



Multi-scale crowd feature detection using vision sensing and statistical mechanics principles

Banafshe Arbab-Zavar¹ · Zoheir A. Sabeur²

Received: 24 September 2019 / Revised: 20 February 2020 / Accepted: 23 March 2020
© The Author(s) 2020

Abstract

Crowd behaviour analysis using vision has been subject to many different approaches. Multi-purpose crowd descriptors are one of the more recent approaches. These descriptors provide an opportunity to compare and categorize various types of crowds as well as classify their respective behaviours. Nevertheless, the automated calculation of descriptors which are expressed as measurements with accurate interpretation is a challenging problem. In this paper, analogies between human crowds and molecular thermodynamics systems are drawn for the measurement of crowd behaviour. Specifically, a novel descriptor is defined and measured for crowd behaviour at multiple scales. This descriptor uses the concept of *Entropy* for evaluating the state of crowd disorder. By results, the descriptor *Entropy* does indeed appear to capture the desired outcome for crowd entropy while utilizing easily detectable image features. Our new approach for machine understanding of crowd behaviour is promising, while it offers new complementary capabilities to the existing crowd descriptors, for example, as will be demonstrated, in the case of spectator crowds. The scope and performance of this descriptor are further discussed in detail in this paper.

Keywords Crowd dynamics · Crowd behaviour detection · Multi-scale crowd features · Group detection and tracking · Video analysis · Statistical mechanics

1 Introduction

Various physical analogies and modelling approaches have been used in dealing with crowd motion modelling. Some of the more popular modelling analogies in this domain include cellular automata, social force model and fluid mechanics. These methods each have their own functionalities and limitations. Cellular automata (CA) have been used to simulate crowd dynamics in situations such as evacuation [2,7,13,15] or to simulate certain effects such as line formation in the crowd [30]. CA does not aim to capture all the microscopic dynamics, but only the one which is necessary to produce a certain macro effect. The social force model is another pop-

ular method for crowd simulation [9,10]. The social force model is based on a simple concept wherein individuals move according to their goals and environmental constraints. It is assumed that each individual has a desired direction and velocity, while seeking to keep a social distance from other members of the crowd as well as avoid hitting boundaries. The accuracy of the social force model is directly dependent on the accuracy of the estimated desired velocities, which is in itself a challenging problem. Fluid mechanics has also been investigated for modelling pedestrian motions. Henderson was the first to propose a gas kinetic model for pedestrian flows [11]. Using this basis of a Boltzmann-like gas kinetic model, Helbing [8] developed a special theory for pedestrians, distinguishing between different groups within the crowd with different types of motions and goals. However, these works do not include any experimentation with real crowds using vision or any other in situ observation or measurement.

In addition to the above-mentioned modelling approaches, where models impose hypothetical structures that are controlled by sets of parameters on crowd motion, other approaches for understanding crowd behaviour are achieved

✉ Banafshe Arbab-Zavar
baz1f09@soton.ac.uk

Zoheir A. Sabeur
zsabeur@bournemouth.ac.uk

¹ Department of Electronics and Computer Science, University of Southampton, Southampton, UK

² Department of Computing and Informatics, Bournemouth University, Bournemouth, UK

through observation and learning from reoccurring crowd patterns of motion. Topic models have been used successfully for learning these patterns in an unsupervised fashion. In this, they have been used to learn the semantic (spatial) regions within a crowd scene [29,38]. Other methods for detecting and segmenting semantic regions have also been proposed [5,17,21,27,28]. Other approaches for analysing crowd behaviour include agent-based methods [32,36], where the behaviour of pedestrians as individuals is considered and modelled in relation with the rest of the crowd; appearance-based approaches [22] using crowd behaviour priors in the form of image patches which are learned offline; and methods which look at groups and group activities within the crowd [4,20,33].

One of the first works which proposed a descriptor for crowd and demonstrated its usefulness in analysing the behaviour of the crowd is by Zhou et al. [35,37]. Crowd *Collectiveness* was introduced as a measure of the degree of individuals acting as a union in a collective motion. Collectiveness is based on two key properties of collective motion: (i) behaviour consistency in neighbourhoods; (ii) global consistency among non-neighbours. Using collectiveness as a method for detecting groups within the crowd was also proposed [37]. The detection of groups within a crowd was further studied using the concepts of *coherent motion* [31]. A number of group-level crowd descriptors were then introduced by Shao et al. [25]. These descriptors include *Stability*, *Uniformity* and *Conflict*. Our approach bears similarities to these works, in that crowd descriptors are sought to assist understanding of the behaviour of the crowd.

Following the methods that consider a crowd of people to resemble a physical system, we propose *Entropy* as an additional descriptor for crowd analysis. Entropy has similarities to both collectiveness and stability. However, it is distinctly different in terms of its definition and computation. In practice, as will be demonstrated, these descriptors are suitable in different circumstances. A detailed comparison is made between entropy and collectiveness, while stability is also compared to these both. We will further introduce the concept of *Internal energy* for crowds and offer initial discussions as to its validity as a crowd descriptor.

As for the significance of defining and estimating crowd descriptors, Zhou et al. [37] note that the “lack of universal descriptors” to characterize crowd behaviour(s) is the main reason behind the inadequacy of most surveillance technologies for automatically detecting crowd behaviour(s) across different scenes. Crowd descriptors, especially when they are used as a set of features, provide the generality of approach which is needed to handle different types of crowds and different types of crowd behaviours. This is in contrast to the modelling approaches which have been developed to investigate specific crowds and specific behaviours. Different descriptors are therefore required to express the

various aspects of crowd behaviour. In our study, we propose a novel and complementary descriptor for meso-scale crowd description. This is inspired from the fundamental principles of statistical mechanics, where the macroscopic properties of gases are derived from the statistical motion realisation of their constituent microscopic molecules (microscopic particles). The similarity can be found with crowd, where micro-scale (individual) motions within the crowd can influence the overall behaviour(s) of the crowd at the macro-scale levels. In this, individual motions are observed and utilized to measure crowd descriptors at macro-level. This provides a method to quantify entropy as a computer vision descriptor for crowds. Furthermore, it opens up an opportunity to explore the knowledge of statistical mechanics for the benefit of crowd behaviour analysis from vision.

The use of an ensemble of particles for modelling people was introduced in the initial theoretical studies [8,10,11] and was further utilized for automated visual analysis of crowd behaviour. Some of the works mentioned in this section use a similar framework [1,5,18]. A survey work by Moore et al. [19] refers to this as particle-based framework and reviews the benefits and scope of such an approach.

Our contributions include (i) the introduction of entropy as a complimentary descriptor for crowd analysis; and (ii) a new approach for unusual behaviour detection in crowds via the *crowd space*. A similar method can also be used for defining multiple crowd states and thereby detecting a change in the crowd state.

In the next section, a new crowd feature space is introduced. In this, three features of *Structure*, *Energy* and *Translation* are intuitively identified to facilitate the understanding of the state of a crowd and its behaviour. Also, these features can be used directly to evaluate the usualness or the unusualness of this behaviour. Section 3 provides a detailed description of structure and its usage in context of the study of crowd. It is shown here how this descriptor can be mapped onto the statistical mechanics principles of entropy. The scope and comparisons to the other descriptors are also covered. Discussions on sub-groups and homogeneity in sub-groups, as well as discussions on internal energy as a crowd descriptor, can be found in Sect. 4. Section 5 looks into unusual behaviour detection of crowds in real settings and within a context. Finally, conclusions are summarized in Sect. 6.

2 The crowd features

In our approach, we assume that a force keeps crowd members together. The strength of connections between the members will be referred to as *Structure*. Irrespective of the strength of connections, the crowd may be in an excited state (high energy) or a calm state (low energy). This feature is

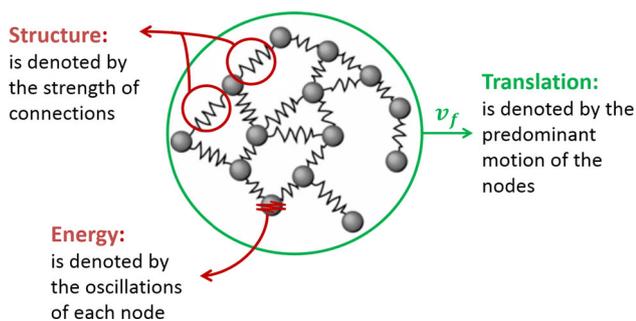


Fig. 1 Crowd features

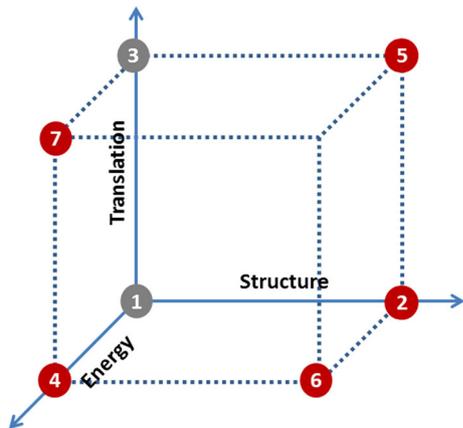


Fig. 2 Crowd space with hypothetical examples

defined as *Energy*. It is also possible to consider that the whole crowd moves in space. This is referred to as *Translation*. As will be discussed in greater detail later, the above features are inspired and translated into statistical mechanics concepts of entropy, internal energy and flow. Figure 1 depicts a visual representation of these features.

Following the introduction of these features, a three-dimensional crowd space can be defined. Figure 2 shows a representation of the structure–energy–translation crowd space. In this, the cube represents a space of normalised parameters representing the state of the crowd system. Table 1 offers a set of hypothetical examples of different types of crowds, while Fig. 2 shows where these would reside in the crowd space.

Figure 2 shows a number of examples where the different crowd types can be differentiated using the crowd feature space. Changes in the state of a crowd may also be detected using this feature space.¹ Further, unusual behaviours can be detected using this space by defining a sub-space wherein the crowd is expected to reside. Figure 3 illustrates how the state of crowds can be correctly monitored using these features.

¹ Given that these changes are within the scope of the features. The scope is addressed in the following sections.

Table 1 Hypothetical examples of crowds

1&3	Structure and energy both at zero is not an intuitive state. It is similar to having a gas at absolute zero
2	Very unmotivated spectator crowd at a very boring game
4	A group of panicking people locked in a room
5	People on an escalator
6	Crowd at a football match celebrating a goal
7	A sparse panicked crowd escaping in one general direction

As shown in Fig. 2, a crowd may reside in any location in the crowd space. However, for any given situation or context there would be an expectation of where the crowd *should* reside. A divergence from this expected or desired position can be considered as an unusual crowd behaviour. Figure 3 shows the envisaged sub-spaces of usual (expected) crowd behaviour spaces in various situations and crowds. By mapping the crowd onto the crowd space and learning the limits of usual behaviour, a crowd with unusual behaviour can be defined as a crowd which does not fall within such limits.

3 Entropy of crowd

The concept of entropy is based on the generic observation that there are many more ways for a system of microscopic particles to be disordered than to have a certain specific order. While manifesting a specific macroscopic state, it is more probable for such system, or statistical ensemble to assume a level of disorder. If a certain order is observed, given that it would have been unlikely that such order is obtained at random, it can be concluded with certain confidence that there were other forces at play which enforced such arrangement.

In this section, one assumes that the crowd is a homogeneous system or otherwise the concepts are considered for homogeneous groups within the crowd. Detection of homogeneous groups is achieved through the use of *Collective merging* [37]. These meso-scale groups are tracked in consecutive frames. This is further discussed in Sect. 4.

In classical statistical mechanics theory, entropy, S , is the measure of mechanical disorder for a system of microscopic particles. It is defined in the following way:

$$S = -K \sum_i p_i \ln p_i \tag{1}$$

where, for a system with a discrete set of microstates, p_i is the probability of occurrence for microstate i and K is the

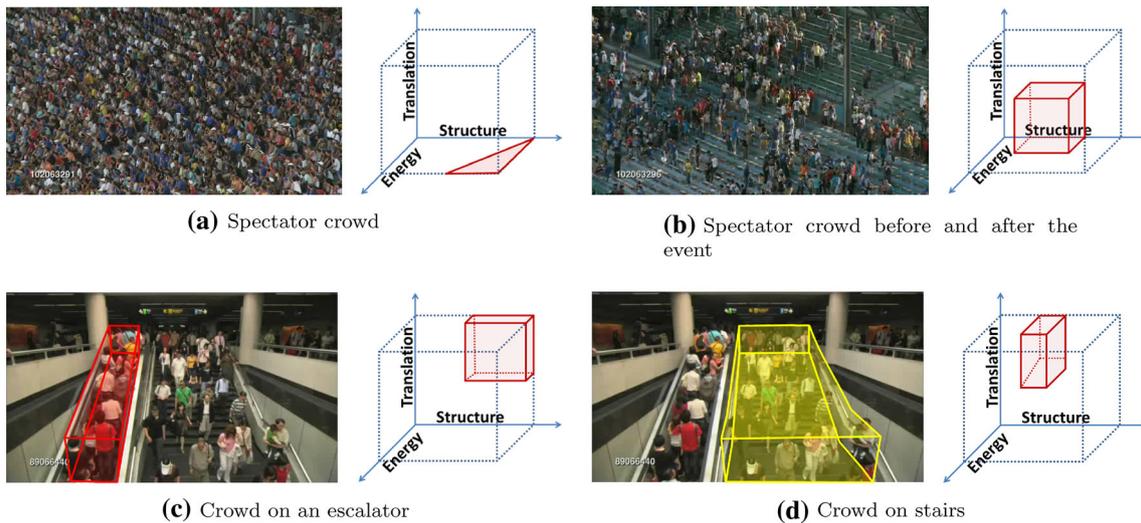


Fig. 3 Usual behaviour sub-spaces are shown on the right of each crowd example. **a** Spectator crowd is denoted by variant levels of energy at high structure with no translation. **b** When arriving or departing, spectator crowd has lower structure but significant translation. **c** A crowd

on an escalator is a good example of a low-energy crowd. **d** A crowd on stairs has smaller structure in comparison with (c) since each individual is moving at its own pace and more energy since each individual moves its limbs

Boltzmann constant. Similarly, entropy (mostly denoted by H) is adopted as a measure of uncertainty in information theory:

$$H = - \sum_i p_i \log_b p_i. \tag{2}$$

For both the above-mentioned theories, entropy leads to understanding of the overall macroscopic state of a system of microscopic particles, by calculating the statistical realisation of their microscopic states. The initial definition of entropy in classical statistical mechanics, $S = k_B \ln W$, connects entropy directly to the number of microstates, W , which corresponds to the macroscopic state of the system.

Considering the states of matter which include solid, liquid and gas, entropy for these states can be understood intuitively. In a solid, molecules oscillate around a fixed point, while entropy remains relatively low. In a liquid system, molecules move relatively freely while keeping certain distances from one another. In such case, entropy is usually higher in value than that of a solid system. Finally in a gaseous system, the constituent molecules can freely move anywhere, which leads to the highest values of entropy. In other words, higher manifested values in entropy are observed when the uncertainty on the position of the constituent molecules of matter increases.

One of the challenges in evaluating the value of entropy is that for each crowd example only a limited subset of all possible microstates are observed. Therefore, it is not possible to count the number of microstates or directly calculate their probabilities. For this, an extra step is devised to infer a

model for all possible microstates using the set of observed microstates.

3.1 Calculation of entropy using a microstate model

We define the entropy of a crowd as the joint entropy of N_p individuals who are scattered in N_l locations with a probability mass function f_{Y_i} on a discrete random variable, Y_i , defined at each spatial bin, l_i .

The joint entropy of two ensembles X and Y is [16]

$$H(X, Y) = \sum_{x,y \in \mathcal{A}_X \mathcal{A}_Y} P(x, y) \log \frac{1}{P(x, y)} \tag{3}$$

where both X and Y are triples. X is a triple $(x, \mathcal{A}_X, \mathcal{P}_X)$ where x is the value of a random variable, which takes on one of a set of possible values, $\mathcal{A}_X = \{a_1, a_2, \dots, a_l\}$, having probabilities $\mathcal{P}_X = \{p_1, p_2, \dots, p_l\}$. Similarly, Y is a triple $(y, \mathcal{A}_Y, \mathcal{P}_Y)$.

Thereby, the entropy of a crowd can be described as

$$\begin{aligned} H(X_1, \dots, X_{N_p}) \\ = - \sum_{x_1 \in \mathcal{L}_X} \dots \sum_{x_{N_p} \in \mathcal{L}_X} P(x_1, \dots, x_{N_p}) \log P(x_1, \dots, x_{N_p}) \end{aligned} \tag{4}$$

where X_k is a triple $(x_k, \mathcal{L}_X, \mathcal{P}_{X_k})$. x_k takes on one of a set of possible values, $\mathcal{L}_X = \{l_1, l_2, \dots, l_{N_l}\}$, having probabilities $\mathcal{P}_{X_k} = \{p_{k,1}, p_{k,2}, \dots, p_{k,N_l}\}$, with $P(x_k = l_i) = p_{k,i}$. Two approaches are considered here to evaluate $H(X_1, \dots, X_{N_p})$.

3.1.1 Approach 1: Complete enumeration

First, the complete enumeration of all possible microstates is considered using the ones which have been observed to calculate f_{Y_i} s. The joint probabilities, $P(x_1, \dots, x_{N_p})$, in Eq. (4) are the other unknowns. While these probabilities can be calculated using the probability mass functions f_{Y_i} , assumptions regarding the dependency of the individuals need to be made. The computation cost is of the order $O(N_l^{N_p})$. This is the permutation of N_p individuals scattered in N_l locations. For each of these permutations, the joint probability $p(X_1, \dots, X_{N_p})$ needs to be calculated.

The validity of this approach may be contested since the probability mass functions f_{Y_i} are calculated using a limited sample set of observed microstates and it is prone to overfitting the model to the observed sample set. Thus, relaxing some of the conditions in this model may be favourable for a better coverage of the space of all possible microstates.

3.1.2 Approach 2: Preserving the density pattern

One of the assumptions in the above approach concerns the dependence between the positions of the individuals. In the example below, it will be shown that although there is reason to believe that these positions are dependent, sufficient information is not available to understand their dependencies in an unbiased manner.

In support of the dependency argument, consider that people tend to keep certain distances from each other, the so-called personal space. Also depending on the relationships between the individuals, they may tend to group together or avoid others. From a different point of view, consider that a certain macrostate has been observed in a crowd: a number of clusters of people are observed in different locations. There may be different causes for this effect. Hypothesis *A*: some physical locations are more desirable than others, and people cluster in them for that reason. Hypothesis *B*: there is some social relationship between members of the crowd, and they cluster together due to that relationship. In Hypothesis *B*, the act of clustering is important, while the cluster positions are random. Furthermore, Hypothesis *C* can be added to accommodate the combination of the other two hypotheses. However, sufficient information is not given in favour of either hypothesis *A*, *B* or *C* in the above example.

Therefore, we propose that when analysing crowd formation through a few correlated frames a simpler model which exhibits similar outcomes is adopted. We hypothesize that a pattern is formed in the crowd if each individual is bounded by the same pattern. In this model, apart from the locations of people, which are considered to be independent, the individuals are considered to be identical. As a result of this approach, the calculation of entropy simplifies.

Let $n_{i,j}$ be the number of times that individual j has been observed in bin l_i in N_f frames (N_f is the number of frames in a chosen time window). The probability of selecting this bin, l_i , by individual j is

$$P(x_j = l_i) = \frac{n_{i,j}}{N_f} \tag{5}$$

Given that the location of individuals is considered independent and no distinction applies between individuals, the probability of any individual selecting bin l_i is the same as any other individual. Thus, the probability of selecting bin l_i , $P(x = l_i)$, is estimated in the following way:

$$\begin{aligned} P(x = l_i) &= \frac{\sum_{k=1}^{N_p} P(x_k = l_i)}{N_p} \\ &= \frac{\sum_{k=1}^{N_p} \frac{n_{i,k}}{N_f}}{N_p} = \frac{\sum_{k=1}^{N_p} n_{i,k}}{N_f N_p} = \frac{n_i}{N_f N_p} \end{aligned} \tag{6}$$

where n_i is the sum of all density counts at bin l_i in N_f frames. Since the locations of individuals are independent, the joint entropy of the crowd, $H(X_1, \dots, X_{N_p})$, simplifies as

$$H(X_1, \dots, X_{N_p}) = \sum_{k=1}^{N_p} H(X_k) \tag{7}$$

Also, note that the locations of all the individuals are based on the same location probabilities, $P(x = l_i)$. Thus,

$$H(X_1) = H(X_2) = \dots = H(X_{N_p}), \tag{8}$$

$$H(X_1, \dots, X_{N_p}) = N_p H(X) \tag{9}$$

where X is a triple $(x, \mathcal{L}_X, \mathcal{P}_X)$, the outcome x is the value of a random variable which takes on one of a set of possible values, $\mathcal{L}_X = \{l_1, l_2, \dots, l_{N_l}\}$, having probabilities $\mathcal{P}_X = \{p_1, p_2, \dots, p_{N_l}\}$, with $P(x = l_i) = p_i$ as was defined in Eq. (6). The crowd entropy in Eq. (8) has a time complexity of $O(N_l)$. (This is the time required to calculate p_i s using a constant time window size.) In other words, the entropy can be computed in linear time.

3.2 Pre-processing

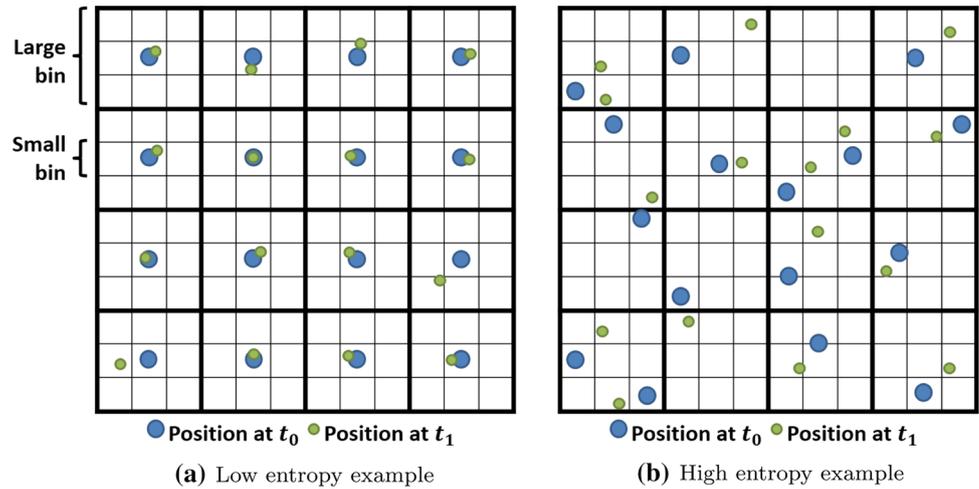
Three pre-processing stages should be considered before crowd entropy is calculated:

Real-world locations; The locations of individuals in an image have been subjected to projective transform. The severity of the distortion caused by this transform is relative to the angle between camera's image plane and the scene's ground plane. Ideally, this angle would be zero. This is specifically when the camera is placed overhead and looking down at



Fig. 4 The disruptive effect of projective transform intensifies from left to right. All images are from the data-driven crowd analysis data set [23]

Fig. 5 Low and high entropy with different bin sizes



the crowd. The location of individuals becomes increasingly skewed as this angle increases. Figure 4 shows three examples where the disruptive effects of projective transform are increasing from left to right.

Given the head locations of the individuals, the real-world positions can be retrieved using the camera calibration matrix and assuming an average height for the entire crowd. This is done through head-height plane homography transform [24]. However, the problem of head detection has proven difficult in the context of crowds. An alternative method using image features is discussed in Sect. 3.5. [19] also noted in their survey paper that side views “are least preferable for particle-based frameworks”. However, a *soft* calibration can be considered in the case of features as was also demonstrated by Zhou et al. [37].

Internal position density map; In order to calculate entropy, the internal position of each individual within the crowd, x_i , is required. If the crowd is stationary, then the observed position, x_o , is equal to the internal position ($x_i = x_o \iff v_f = 0$). However, if the crowd is moving with a flow velocity, v_f , the change in the internal position in a time step dt can be calculated as

$$dx_i = dx_o - v_f dt. \tag{10}$$

Internal position density map; Once the internal positions of individuals are known, an internal density map can be created. Note that the width of the density map bins, w_{bin} , is a significant parameter in the calculation of entropy. In this, a too large a bin will mask the very information that entropy is aiming to extract; with a too large a spatial bin, a gas and a solid may appear similar the way they uniformly occupy the space, while a too small a bin will be prone to noise. This is illustrated in Fig. 5. This figure shows two entropy levels with Fig. 5a low entropy and Fig. 5b high entropy. A time window of two consecutive frames is also depicted with the blue circles representing the position of particles at time t_0 and green circles for positions at time t_1 . The spatial gridding was done using two bin sizes: large bins and small bins. Please note that each spatial bin only counts the number of particles which lands on that bin, while the location of the particle within the bin is inconsequential. Conceptually, entropy for Fig. 5a when observed by the large bin is zero, since there is no difference between the two observed microstates and the particles appear stationary. The oscillations are better observed with the small bin where two of the particles are observed in new bins in t_1 . In Fig. 5b, depicting a large entropy, the large bin only observes three out of 16 particles to have moved between t_0 and t_1 , while this number is

15 out of 16 for the small bin. As a larger time window is considered, it is expected that the particles of example (b) would populate the available space, while the particles of example (a) are expected to oscillate around the original location. This effect cannot be observed by the large bin since it even observes the example (a) with a two-frame time window as uniformly populating the available space.

3.3 Normalisation of Entropy

Non-normalised entropy can only be used to compare crowds which are composed of the same number of individuals and have the same spatial extent. Since these conditions are rarely met, the normalisation of entropy becomes a necessary step to achieve.

Specific entropy; Specific entropy is the entropy per unit of mass. Assuming each individual has a unit of mass, the specific entropy, H_k , will be the entropy of one individual in this crowd:

$$H_k = H(X) \tag{11}$$

where X is a triple $(x, \mathcal{L}_X, \mathcal{P}_X)$, as in Eq. (8).

Specific entropy per unit of area; Entropy is maximized if \mathcal{P}_X is uniform [16]: $H(X) \leq \log |\mathcal{L}_X|$ with equality achieved iff $\forall i \in \{1, \dots, N_l\}, p_i = \frac{1}{|\mathcal{L}_X|} = \frac{1}{N_l}$.

It can be seen that the maximum value of entropy increases with the increase in the number of spatial bins, N_l . To account for this, we borrow a concept called *redundancy* from information theory. Redundancy is a measure of the amount of wasted space when coding and transmitting data. The redundancy of X , $R(X)$, on alphabet \mathcal{A}_X measures the fractional difference between $H(X)$ and its maximum possible value:

$$R(x) = 1 - \frac{H(X)}{\log |\mathcal{A}_X|} \tag{12}$$

Complementary to the concept of redundancy is *efficiency*, where redundancy and efficiency of a code add up to one. In this case, our notion of normalised specific entropy, h_k , is analogous to efficiency:

$$h_k = \frac{H_k}{\log N_l} \tag{13}$$

Minimum entropy; The minimum value for entropy is theoretically equal to zero. This is when only one microstate is possible for the system, and therefore the probability of that microstate to occur is one. We do not differentiate between individuals, and the probability of their presence at each location is calculated from the density map of the entire crowd. Thus, except if the entire crowd is concentrated at one spatial bin (which does not sound like a proper behaviour for a crowd if the bin size is set correctly), the minimum value of zero

is not obtainable. Instead, the obtainable minimum value of entropy is dependent on the initial density map, which in turn depends on the number of individuals, their sparseness and the bin sizes. It is desirable to assign a small entropy to a crowd that holds its structure, no matter how dense or sparse that structure may be. In this, the focus should be on the deviation of the crowd from its original arrangement. The minimum entropy is assumed to be that of the initial state (with window size zero). This normalises for density and sparsity of the crowd. A crowd for which the members hold their initial positions and just oscillate within the bounds of their respective positions the entropy is considered to be minimum within that time window. The entropy of this crowd is mapped onto zero entropy. In other words, only if the same structure is repeatedly replicated the entropy is considered to be zero. In practice, as the time windows get larger, uncertainty and noise build up and generally entropy grows with the increase in the size of the time window. Therefore, in real examples zero entropies do not occur. Similarly, the uniform coverage of the spatial bins will not be achieved in real examples and so is the entropy value of one. A word of caution: it is possible that in the initial state the particles are nearly uniformly distributed. In such cases, the difference between the minimum and maximum entropy is very small. This is generally a cue to incorrect bin size. An example of this was seen with the large bin in Fig. 5. The minimum entropy is thereby defined as

$$h_{min} = - \sum_{i=1}^{N_l} p_{0_i} \log p_{0_i} \tag{14}$$

where p_{0_i} is the probability of location l_i being occupied in the initial frame. Thereby, the normalised, scaled, specific entropy, \tilde{h}_k , is defined as

$$\tilde{h}_k = \frac{H_k - h_{min}}{\log N_l - h_{min}} \tag{15}$$

The normalised, scaled specific entropy, \tilde{h}_k , will be referred to as entropy hereafter.

3.4 Experimental results

Three crowd examples have been used in order to demonstrate the proposed method for conceptualising crowd as a statistical mechanical system. Experiment A (exp A) shows a crowd of people going down a staircase. The motion of the crowd in this example is unidirectional. Figure 6a shows one frame example of this crowd. It depicts an indoor scene with artificial lighting, while the crowd is viewed from an oblique frontal view. Figure 6b shows the second crowd example (exp B). This focuses on people on an escalator which is located on the left-hand side of the same video footage. Here, the

Fig. 6 Crowds with various levels of entropy

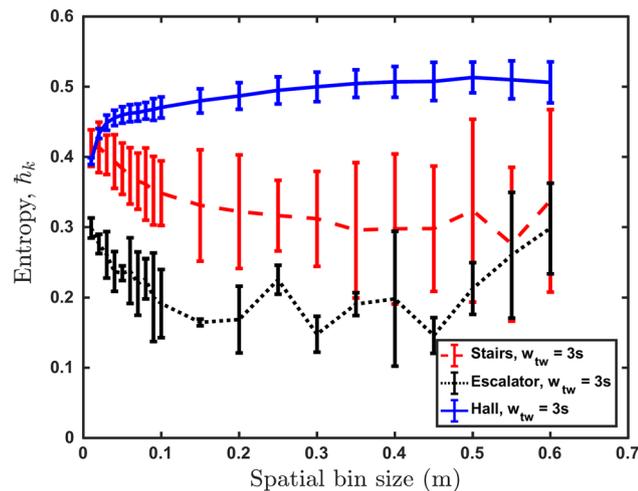
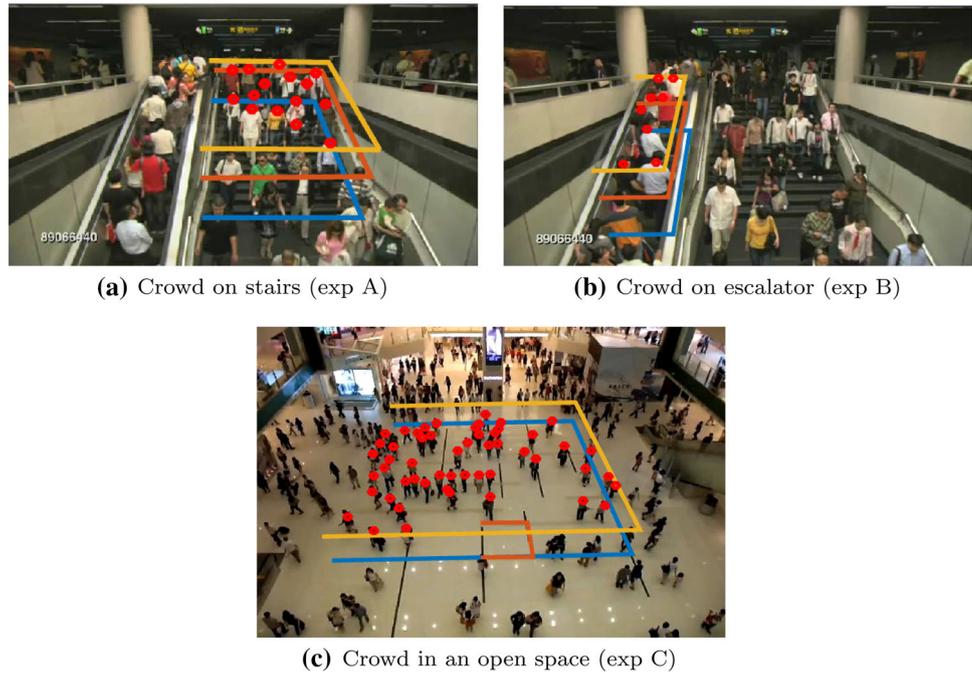


Fig. 7 Experiments with a 3 s time window ($w_{tw} = 3s$)

pedestrians are mostly standing still while the escalator carries them upwards. Finally, Fig. 6c shows a larger crowd in an open indoor space (shopping mall) with pedestrians moving in various directions (exp C). Both exp A and B reside in a crowd footage from the data-driven crowd analysis data set [23]. This video is captured at a resolution of 640×360 pixels and comprises 1155 frames at 25 frames per second (fps). Exp C uses a video footage from the Collective Motion Database [35] and has a resolution of 1000×670 pixels with 600 frames captured at 25 fps.

It is expected that: (i) the crowd in exp B, Fig. 6b, has the smallest entropy; (ii) the crowd in exp A, Fig. 6a, has a larger entropy than the crowd in exp B but still smaller than that of

the crowd in exp C, Fig. 6c. The largest entropy is envisaged for the crowd in exp C.

In these experiments, the respective figures show three calibration planes. In this, the orange plane is the reference plane which is manually drawn. The blue and yellow planes are the ground-level and head-level planes, respectively. These are projected back to the image plane after calibration. The red circles show the position of the individuals' heads on the head-level plane. Entropy was initially calculated using manually labelled heads. These were projected into the ground plane [24]. For this, a pre-processing step with a head detection algorithm was assumed to be present. Experiments were carried out for varying time window sizes (w_{tw}) and spatial bin widths (w_{bin}). The results confirmed the hypothesis with

$$\hat{h}_k(X_{expC}) > \hat{h}_k(X_{expA}) > \hat{h}_k(X_{expB}). \quad (16)$$

Figure 7 shows the results, where a time window size of 3 s is used. It can be seen that the order of entropy values is as expected and the separation within the error bars between the various experiment crowds is mostly achieved. This figure also demonstrates the effects of spatial bin size, where bins in the range of [0.01 m, 0.6 m] are investigated. It can be seen that the smallest bin sizes do not offer a good separation between the crowds. The same also goes for the larger bin sizes. The best separation is achieved for bin sizes within the range [0.04 m, 0.2 m]. As the bins get larger, the entropy becomes unstable for the escalator case. This can be attributed to the small volume of the escalator crowd as well as that too large bin sizes are not sensitive enough to the differences in individuals' motions. It is also observed



Fig. 8 n_i -maps for exp A

that larger time windows offer better separation. However, it must be noted that due to observing a non-stationary crowd with a stationary camera, it is possible that the crowd or the section of the crowd which is being analysed would move beyond the camera field of view. The results for exp B, when analysed with a 5 s time window, may be less reliable for this reason.

However, as it transpired, obtaining a good tracking of heads with a generic algorithm for different crowd examples was an elusive task. Thus, a series of image features that are detected readily and tracked easily were considered as the initial step. The immediate concern would be that the features are not necessarily from the head area; they can be from different parts of the body. However, if the crowd is dense enough most features will be from the head region. But since we deal with crowds that are not sufficiently dense, the mapping of the features onto the ground plane is problematic. We have experimented with masking the non-head plane regions to eliminate those features which are defiantly not on the head plane and assume that the rest of the features are on the head plane. However, this is a very naive assumption and introduces large errors in the position of features. Depending on the specifics of the example, these errors may be more disruptive than the distortion caused by the projective transform. This issue will be discussed further in the next section.

3.5 Entropy via image features

Corner feature detection using a method which was introduced by [26] is used for feature detection in images. These features are specifically designed to be suitable for tracking. If a background image is available, the detected features are compared against the features which are detected on the background and the background features are removed from the list of detected features. As mentioned before, a mask for the head plane is used to eliminate all the features which can-

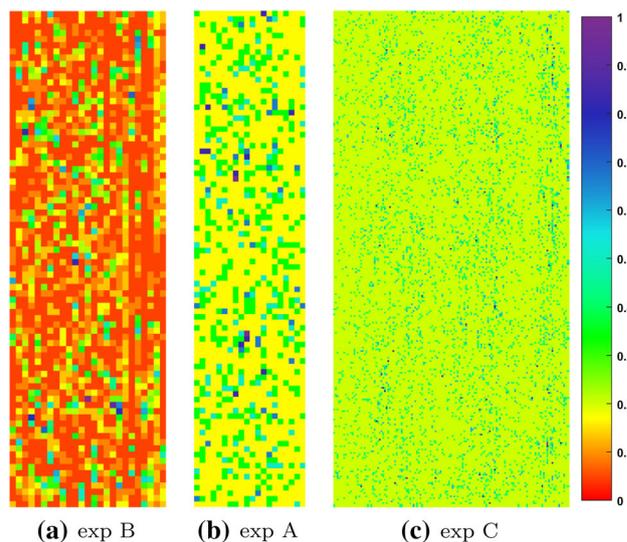


Fig. 9 Profiles for the three examples in the order of increasing entropy: $\hat{h}_k(X_{expB}) < \hat{h}_k(X_{expA}) < \hat{h}_k(X_{expC})$. (Best viewed in colour) (colour figure online)

not be on the head plane. The remaining features are assumed to be on the head plane and are mapped onto the ground plane. Entropy is calculated as before.

A visual and intuitive description of how the algorithm works is shown in Figs. 8 and 9. In Eq. 6, n_i was defined as the sum of all density counts at bin l_i in N_f frames. It can also be seen from this equation that the p_i s which determine the value of entropy are linearly dependent on these n_i s. An image showing all these n_i s where the image intensity at location l_i is dependent on the value of n_i is referred here to as n_i -map. Note that the locations on the n_i -maps are the internal positions of the features which are projected into the ground plane. Figure 8 shows the n_i -map for exp A (stairs) over a 2s time window. Since the locations with $n_i = 0$ do not affect the value of entropy, condensed versions of the n_i -maps for all the three experiments are also shown in Fig. 9. We shall call these, condensed n_i -maps, *profiles*.

Figure 9 shows the profiles for exps A&B&C in the order of increasing entropy from left to right. This effect (increasing entropy) can be seen visually. In these, the probability of feature occurrence is linearly dependent on the value of the pixels. In Fig. 9a, most of the pixels are very low-valued (red in colour). Thus, they have low probability of feature occurrence. However, note that all the points in the profile are nonzero. In contrast, there are also some isolated high-valued pixels. (These can be viewed as peaks of probability function.) In fact, they offer a sound hypothesis for features' respective locations. The background pixels have higher values in Fig. 9b (yellow in colour), meaning that the probability of feature occurrence is more evenly distributed over spatial bins. However, there are still many high-valued points (peaks) where the probability of occurrence is higher.

Table 2 Normalising entropy values for the crowd examples

	Exp B	Exp A	Exp C
h_{\min}	4.70	4.54	6.88
$h_{\max} = \log N_l$	7.60	7.50	10.64
H_k	7.24	7.38	10.59
\hat{h}_k	0.876	0.962	0.988

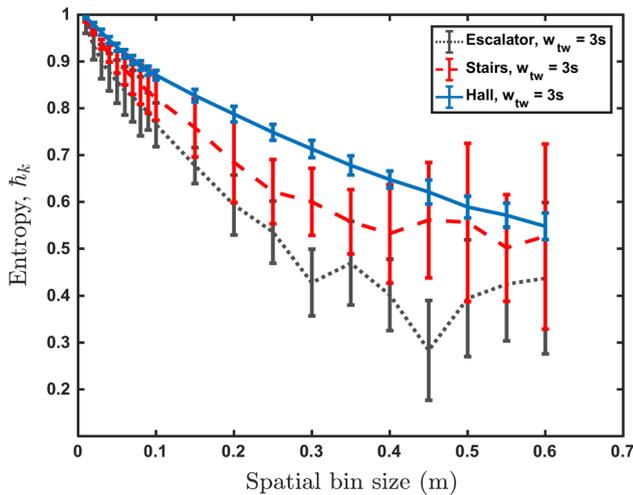


Fig. 10 Experimenting with image features and calibration at $w_{tw} = 3s$

In Fig. 9c, the background is yet higher in value (green in colour), while the peaks are less prominent. In this example, the probability of feature occurrence is more evenly distributed and thus high values of entropies are expected.

Table 2 shows the normalised respective entropies which are calculated for these examples. The normalisation values h_{\min} and h_{\max} affect the result significantly. Also, it can be seen that the values for normalised entropies are very high. This is due to the small size of the bins being used. In Fig. 10, these results reside in the upper left corner of the graph. Small bin sizes are depicted for more intuitive visualisation.

Figure 10 shows the detected entropy of the three examples using image features. The level of separation between the entropies is understandably lower. This is due to the noise which is introduced by replacing head detection with feature detection and the added distortion which corresponds to assuming feature points are on the head plane. The mean value separation still holds for all the bin sizes. It was initially noted that the distortion introduced by an approximate ground plane projection might be more disruptive than that which has been originally introduced by the projective transform. Therefore, the results for entropy via image features using image coordinates were also provided.² These

² Note that the entropy values need to be compared at corresponding bin sizes. Therefore, the reference plane is used to find a mapping between

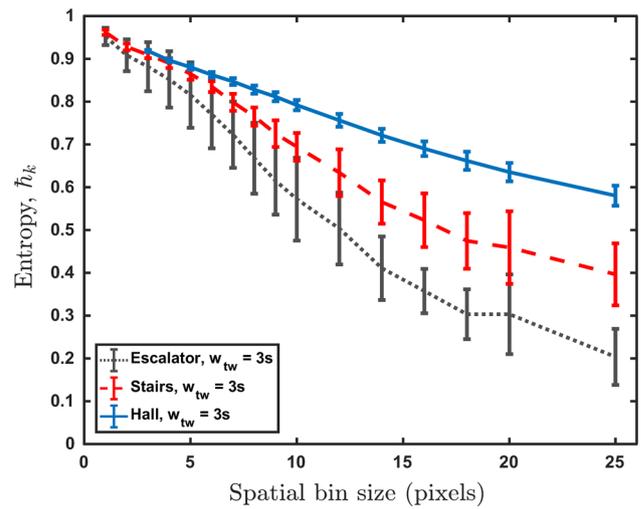


Fig. 11 Experimenting with image features without calibration at $w_{tw} = 3s$

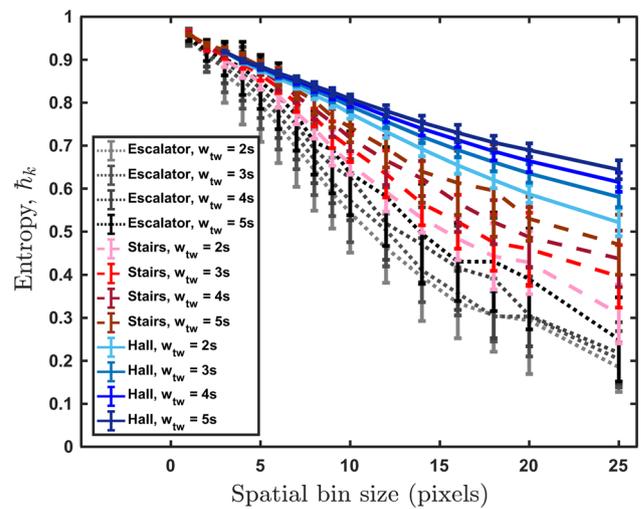


Fig. 12 Experimenting with image features and different time windows

are shown in Fig. 11. It can be seen that the results are improved and the separation is mostly achieved for the three experiments. The effect of using larger time windows is demonstrated in Fig. 12. As was described before, when larger windows are considered, more variability is observed together with the natural build-up of noise. Therefore, the value of entropy increases. However, it is worth mentioning that in the case of the experiments shown in Fig. 12, where no ground plane mapping is used, the results remain consistent. In those, the mean value separation is obtained between the entropy values of the experiments at various time window sizes.

the corresponding bin sizes between exps A, B and C. For example, the bin size of 3 pixels in exp A&B corresponds to bin size of 1.9 pixels in exp C, all corresponding approximately to a bin size of 5 cm.

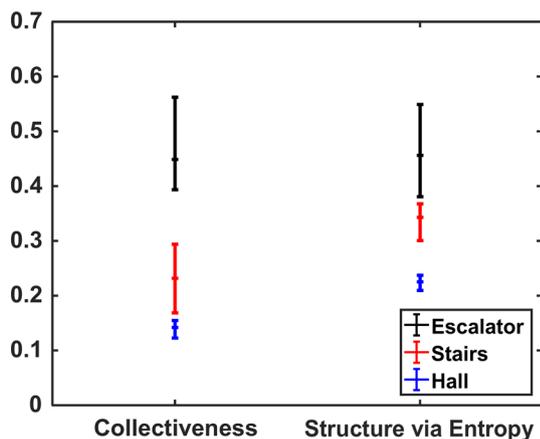


Fig. 13 Comparing collectiveness and entropy for structure

As an example of execution time, the video containing both the escalator and stairs experiments is processed at a mean rate of 10 fps with the spatial bin size of 16 pixels. This is using an Intel Core i7-2600 CPU at 3.40GHz. One notes that the execution speed decreases with the increasing number of crowd clusters in the frame. (Crowd clusters are discussed in Sect. 4.3.) On the other hand, the speed increases as a result of using larger spatial bins.

3.6 Entropy versus collectiveness

Collectiveness is a measure of collective motion that is introduced by Zhou et al. [37]. They define it as follows: “Collectiveness describes the degree of individuals acting as a union in collective motions”. Collectiveness seeks collective manifolds wherein consistent motion is observed in neighbourhoods, while global consistency among non-neighbours is obtained through intermediate individuals in neighbourhoods on the manifold. Collectiveness assigns values in the range [0, 1] to a given crowd. It requires setting a parameter, K , which defines the range of neighbourhoods in the given experimented crowd.

Collectiveness bears similarities with entropy. In order to be able to compare collectiveness with entropy directly, the notion of *structure* is introduced. As noted, entropy is basically a measure of disorder, while structure can be described as a measure of order. For a normalised entropy ranging within an interval [0, 1], structure and entropy are complementary and add up to unity: $s_k = 1 - \hat{h}_k$, where s_k is the normalised structure. Figure 13 shows a comparison between collectiveness and structure (via entropy using image features with no ground plane projection). It can be seen that collectiveness also achieves separation between these examples. Although entropy finds a larger distinction between exp A (Stairs) and exp C (Hall), collectiveness finds exp B (Escalator) and exp A (Stairs) more distinct. This is an early sign

that depending on the sample which is to be analysed one or the other method may be more effective. The most important factors which may contribute here are: (i) the density and behaviour of crowd; (ii) camera view angle and spatial resolution; and (iii) structure of the environment. It should be mentioned here that both collectiveness and entropy values depend on the respective adopted parameters of these methods. These include K for collectiveness and spatial bin sizes (w_{bin}) for entropy (temporal window, w_{tw} , is not that significant). Here, a mid-range k ($k = 20$) is used to produce the collectiveness results and w_{bin} is subsequently chosen to produce similar values for the structure in the escalator example and then used to evaluate the other two examples.

Figure 14 shows an example where collectiveness fails to produce stable and reliable results. It is worth noting that collectiveness is essentially a different concept from that of entropy. Collectiveness is best for analysing crowds with discernible motions in the form of flows and limited oscillatory motions. Figure 14 depicts an example of a stadium, wherein the initial state of the crowd is calm with sparse incoherent motions. However, an event which may occur on the pitch may trigger increased level of excitement of the crowd in the stadium arena.³ Figure 14c, d shows the values of collectiveness and entropy in the crowd for illustration. Here, the dotted red line indicates the time of the event, while the volatility of the crowd increases before the event in anticipation. In this circumstance, collectiveness does not seem to provide intuitive results. The initial state of crowd has small amounts of motions, meaning that any small group with more significant motion can override the value for the collectiveness. Further, in the absence of such groups collectiveness becomes unstable as it tries to connect incoherent sparse motions within the crowd. In contrast, entropy clearly captures the increased volatility and the change in the state of the crowd.

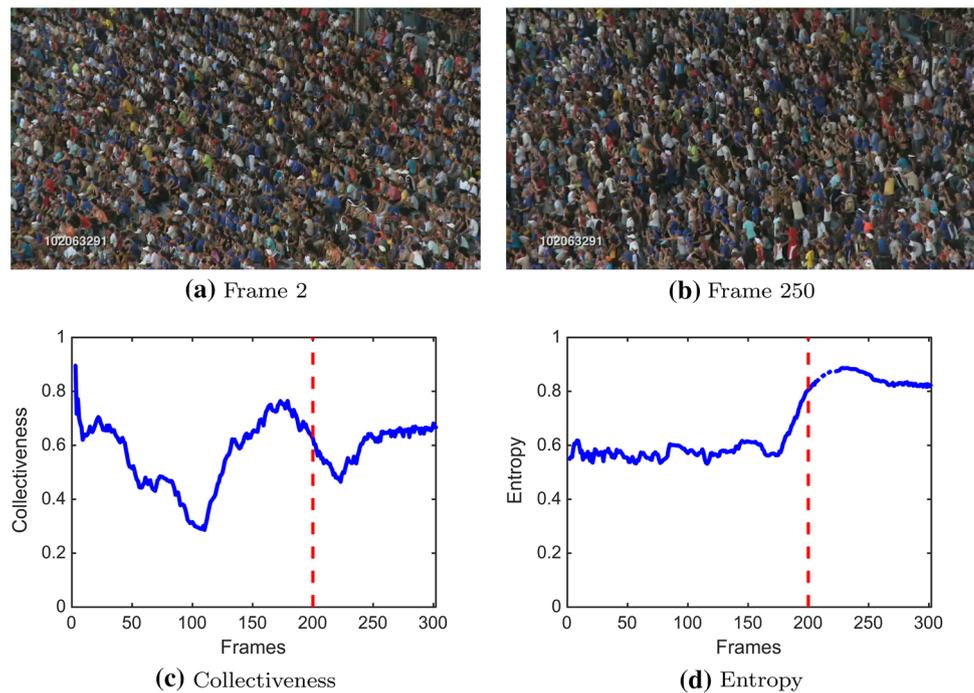
4 Discussions and future work

4.1 Other crowd descriptors

The other relevant crowd descriptor which has been recently proposed by Shao et al. [25] is *Stability*. This descriptor is defined as the property which characterizes “whether a group can keep internal topological structure over time”. Stability is a composite descriptor, and it is computed using three components that each assess one of the following stability criteria for the group members. In this, the stability of the group is measured via the stability of its members. The stable members are assumed to: (i) maintain a similar set of nearest neighbours;

³ This video footage is from the data-driven crowd analysis data set by [23]. It has a resolution of 640×360 pixels and comprises 850 frames at 25 fps.

Fig. 14 Entropy versus collectiveness for complex motion scenes. The red dotted line indicates the time of the event (colour figure online)



(ii) keep a consistent topological distance with neighbours; and (iii) be less likely to leave their nearest neighbour set. Shao et al. have compared stability with collectiveness and found a weak positive correlation between the two. It has also been shown that groups with similar collectiveness can have very different stabilities. With its focus on measuring the stability of each member, this descriptor is very useful for measuring small groups, but less suitable for dense crowds viewed at distance. Also, like collectiveness, stability was found not to be suitable for mostly stationary crowds with random oscillatory motions (e.g. spectator crowds) due to its reliance on tracklets. Stability has been shown to provide promising results alongside other descriptors for the applications such as crowd monitoring, crowd classification and retrieval. However, a detailed analysis of the behaviour of this descriptor in different crowd examples was not shown.

4.2 Internal kinetic energy

The internal energy of a crowd as a thermodynamic system, U , can possibly be used as a measure of how excited the crowd is. Irrespective of its entropy, a crowd may be in an excited/agitated state (high energy) or a calm state (low energy). On this note, it is worthy to point out that in thermodynamics, entropy and internal energy are both state variables. U is composed of two components:

$$U = U_{\text{kinetic}} + U_{\text{potential}}. \tag{17}$$

U_{kinetic} can be computed as

$$U_{\text{kinetic}} = \frac{1}{2} m v_{rms}^2. \tag{18}$$

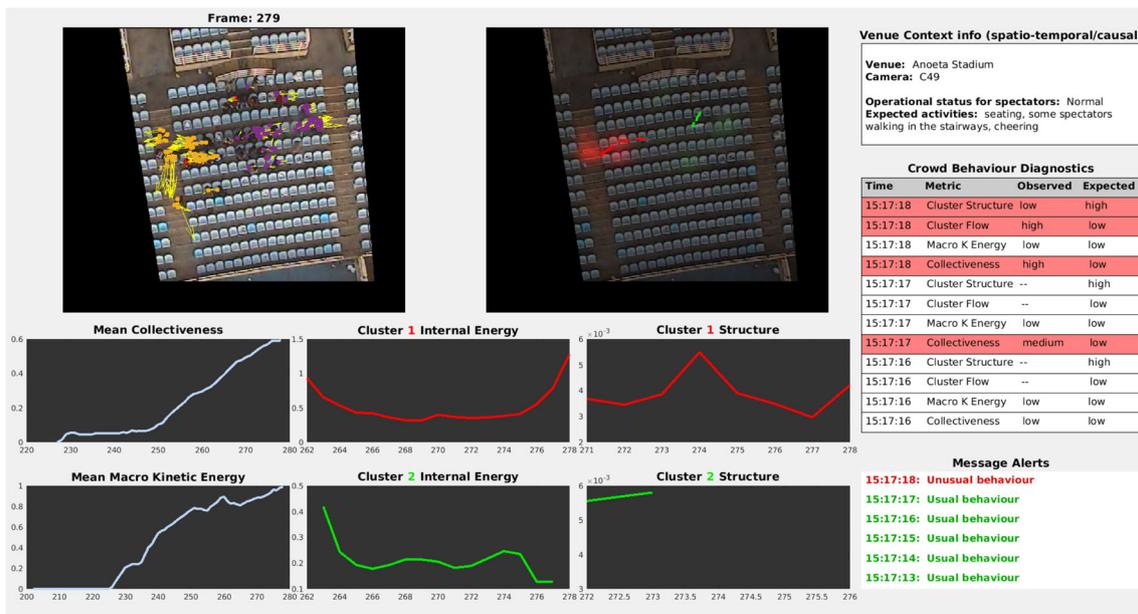
v_{rms} is the square root of the mean of the squares of internal velocity, v_i , of the particles ($v_{rms} = \sqrt{v_i^2}$). Having extracted the subgroups in the crowd and detected their flow, the internal velocity for a particle j at time t is $v_i(j, t)$:

$$v_i(j, t) = v_o(j, t) - v_f(x_i(j, t)) \tag{19}$$

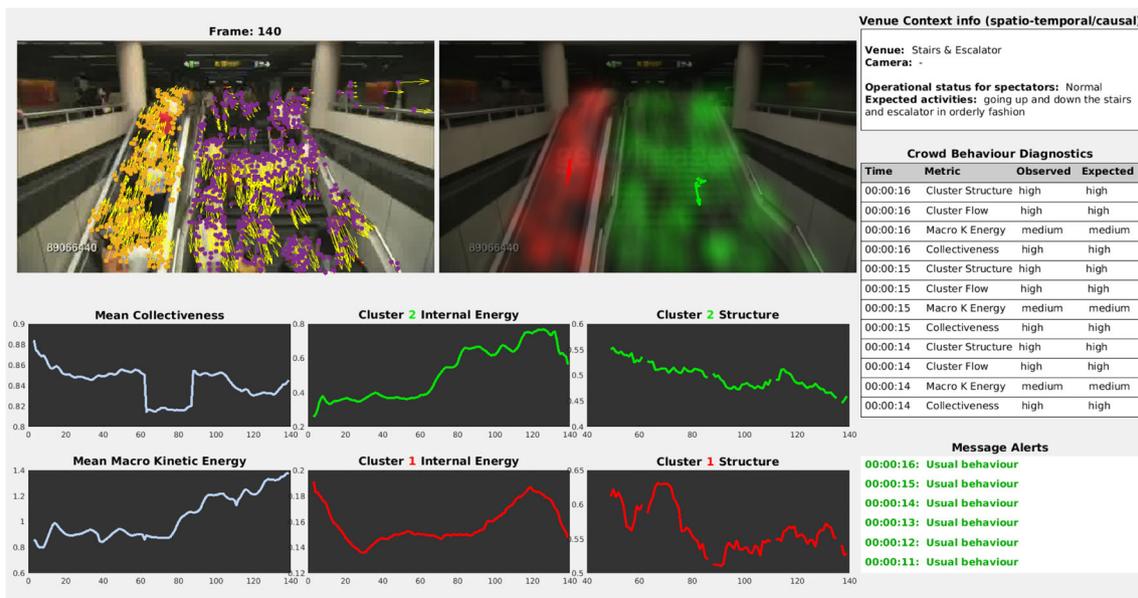
where v_o is the observed velocity and v_f is the sampled flow velocity at location $x_i(j, t)$, which is the internal location of particle j at time t .

In many occasions, sufficient information can be gathered using solely the U_{kinetic} . For example, generally for a gas at higher temperatures and lower pressure the potential energy due to inter-molecular forces becomes less significant when compared with the internal kinetic energy of the particles: $U \sim U_{\text{kinetic}}$.

$U_{\text{potential}}$ is a significantly more complex value to calculate. Two of the most prominent pedestrian modelling approaches took inspiration from gas kinetic theory where the focus is on the kinetic energy of the crowd [8,12] with the considerations that if there are more than one phase (gas, liquid, solid) present potential energy needs to be considered [12]. Hughes [14] defines the crowd potential energy as the “common sense of the task the pedestrians face to reach their common destination”. However, a directly measurable value



(a) Detecting unusual behaviour of a crowd in Anoeta Stadium



(b) Detecting unusual behaviour of a crowd on the stairs and escalator example

Fig. 15 Detecting unusual behaviour of a crowd

has not been defined. We will address $U_{\text{potential}}$ calculation in our future works.

4.3 Homogeneity and multi-scale descriptors

Entropy and internal energy are calculated at meso-scale (sub-group) level within the crowd. Collective merging [37] has been used here as the starting point for the detection and tracking of the sub-groups within the crowd. Collective

merging has two tuning parameters: α , which indicates the scale of the cluster of interest, and K , which is a parameter for collectiveness that indicates the spatial extent of a pedestrian in pixels. α and K control the scale of the sub-groups of interest. This highlights the need for having a pre-knowledge about the crowd and the scale of the desired behaviour analysis. The detection of putative crowd clusters is performed for each pair of consecutive frames using collective merging. Further, a mapping is made between the detected clusters

in consecutive frames based on their population, motion and feature points, and thereby the clusters are tracked for as long as they are in the field of view. However, there is an inherent ambiguity in determining the sub-groups within a crowd, as was denoted by [4] and [24]. Ideally, crowd attributes are assigned to the parts of the crowd which are homogeneous with regard to that attribute. Different groups can be detected within a crowd depending on the attribute which guides this segmentation. For example, it is possible to detect the segments of the crowd which have similar energy levels. Note that these segments may be different from the ones detected using entropy for example. The most intuitive and common basis for finding groups within a crowd is through detecting segments of the crowd that demonstrate collective motion [25,31,34,37]. Note that this is different from the social group formations as described by [20]. Here, the members of the detected groups do not necessarily have social attachments. Consider the example of a marathon run where the entire participant population can be considered as one group. Using the idea of collective motion as denoted by [37] is similar to segmentation based on flow.

5 Unusual behaviour detection

The work here has been performed within the eVACUATE project [6]. Its goal is to facilitate the safety and security of crowds as they are evacuated from confined spaces. This includes a holistic situation awareness and guidance system for sustaining the active evacuation route under different crowd evacuation scenarios. The work described here contributes to the situation awareness functionality of the system by detecting usual/unusual behaviour of crowd using computer vision with added context awareness. The earlier works have been published in two conference papers [3,24].

A series of experiments have been performed within the eVACUATE project, while looking into different crowd behaviour scenarios. These included experiments in an airport, a metro station and a stadium. In these, a context is established for a given crowd taking into account the event and the spatial characteristics of the venue. For instance, for a crowd at a football match, the event is the match and the venue is the stadium. During the match, it is expected that the crowd will be mostly seated. Furthermore, one of the main features of a crowd at a stadium is that they are prone to excitement. Thus, a wide range of internal energy levels is also expected for this crowd.

As an example, a series of experiments have been performed in the Anoeta Stadium, San Sebastian. Different scenarios and events have been enacted and recorded during

the evacuation of a crowd.⁴ Initially, different social groups have been established in the crowd. In this, the crowd were segmented into groups of 2, 3 and 4, while some were directed to act as individuals. Each group was asked to appoint a group leader, and a susceptibility level to be led is also assigned to each member of the group as a personality trait. A number of actors were also used in some of the scenarios to initiate certain behaviours in the crowd.

Examples of our prototype system are provided in Fig. 15 as a proof of concept. Here, the notion of crowd space has been used to define the thresholds of usual behaviour within a context. In our future work, we will look into automatically setting these thresholds and tuning parameters such as bin size for the evaluation of entropy. Figure 15a is a screen capture of the system detecting unusual behaviour of crowd in one of the experiments which we conducted at the Anoeta Stadium. Results from the same system detecting unusual behaviour in the stairs and escalator example are shown in Fig. 15b.

6 Conclusions

A new crowd descriptor has been introduced to characterize the behaviour of people using vision measurements. This descriptor is inspired from properties of statistical molecular systems and entropy. The quantification of this descriptor has been investigated, and alternative methods explored. Experiments have been performed on example crowds from two publicly available data sets and an in situ data set generated for this work. The descriptor, *entropy*, is shown to capture the desired outcome for entropy of crowd. It achieves this consistently throughout several experiments while using easily detectable image features. The effects of projective transform and mitigation strategies using calibration have been investigated. It has also been shown that *entropy* offers complementary capabilities to the set of existing crowd descriptors including *collectiveness*.

In our future work, we shall explore *Internal Energy* as a crowd descriptor. Also, we shall systematically investigate more on the performance of the currently defined crowd descriptors. The descriptors are evaluated against the inherent characteristics of crowd, such as its density and homogeneity as well as the recipient environment in which it moves. Visual variations of video footage such as view angle and lighting conditions will also be considered. Another interesting avenue which we will explore is to predict crowd behaviour through understanding the nature of the mechanics of groups

⁴ Ten scenarios have been enacted. These have been captured from above with a 756 × 568 resolution at 8 fps with the mean length of 66 s per experiment.

and their potential dispersion in context of the venue spaces, boundaries and temporal constraints.

Acknowledgements The authors are very grateful to Real Sociedad FC, at San Sebastian, Spain, for authorizing us to generate valuable in situ data at the Anoeta stadium for our research, without which we would have not achieved this research outcome at IT Innovation. This research work is partially funded by the EC under the eVACUATE project, grant number FP7-EC-313161.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Ali, S., Shah, M.: Floor fields for tracking in high density crowd scenes. In: European Conference on Computer Vision, pp. 1–14. Springer, New York (2008)
2. Burstedde, C., Klauck, K., Schadschneider, A., Zittartz, J.: Simulation of pedestrian dynamics using a two-dimensional cellular automaton. *Phys. A* **295**, 507–525 (2001)
3. Correndo, G., Arbab-Zavar, B., Zlatev, Z., Sabeur, Z.A.: Context ontology modelling for improving situation awareness and crowd evacuation from confined spaces. In: Environmental Software Systems. Infrastructures, Services and Applications, pp. 407–416. Springer, New York (2015)
4. Ge, W., Collins, R.T., Ruback, R.B.: Vision-based analysis of small groups in pedestrian crowds. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(5), 1003–1016 (2012)
5. Ali, S., Shah, M.: A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. In: 2007 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–6. IEEE (2007)
6. Evacuate eu fp7 project. <http://www.evacuate.eu/>. Online; Accessed 3 January (2017)
7. Guo, R., Huang, H.: A mobile lattice gas model for simulating pedestrian evacuation. *Phys. A* **387**, 580–586 (2008)
8. Helbing, D.: A fluid-dynamic model for the movement of pedestrians. *Complex Syst.* **6**, 391–415 (1992)
9. Helbing, D., Farkas, I., Vicsek, T.: Simulating dynamical features of escape panic. *Nature* **407**, 487–490 (2000)
10. Helbing, D., Molnar, P.: Social force model for pedestrian dynamics. *Phys. Rev. E* **51**, 4282–4286 (1995)
11. Henderson, L.F.: The statistics of crowd fluids. *Nature* **229**, 381–383 (1971)
12. Henderson, L.F.: On the fluid mechanics of human crowd motion. *Transp. Res.* **8**(6), 509–515 (1974)
13. Heneina, C.M., White, T.: Macroscopic effects of microscopic forces between agents in crowd models. *Physica* **373**, 694–712 (2007)
14. Hughes, R.L.: A continuum theory for the flow of pedestrians. *Transp. Res. B: Methodol.* **36**(6), 507–535 (2002)
15. Kirchner, A., Nishinari, K., Schadschneider, A.: Friction effects and clogging in a cellular automaton model for pedestrian dynamics. *Phys. Rev. E* **67**(5), 056122 (2003)
16. MacKay, D.J.: Information Theory, Inference, and Learning Algorithms. Cambridge University Press, Cambridge (2003)
17. Mehran, R., Moore, B.E., Shah, M.: A streakline representation of flow in crowded scenes. In: European Conference on Computer Vision, pp. 439–452. Springer, New York (2010)
18. Mehran, R., Oyama, A., Shah, M.: Abnormal crowd behavior detection using social force model. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 935–942 (2009)
19. Moore, B.E., Ali, S., Mehran, R., Shah, M.: Visual crowd surveillance through a hydrodynamics lens. *Commun. ACM* **54**, 64–73 (2011)
20. Moussaïd, M., Perozo, N., Garnier, S., Helbing, D., Theraulaz, G.: The walking behaviour of pedestrian social groups and its impact on crowd dynamics. *PLoS ONE* **5**(4), e10047 (2010)
21. Ozturk, O., Yamasaki, T., Aizawa, K.: Detecting dominant motion flows in unstructured/structured crowd scenes. In: 2010 20th International Conference on Pattern Recognition (ICPR), pp. 3533–3536. IEEE (2010)
22. Rodriguez, M., Sivic, J., Laptev, I., Audibert, J.Y.: Data-driven crowd analysis in videos. In: 2011 International Conference on Computer Vision, pp. 1235–1242. IEEE (2011)
23. Rodriguez, M., Sivic, J., Laptev, I., Audibert, J.Y.: Data-driven crowd analysis in videos. In: Proceedings of the International Conference on Computer Vision (ICCV) (2011)
24. Sabeur, Z.A., Doulamis, N., Middleton, L., Arbab-Zavar, B., Correndo, G., Amditis, A.: Multi-modal computer vision for the detection of multi-scale crowd physical motions and behavior in confined spaces. In: Advances in Visual Computing, pp. 162–173. Springer, New York (2015)
25. Shao, J., Loy, C.C., Wang, X.: Learning scene-independent group descriptors for crowd understanding. *IEEE Trans. Circuits Syst. Video Technol.* **27**(6), 1290–1303 (2017)
26. Shi, J., Tomasi, C.: Good features to track, pp. 593–600 (1994)
27. Solmaz, B., Moore, B.E., Shah, M.: Identifying behaviors in crowd scenes using stability analysis for dynamical systems. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(10), 2064–2070 (2012)
28. Wang, W., Lin, W., Chen, Y., Wu, J., Wang, J., Sheng, B.: Finding coherent motions and semantic regions in crowd scenes: a diffusion and clustering approach. In: European Conference on Computer Vision, pp. 756–771. Springer, New York (2014)
29. Wang, X., Ma, K.T., Ng, G.W., Grimson, W.E.L.: Trajectory analysis and semantic region modeling using a nonparametric Bayesian model. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–8 (2008)
30. Weng, W.G., Chen, T., Yuan, H.Y., Fan, W.C.: Cellular automaton simulation of pedestrian counter flow with different walk velocities. *Phys. Rev. E* **74**(3), 036102 (2006)
31. Wu, Y., Ye, Y., Zhao, C.: Coherent motion detection with collective density clustering. In: Proceedings of the 23rd ACM International Conference on Multimedia, pp. 361–370. ACM (2015)
32. Yi, S., Li, H., Wang, X.: Understanding pedestrian behaviors from stationary crowd groups. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015)
33. Yücel, Z., Miyashita, T., Hagita, N.: Modeling and identification of group motion via compound evaluation of positional and directional cues. In: 2012 21st International Conference on Pattern Recognition (ICPR), pp. 1172–1176. IEEE (2012)
34. Zhou, B., Tang, X., Wang, X.: Coherent filtering: detecting coherent motions from crowd clutters. In: Computer Vision–ECCV 2012, pp. 857–871. Springer, New York (2012)
35. Zhou, B., Tang, X., Wang, X.: Measuring crowd collectiveness. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2013)

36. Zhou, B., Tang, X., Wang, X.: Learning collective crowd behaviors with dynamic pedestrian-agents. *Int. J. Comput. Vis.* **111**(1), 50–68 (2015)
37. Zhou, B., Tang, X., Zhang, H., Wang, X.: Measuring crowd collectiveness. *IEEE Trans. Pattern Anal. Mach. Intell.* **36**, 1586–1599 (2014)
38. Zhou, B., Wang, X., Tang, X.: Random field topic model for semantic region analysis in crowded scenes from tracklets. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3441–3448 (2011)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Banafshe Arbab-Zavar received her B.Sc. in computer engineering from the Ferdowsi University of Mashhad, Iran, in 2004 and her Ph.D. in image processing from the School of Electronics and Computer Science, University of Southampton, UK, in 2009. Since then, she has been conducting research at the University of Southampton. Her research interests include image, video and signal processing and feature extraction. She has a published track record in many application areas such as biometrics, human and crowd behavior, tsunami detection and visual detection and tracking.

Zoheir A. Sabeur received his M.Sc. and Ph.D. in Theoretical Physics from the University of Glasgow, UK, in 1986 and 1990, respectively. In 2019, he joined, as Associate Professor of Data Science, the Department of Computing and Informatics, Bournemouth University, UK. Prior to this appointment, he worked as Science Director and leader of his Big Data research group at IT Innovation Centre, School of Electronics and Computer Science, University of Southampton, UK (2009–2019). He is also Visiting Professor of Data Science at the Colorado School of Mines, Golden, Colorado, USA, since 2017. His research interest is focused on data science and artificial intelligence. Specifically, he specializes in human and natural processes behavior detection and understanding using machine intelligence. He additionally researches on data fusion and reasoning for the advancement of critical decision-support systems and situation awareness. This is applied in the domain of physical and cybersecurity. He is Fellow of the British Computer Society and Chair of the Discrete Global Grid System (DGGs) Specification and Domain Working Groups at OGC international. He has published more than 120 papers in science and engineering journals, conference proceedings and books.