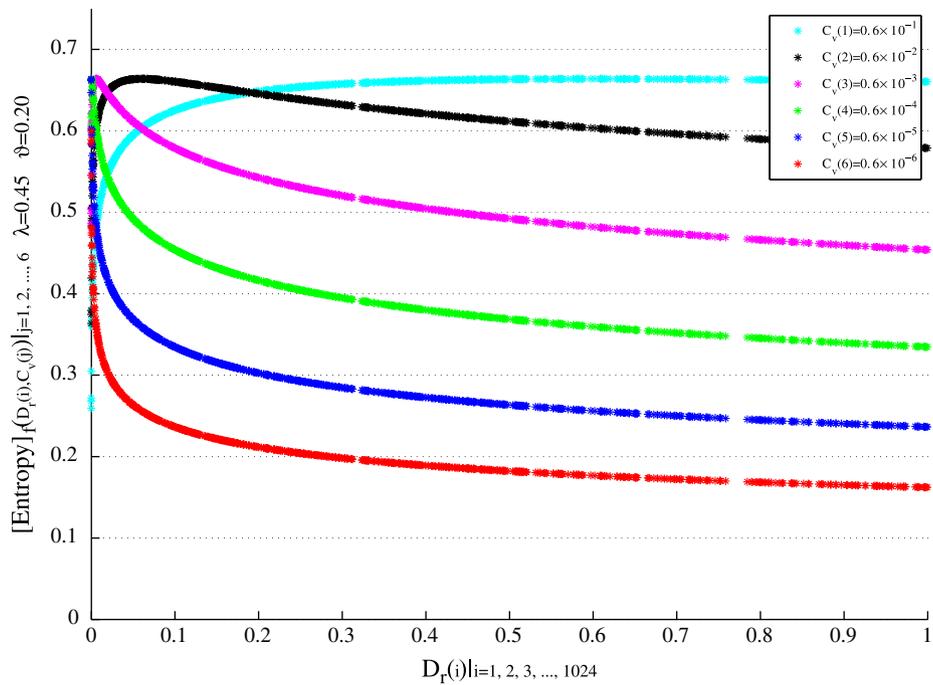


Fig. 11. Simulation results of 1024 people running towards each time one of $C_v(j)|j = 1, 2, \dots, 6$ directions at $D_r(i)|i = 1, 2, 3, \dots, 1024$ various speeds using Eq. (11) as well as $\lambda = 0.45$ and $\vartheta = 0.20$. Data have been plotted on performing an ascending order of $D_r(i)$ values. Correct estimation or workable range may go less than some 10^{-6} and so 1000 times better performance.

Table 3

Entropy estimation of the simulated situations as simulated in Figs. 5–7.

Diverse cases	Required parameters					[Entropy] _f				Conclusive remarks
	C_r Eq. (3)	D_r Eq. (6)	$p(c_r)$ Eq. (8)	$p(d_r)$ Eq. (9)	E_f Eq. (10)	$\lambda = 0.4 \vartheta = 0.2$	$\lambda = 0.5 \vartheta = 1.5$	$\lambda = 0.5 \vartheta = 2.2$	$\lambda = 0.6 \vartheta = 2.5$	
Fig. 5(I)	0.0123	0.1525	0.0748	0.9252	0.2658	0.6020	0.3208	0.2199	0.1221	Normal
Fig. 5(II)	0.8950	0.1386	0.8659	0.1341	0.3940	0.6319	0.5033	0.4470	0.2951	Abnormal
Fig. 6(I)	0.0123	0.1399	0.0809	0.9191	0.2810	0.6031	0.3435	0.2449	0.1391	Normal
Fig. 6(II)	0.8950	0.1570	0.8508	0.1492	0.4214	0.6401	0.5387	0.4966	0.3386	Abnormal
Fig. 7(a)	0.0063	4×10^{-6}	0.9994	6×10^{-4}	0.0051	0.3438	0.0029	5×10^{-5}	7×10^{-6}	Normal
Fig. 7(b)	0.0063	0.0451	0.1230	0.8770	0.3729	0.6302	0.4736	0.4065	0.2612	Abnormal
Fig. 7(c)	0.0063	3×10^{-6}	0.9995	5×10^{-4}	0.0039	0.3301	0.0017	3×10^{-5}	4×10^{-6}	Normal
Fig. 7(d)	0.0063	0.0356	0.1508	0.8492	0.4240	0.6413	0.5412	0.5001	0.3417	Abnormal
Fig. 7(e)	0.9735	0.2959	0.7669	0.2331	0.5430	0.6546	0.6775	0.6985	0.5412	Abnormal
Fig. 7(f)	0.9735	0.0351	0.9652	0.0348	0.1510	0.5629	0.1530	0.0693	0.0313	Normal

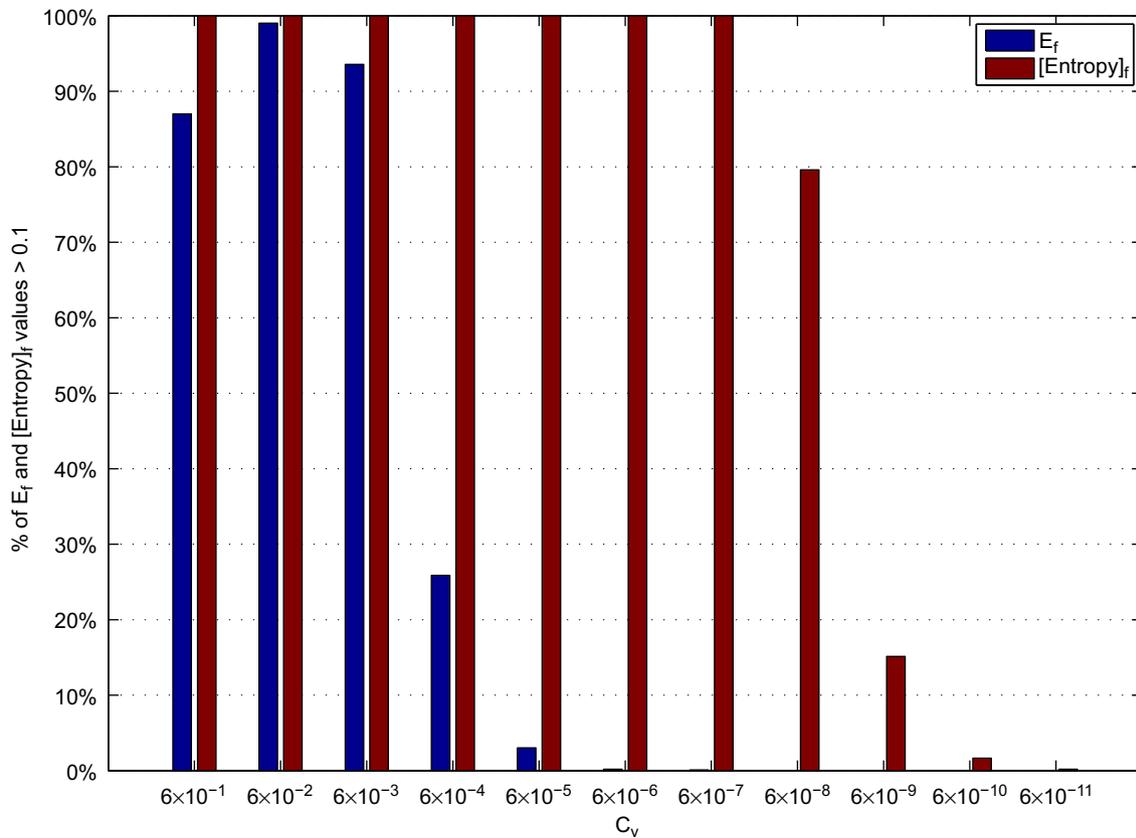


Fig. 12. Percentages of E_f and $[Entropy]_f$ values in their individual D_r and C_v varying distributions which exceed $\xi = 0.1$ applying the simulation results with wider C_v range in Figs. 8 and 11. Usable values of both E_f and $[Entropy]_f$ decrease on decreasing C_v , nearly after 10^{-3} and 10^{-8} , respectively. Thus $[Entropy]_f$ gives 10^5 times better result at $\xi = 0.1$ than that of E_f .

essentially 74.12% and 88.57% of the values fall into less than 0.1 and 0.2 threshold labels, respectively. Cutoff like 0.1 is often impractical in both simulation and real world video applications for anomalies detection, as a consequence more than 74.12% of E_f values fall into simple not-workable range. Conversely, from Fig. 11 at the same situation in the distribution of $[Entropy]_f$ all values are in running order. Accordingly, Eq. (11) reasons out the incompleteness and imperfection of Eq. (10) (e.g., Fig. 8) by improving the workable range in a clearly noticeable manner (e.g., Fig. 11).

How does the explicit normalization improve the stability of E_f and in which extent? Overall pleasantness resulting from explicit normalization would not be quickly seen from Figs. 8 and 11. Figs. 9 and 10 may not help too much in this talking point, since the value of E_f is fixed in both cases. In order to pay close attention to the overall improvement factors of $[Entropy]_f$ with respect to E_f , we have investigated two measures namely Γ_ξ and A_ξ where the cutoff ξ would see with attention the upper borderline of the not-workable range in the distributions of E_f and $[Entropy]_f$. The Γ_ξ affords access to information how many E_f and $[Entropy]_f$ values, measured in percentage, run into workable or not-workable ranges under certain cutoff ξ in the distributions of E_f and $[Entropy]_f$. In a formalistic manner Γ_ξ can be formulated as

$$\Gamma_\xi = \frac{\text{Average \% of } [Entropy]_f \text{ values which exceed } \xi}{\text{Average \% of } E_f \text{ values which exceed } \xi} \quad (12)$$

The A_ξ provides the ratio of the C_v value at any given percentage of E_f values in the distribution of E_f which exceed ξ and the C_v value at the same percentage found in $[Entropy]_f$ distribution

which exceed ξ . Formally A_ξ can be defined as

$$A_\xi = \frac{C_v \text{ value at any given maximum \% of } E_f \text{ values which exceed } \xi}{C_v \text{ value at the same \% found in } [Entropy]_f \text{ distribution which exceed } \xi} \quad (13)$$

Taking into account the simulation results in Figs. 8 and 11 considering C_v range up to and including 10^{-11} , the derived Fig. 12 bears witness the percentage of E_f and $[Entropy]_f$ values in their corresponding D_r and C_v varying distributions which go beyond $\xi = 0.1$. After certain point usable values of both E_f and $[Entropy]_f$ decrease on decreasing C_v . In case of $\xi = 0.1$ the decay starts approximately from 10^{-3} and 10^{-8} for E_f and $[Entropy]_f$, respectively. This phenomena provides $\frac{10^{-3}}{10^{-8}} = 10^5$ times better quality of result for $[Entropy]_f$ at cutoff label $\xi = 0.1$. But in case of $\xi = 0.3$ the similar decay starts approximately from 10^{-2} and 10^{-5} for E_f and $[Entropy]_f$, respectively and thus the gain is 10^3 . Yet, average percentage of E_f and $[Entropy]_f$ values which tower above $\xi = 0.1$ are 28.0717 and 72.4165, respectively. Applying Eq. (12) the improvement factor puts through $\Gamma_{0.1} = \frac{72.4165}{28.0717} = 2.5797$. On the other hand, at $C_v = 6 \times 10^{-6}$ the maximum percentage of E_f values which exceed $\xi = 0.1$ is 0.1953. In the distribution of $[Entropy]_f$ exactly 0.1953 can be found at $C_v = 6 \times 10^{-11}$. On applying Eq. (13), this phenomena gives evidence of $A_{0.1} = (6 \times 10^{-6}) / (6 \times 10^{-11}) = 10^5$ times better quality of outcome for $[Entropy]_f$ at cutoff label $\xi = 0.1$, i.e., if the upper bound of not-workable range is $\xi = 0.1$, then improvement factor $A_{0.1} = 100\,000$.

Fig. 13 describes in vivid detail of improvement factors of the explicit normalization resulting from the simulation results with C_v limit up to and including 10^{-11} in Section 2.5 along with

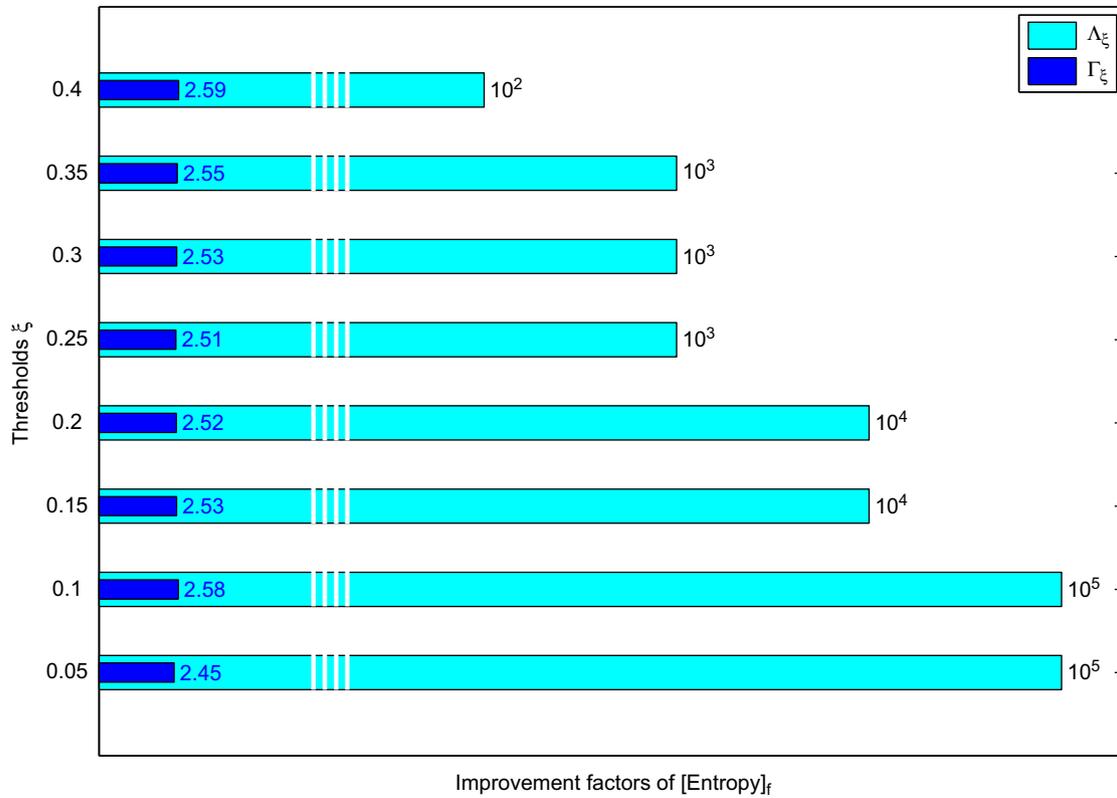


Fig. 13. Imaginable various ξ values and their corresponding improvement factors came into possession of explicit normalization deeming simulation results with wider C_r limit in Section 2.5, Eqs. (11)–(13) with $\lambda = 0.45$ plus $\vartheta = 0.20$. Bar ellipsis exhibits large values exhibition.

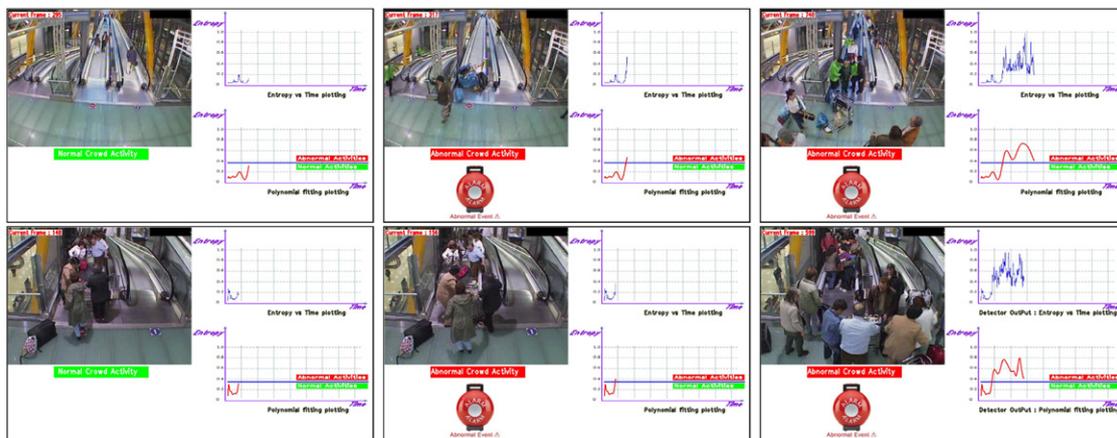


Fig. 14. Two sample videos from the escalator dataset: first row concerns a person falling episode on the escalator egress; second row presents an aberrant situation caused by a wheel broken trolley.

Eq. (11)–(13) plus $\lambda = 0.45$ as well as $\vartheta = 0.20$. Without any shadow of doubt from Fig. 13 it can be easily concluded that $[Entropy]_f$ in Eq. (11) has a better performance.

2.8. Polynomial fitting

We can apply a threshold on the obtained *Entropy* measures data to get a decision of normal or abnormal event frame. But any discrete value of *Entropy* which exceeds a predefined threshold T_E is not a clear evidence of abnormal event frame. It may frequently fear that at least one attribute (e.g., an outlier) may have been severely corrupted by a mistake or error (e.g., tracking calculation

errors) which would lead an erroneous decision of the normal or abnormal event frame. An outlier is a sample that is very different from the average sample in the dataset. An outlier may be an ordinary sample, but of which at least one attribute has been severely corrupted by a mistake or error (e.g., tracking calculation errors). An outlier may also be a bona fide sample, that simply turns out to be exceptional. To minimize this outlier problem, a polynomial fitting would be a good solution. Runge’s phenomenon [22] shows that lower-order polynomials are normally to be preferred instead of augmenting the degree of the interpolation polynomial, even if some of the badness of this interpolation may be overcome by using Chebyshev polynomials [7] instead of

Table 4 Performance evaluation of the method using escalator dataset. G_{V_s} and D_{V_s} mark ground truth and first detected atypical frames of some video V_s , respectively. $[T_E]_{V_s}$ denotes T_E of V_s .

Various measures	Video streams ($V_s, s = 1..29$)																												
	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	V_{10}	V_{11}	V_{12}	V_{13}	V_{14}	V_{15}	V_{16}	V_{17}	V_{18}	V_{19}	V_{20}	V_{21}	V_{22}	V_{23}	V_{24}	V_{25}	V_{26}	V_{27}	V_{28}	V_{29}
D_{V_s}	309	155	24	93	16	(a)	37	92	(b)	169	95	183	256	(c)	338	29	(d)	147	139	55	123	(e)	80	63	212	49	207	(f)	119
$[T_E]_{V_s}$.38	.36	.34	.41	.36	.43	.39	.40	.42	.35	.37	.44	.37	.39	.42	.38	.33	.37	.43	.35	.33	.44	.45	.42	.38	.39	.41	.45	.37
G_{V_s}	312	158	28	99	13	54	31	97	62	175	99	179	261	923	331	33	553	151	144	52	128	71	86	67	217	56	211	411	125
Δ_{V_s}	9	9	16	36	9	-	36	25	-	36	16	16	25	-	49	36	-	16	25	9	25	-	36	16	25	49	16	-	36
Mean errors	$\text{Root Mean Squared Error } (\Psi) = \sqrt{\frac{1}{29} \sum_{s=1}^{29} \Delta_{V_s}}$, where $\Delta_{V_s} = (G_{V_s} - D_{V_s})^2 \Rightarrow \Psi = \sqrt{\frac{571}{29}} \approx 5$ $\text{Mean Absolute Error } (\Phi) = \frac{1}{29} \sum_{s=1}^{29} G_{V_s} - D_{V_s} \Rightarrow \Phi = \frac{11}{29} \approx 5$																												

equidistant points. Accordingly, we can apply some lower degree (e.g., 5) of polynomial fitting on the obtained *Entropy* measures data. As a consequence a more reliable, stable, workable, palatable, and much less erroneous and sensitive measures over the originally obtained *Entropy* measures data for a decision of normal or abnormal frame can be gained.

2.9. Threshold estimation

The decision of normal or abnormal frame can be taken either *static way* by comparing with polynomial fitting data with a predefined threshold T_E or *dynamic way* by detecting considerable sudden changes of the polynomial fitting data over time. In static way, a predefined threshold T_E , calculated from video which contains exclusively normal activities, can differentiate each frame with respect to its estimated *Entropy* whether it is normal or exceptional. An abnormal frame can be detected *if & only if* $Entropy > T_E$, otherwise normal frame. The T_E (also *Entropy*) depends on the controlled environment (video stream V_s), specifically the remoteness of the camera to the scene, the orientation of the camera, the type and the position of the camera, lighting system, density of the crowd, etc. In general, the more is the remoteness between the camera and the scene, the less is the considerable amount of optical flows and blobs. In case of escalator, T_E also places trust on the escalator type and position. Looking on these facts, we have at least one threshold for a video stream. If we have \mathcal{N} video streams, which are the case in sites e.g., airports, banks, hospitals, hotels, concerts, cinema halls, parking places, political events, shopping malls, subways, stations, schools, sporting events, town centers, etc., then we put forward at least \mathcal{N} thresholds. If the video stream V_{s-1} (where $s-1 \in \mathcal{N}$) leaves for another V_s (where $s \in \mathcal{N}$), then the threshold T_E of V_s will be made over through the use of

$$[T_E]_{V_s} = \arg \max_{f=1..m} [Entropy]_f + \arg \min_{f=1..m} \left[\frac{1}{(2\pi)^2} \sum_{k=0}^{\infty} \frac{(-1)^k (Entropy)^{2k+1}}{k!(2k+1)} \right]_f \tag{14}$$

where m is the number of frames in the video V_s and second term indicates some minimum *Gaussian error*, which is added for a good estimation of the threshold.

3. Experimental results and discussion

To conduct experiments, we have principally made use of the *Escalator dataset* [20] and the two datasets as operated by [19] so-called, respectively, the *UMN dataset* [21] and the *Web Dataset* [19]. Routinely, ζ limits $0.65 \leq \zeta \leq 1$ and $n=2000$. Adjacent to camera region, $\zeta = 0.65$ suits well while ζ bears 1 at opposite end. User friendly parameters have been selected as $\lambda=0.5$ and $\theta = 1.20$.

3.1. The escalator dataset

Escalators have become an imperative part of urban life. Escalator related injuries occur infrequently but may result in significant trauma. In 2000, the accident rate for escalator riding was about 0.815 accidents per million passenger trips through Taipei Metro Rapid Transit heavy capacity stations [8]. There are approximately 7300 escalator-related injuries in the United States each year [9].

However, our used *Escalator dataset* [20] consists of 29 real videos of total duration 15 min ($\approx 285\ 608$ frames), taken in spanning days and seasons, of frame size 640×480 pixels, collected by cameras installed in an airport to monitor especially



Fig. 15. The algorithm has hardly effect on handling occlusion anomalies e.g., (a)–(f).

the escalator exits, provided by a video surveillance company under the MIAUCE project [20]. The videos were used to provide informative data for the security team. Each video stream consists of normal and abnormal events. The normal situations correspond to crowd flows without any eccentric event on the escalator elsewhere. Eccentric events correspond to videos which contain collapsing events mostly in the escalator egresses. Generally, in the videos we have two escalators corresponding to two-way-traffic of opposite directions.

Images in Fig. 14 are the output of the abnormal event detector, depict two crowd scenarios of collapsing events on the escalator exits. First row (V_1 listed in Table 4) depicts a scenario where two persons were standing on the moving escalator and suddenly a trolley became unbalanced and rushed out toward them. One person got away by running and was not run down under the force of trolley, while other was ill-fated. As a result the non-escapee was run down by the runaway trolley, and subsequently fell down at the exit point of the moving escalator. Second row (V_2 listed in Table 4) describes another inconsistent circumstances on the exit point where a wheel from the trolley has suddenly been broken off by the friction during its travel over the escalator. Most of the inconsistent situations were detected by the proposed approach. The detailed evaluation of the proposed algorithm considering static method of thresholding for the provided escalator dataset has been listed in Table 4.

The algorithm has scarcely effect on detecting abnormalities from the video streams 6th, 9th, 14th, 17th, 22nd, and 28th listed in Table 4 as shown their sample frames in Fig. 15(a), (b), (c), (d), (e), and (f), respectively. This is due to the fact that the video sequences include abnormal events occur with occlusion. Accordingly, the estimated *Entropy* obtained from the quantity of extracted information is insufficient to draw out anomalous frames. Of course, it is well known that occlusion handling is an arduous part of optical flow technique. In Table 4, except six videos, the first detected abnormal frame D_{V_s} of some video V_s has been compared with the respective ground truth G_{V_s} and thereof *root mean squared error* Ψ and *mean absolute error* Φ have been

estimated for 23 out of 29 videos. The estimation of $\Psi = 0$ and $\Phi = 0$ corresponds to perfect detection or ideal case or ground truth. In spite of the fact that, the estimated $\Psi \approx 5$ and $\Phi \approx 5$ fall within the fitting range of many computer vision applications along with escalators.

3.2. The UMN dataset

The publicly available dataset of normal and abnormal crowd videos from University of Minnesota [21] comprises the videos of 11 different scenarios of an escape event in three different indoor and outdoor scenes. Duration of the videos is 4.3 min (≈ 7724 frames) and the frame size is 320×240 pixels. Each video consists of an initial part of normal behavior and ends with sequences of the abnormal behavior.

The qualitative results of the abnormal behavior detection for four sample videos (we named d_1, d_2, d_3, d_4 from top to bottom) of UMN dataset, as demonstrated by [19], have been presented in Fig. 16. In all the sample videos, abnormal motion includes a sudden situation when the group of people start running the measured *Entropy* will be higher than that of any other before estimated. Gaussian like curves present the abnormal motions when those groups of people are trying to leave their places with atypical motions. Results report that the proposed method performs something to a greater degree to distinguish abnormal sequences. The results are likely a bit superior to [12,19] in the sense that there is no significant reported false positives. Table 5 provides the quantitative results of a comparison with Ihaddadene et al.'s [12] and Mehran et al.'s [19] results for the same four videos. Most false alarms in [12] have been resulted from the incorrect estimation of influencing features e.g., direction histogram peaks, linear direction variance, etc. Like the approach of Mehran et al.'s [19], our approach does not require an explicit learning period to estimate various parameters of the system. Consequently, there is no serious concern of the reliable learning of unknown parameters which leads to potentially decrease the number of false alarms.



Fig. 16. Qualitative results of abnormal behaviors detection using the proposed framework for the same four sample videos as demonstrated in [19] from the UMN dataset.

Table 5
Comparison of Ihaddadene et al.'s [12] and Mehran et al.'s [19] results.

Approaches	d_1	d_2	d_3	d_4	Ψ	Φ	False alarms
Ihaddadene et al. [12]	478	595	734	689	12	12	9
Mehran et al. [19]	482	593	741	696	16	16	6
Proposed	461	576	718	671	6	6	0
Ground truth case	466	581	724	678	0	0	0

3.3. The Web dataset

We have likewise conducted the experiments on the challenging set of videos that has been used by [19] and placed together from the sites e.g., Getty Images, ThoughtEquity (<http://www.thoughtequity.com>), Google Videos, etc. which contain documentary and high quality videos of crowds essentially in different urban scenes. This dataset consists of video streams with duration 17 min ($\approx 30\,536$ frames) and the frame size is 320×240 pixels. Video sequences of normal crowd scenes are pedestrian walking, marathon running, and so forth as well as abnormal crowd scenes include escape panics, protesters clashing, crowd fighting on the street, etc. Fig. 17 shows five sample videos (we named w_1, w_2, w_3, w_4, w_5 from top to bottom) of the Web dataset [19] where w_1 pedestrian walking, w_2 marathon running, w_3 crowd fighting on the street, and w_4 and w_5 escape panics.

Beyond the crowd aberrant activities detection, the algorithm can monitor illegal traffic activities on the high-way, e.g., car making illegal U-turn. Fig. 18 depicts an illegal U-turn situation which has been picked up by the algorithm.

3.4. Performance evaluation

We have conducted experiments of our proposed approach essentially on three datasets and its effectiveness on those datasets has been listed in Table 6. We have considered two statistical probability measures namely *recall rate* (also called sensitivity) and *precision rate* (also called positive predictive value). Precision is the probability that can be seen as a measure of exactness or fidelity, whereas recall is also the probability that is a measure of completeness. A general desired attribute for a smart detector is that the recall and precision rates from a given video stream are expected to be very high by minimizing the number of false negatives and false alarms, respectively. An average sensitivity of 100% means that the detector recognizes all actual abnormal activities from the video streams and such kind of detector is called perfect or ideal detector, which is yet unknown to the computer vision research community.

The proposed approach has a figurative support on optical flow concepts and therefore overcomes problems with detection and tracking of individual objects. Yet, abnormal events which occur with occlusion have circumscribed the average performance of



Fig. 17. Qualitative results of normal and abnormal behaviors detection using the proposed approach for five sample videos from Web dataset as operated by [19]. First and second rows concern normal activities of pedestrian walking and marathon running, respectively. The leftover rows bear reference to the aberrations: third row demonstrates crowd fighting on the street, the last two rows touch upon escape panics (e.g., overwhelming feeling of fear and anxiety).

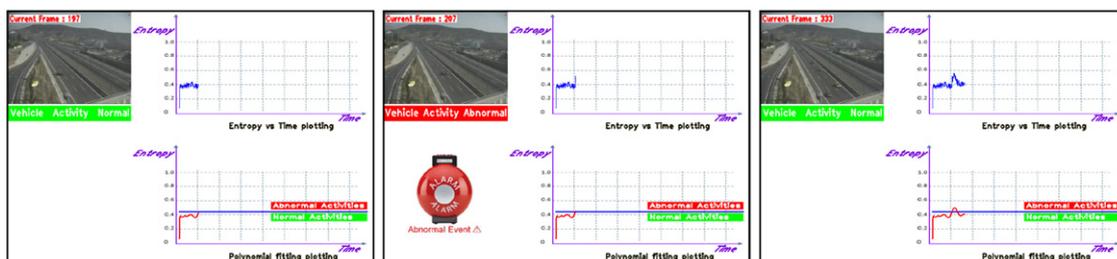


Fig. 18. Example video in which cars are ensuring the regular traffic flow which hints that Entropies are normal; while a car making an illegal U-turn which infers that Entropies are higher, and consequently the illegal traffic activity has been picked up by the pointed framework.

Table 6
The effectiveness of the proposed approach on different datasets.

Datasets	Number of frames	Recall rate (%)	Precision rate (%)
Escalator [20]	285 608	80	86
UMN [21]	7724	94	97
Web (e.g., [19])	30 536	91	89
Average	107 957	88.3	90.6

Table 7
General characteristics. Symbols ✓ and × denote yes and no, respectively.

Different issues	Different approaches			
	Andrade et al. [2,3]	lhaddadene et al. [12]	Mehran et al. [19]	Our approach
Detected anomaly?	✓	✓	✓	✓
Region of interest used?	×	✓	×	✓
Low rate of false alarm?	×	×	×	✓
Tested with real dataset?	×	✓	✓	✓
Tested with simulated data?	✓	×	×	✓

the detector to a certain degree to detect abnormal activities from video streams. Unambiguously, handling of occlusion is a laborious work in optical flow as occluded pixels violate a major assumption of optical flow that each pixel goes somewhere. In theory, the pixels at the occlusion area should not be assigned any flow vector since there is no correspondence available in the other frame. As the proposed approach touches base with optical flow techniques, no flow vector can come to have possession on occlusion areas. Subsequently, the detector has scarcely effect on overlapping standpoints. In spite of that, the mainstream performance of the detector demonstrates an affording satisfaction in many applications of computer vision.

3.5. Comparison with state-of-the-art

Our proposed approach has several important differences from the most related body of works, e.g., Andrade et al. [2,3], lhaddadene et al. [12], and Mehran et al. [19]. A brief overview of some important issues of these research works along with our approach have been listed in Table 7.

Like our approaches, in the work of Andrade et al. [2,3] crowd behavior has been characterized at a global level by using the optical flow of the video sequence. Unlike our approaches, during the learning stage, a reduced order representation of the optical flow was generated by performing PCA on the flow vectors. Afterwards, top few eigenvectors were used as the representative features and spectral clustering was performed to identify the number of distinct motion patterns present in the video. The features in the clustered motion segments were used to train different HMMs which were then used for event detection in crowds. The method was only tested by data obtained from simulation. A general limitation of simulation is that models are typically unstructured and must be developed for problems that are also unstructured. It is often impractical to realistically validate simulation results all of the above. Furthermore, the model building for simulation is customarily costly and time-consuming.

In video surveillance scenes, camera positions, and lighting conditions allow getting a large number of Harris corners that can be easily captured and tracked. Since the most influencing

features (e.g., direction histogram peaks, ordinary linear direction variance, etc.), as proposed by lhaddadene et al. [12], are extracted based on the result of Harris corner detection, these features would be sensitive to textures. For example, if a person wears a grid-dress like cloth, there will be too many corners detected from the region of him/her so that most motion directions (e.g., 50% or more) in that frame are the same as the movement direction of the person. Consequently, features like direction histogram will be distorted in such situation. Moreover, the ordinary linear direction variance is not always accurate. As a result, the false positive as well as false negative will be increased significantly. As our proposed approach considers circular variance C_v (as discussed in Section 2.3.1), it is too accurate and reliable to report any angular change as compared to [12], where along with other measures the ordinary linear direction and direction histogram have been taken into account. Fact is that in case of any angular change circular variance is more skillful than that of linear statistic. For this reason, our approach can detect the U-turn like Fig. 18 successfully, whereas the method of [12] has narrow effect on the detection of such nearly undetectable change on the video. In this regard, our proposed method is again superior to [12].

Many existing methods of abnormal event detection require a learning period to estimate various parameters of the system, and hence reliable learning of unknown parameters is not always accurately possible which could potentially increase the rate of false alarms. For instance, Mehran et al. [19] have introduced an approach to detect and localize abnormal behaviors in crowd videos using social force model. They have presented that their estimated social force model is capable of detecting the governing dynamics of the abnormal behavior, even in the scenes that it is not trained. Nevertheless, significant number of false alarms in their framework are the result of incorrect estimation of social forces. This is a severe shortcoming of their approach. In a contrary manner, the implicit learning period in our approach is threshold, consequently, the false alarm rates are significantly low as compared to their approach.

3.6. Future work directions

Future work would carry on the method to bring about circumstances of overlapping abnormal activities along with some smarter threshold estimation process. Occlusion handling is still one of the major challenges in computer vision. Advances in sensing technologies as well as the increasing availability of computational power and efficient bandwidth usage methods are favoring the emergence of applications based on systems combining multiple cameras and other sensing modalities. Multiple cameras can provide different viewpoints of a region of interest. Since all experiments have been conducted on videos of single fixed camera, it would be interesting to test the approaches with moving single camera datasets or multi-camera datasets. Future work would take into account the dedication of multiple cameras so that videos, like escalators, could be conclusively broken down into its essential features properly in all parts (e.g., commencement, halfway point, and outlet) of an elongated escalator to proclaim the eccentric event if there will exist any. Consequently, the engagement of multiple cameras would help to analyze many region of interests which would be occluded by a single camera.

4. Conclusion

We keyed out a simple but effective approach to detect abnormal activities in video streams upon estimating *Entropy* over time on frames without the motive to track subjects individually or carry out segmentation. The framework is robust

against variable number of subjects in the scenes. We clarified that both *degree of randomness of the directions* (circular variance) and *degree of randomness of the displacements* of interest points are necessary and sufficient crude elements of the defined entropy measure to detect a wide variety of abnormal activities from video streams. Simple simulations were exercised to understand their potential characteristics in both normal and abnormal circumstances. Normalized entropy measure renders the knowledge of the state of anomalousness. Experiments were conducted on more than a few real world video datasets. Both simulation and experimental results reported that entropy measures of the frames over time is a first-rate technique to characterize aberrations in video streams. Future study include improving the abnormal activity detection by incorporating overlapping occurrences of multi-camera as well as some smarter threshold estimation process.

Acknowledgments

Thanks to the MIAUCE project, which is the 6th framework research programme of European Union (IST-2005-5-033715), for supplying escalator dataset.

Thanks also to the reviewers for their appreciative and constructive comments on the draft of this paper.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version of [10.1016/j.patcog.2011.11.023](http://dx.doi.org/10.1016/j.patcog.2011.11.023).

References

- [1] S. Ali, M. Shah, A Lagrangian particle dynamics approach for crowd flow segmentation and stability analysis, in: *Computer Vision and Pattern Recognition (CVPR)*, 2007, pp. 1–6.
- [2] E.L. Andrade, S. Blunsden, R.B. Fisher, Hidden Markov models for optical flow analysis in crowds, in: *International Conference on Pattern Recognition (ICPR)*, 2006, pp. 460–463.
- [3] E.L. Andrade, S. Blunsden, R.B. Fisher, Modelling crowd scenes for event detection, in: *International Conference on Pattern Recognition (ICPR)*, 2006, pp. 175–178.
- [4] A.F. Bobick, J.W. Davis, The recognition of human movement using temporal templates, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23 (3) (2001) 257–267.
- [5] O. Boiman, M. Irani, Detecting irregularities in images and in video, *International Journal of Computer Vision* 74 (1) (2007) 17–31.
- [6] J.Y. Bouguet, Pyramidal implementation of the Lucas Kanade feature tracker, in: *A part of OpenCV Documentation*, Intel Corporation, Microprocessor Research Labs, 2000.
- [7] P.L. Chebyshev, Théorie des mécanismes connus sous le nom de parallélogrammes, *Mémoires des Savants étrangers présentés à l'Académie de Saint-Petersbourg* 7 (1854) 539–586.
- [8] C.F. Chi, T.C. Chang, C.L. Tsou, In-depth investigation of escalator riding accidents in heavy capacity MRT stations, *Accident Analysis & Prevention* 38 (4) (2006) 662–670.
- [9] U. C. P. S. Commission, CPSC Document #5111: Escalator safety, US Consumer Product Safety Commission, Washington, DC, 2003.
- [10] Entropy, Online Etymology Dictionary, <<http://www.etymonline.com>>, 2010.
- [11] W. Hu, X. Xiao, Z. Fu, D. Xie, T. Tan, S.J. Maybank, A system for learning statistical motion patterns, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28 (9) (2006) 1450–1464.
- [12] N. Ihaddadene, C. Djeraba, Real-time crowd motion analysis, in: *International Conference on Pattern Recognition (ICPR)*, 2008, pp. 1–4.
- [13] I. Ivanov, F. Dufaux, T.M. Ha, T. Ebrahimi, Towards generic detection of unusual events in video surveillance, in: *International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2009, pp. 61–66.
- [14] E.T. Jaynes, Information theory and statistical mechanics, *Physical Review* 106 (4) (1957) 620–630.
- [15] I.N. Junejo, O. Javed, M. Shah, Multi feature path modeling for video surveillance, in: *International Conference on Pattern Recognition (ICPR)*, vol. 2, 2004, pp. 716–719.
- [16] Y. Kim, A.C. Kak, Error analysis of robust optical flow estimation by least median of squares methods for the varying illumination model, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28 (9) (2006) 1418–1435.
- [17] B. Lucas, T. Kanade, An iterative image registration technique with an application to stereo vision, in: *International Joint Conference on Artificial Intelligence (IJCAI)*, 1981, pp. 674–679.
- [18] D. Makris, T.J. Ellis, Path detection in video surveillance, *Image and Vision Computing Journal* 20 (2002) 895–903.
- [19] R. Mehran, A. Oyama, M. Shah, Abnormal crowd behavior detection using social force model, in: *Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 935–942.
- [20] P. MIAUCE, This publicly unavailable dataset belongs to the MIAUCE project, EU Research Programme (IST-2005-5-033715) <<http://www.miauce.org>>, 2008.
- [21] University of Minnesota, Unusual crowd activity dataset of University of Minnesota. Available from <<http://mha.cs.umn.edu/movies/crowdactivity-all.avi>>, 2009.
- [22] C.D.T. Runge, Über empirische funktionen und die interpolation zwischen äqui-distanten ordinaten, *Zeitschrift für Mathematik und Physik* 46 (1901) 224–243.
- [23] C.E. Shannon, Prediction and entropy of printed English, *The Bell System Technical Journal* 30 (1951) 50–64.
- [24] J. Shi, C. Tomasi, Good features to track, in: *Computer Vision and Pattern Recognition (CVPR)*, 1994, pp. 593–600.
- [25] C. Stauffer, W.L. Grimson, Learning patterns of activity using real-time tracking, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (8) (2000) 747–757.
- [26] W. Weibull, A statistical distribution function of wide applicability, *Transactions of the ASME: Journal of Applied Mechanics* 18 (3) (1951) 293–297.
- [27] T. Xiang, S. Gong, Video behavior profiling and abnormality detection without manual labeling, in: *International Conference on Computer Vision (ICCV)*, 2005, pp. 1238–1245.
- [28] T. Xiang, S. Gong, Video behavior profiling for anomaly detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30 (5) (2008) 893–908.
- [29] X. Xue, et al., Fudan University at Trecvid 2008, in: *Surveillance Event Detection Pilot* <<http://www-nlpir.nist.gov/projects/trecvid>>, 2008.