

# Crowd Behaviour Understanding Using Computer Vision and Statistical Mechanics Principles



Zoheir Sabeur and Banafshe Arbab-Zavar

**Abstract** Crowd behaviour understanding in computer science is a research discipline which has grown rapidly in recent years. Specifically, we are currently able to generate large and high-resolution observation data through crowd sensing in varieties of spatial environments. This has also given us the advantage to adopt computer vision methods for detecting human behaviour. In this study, we adopted statistical mechanics principles with analogies of entropy and kinetic energy in classical molecular gases to derive features which describe crowd motions. These are implicitly measured, as basis for understanding *behaviour*, using a holistic three-dimensional representation, of crowd features including *structure*, *energy* and *translation*. As a result, we measured those features using computer vision in the view of machine understanding crowd behaviour. *Usual behaviour* is established from our expected crowd motions in context of the specific recipient spaces of our experiments. The behaviour which does not fall within the expected usual behaviour measurement is considered as an *unusual behaviour*. This research work was initiated in 2013 under the eVACUATE project, while it is currently ongoing under the S4AllCities project since 2020.

## 1 Introduction

The most recent development in understanding crowd behaviour using machine intelligence over the last two decades has been boosted by the affordability and availability of smart sensing for observing humans in spatial environments. In particular, the deployment of smart CCTV cameras for monitoring the safety and security of citizens in public spaces has become the regulated norm of security modus operandi in the majority of modern cities around the world. But in order to reach scalability in the surveillance of large crowds' critical safety in public spaces,

---

Z. Sabeur (✉) · B. Arbab-Zavar

Department of Computing and Informatics, Faculty of Science and Technology, Bournemouth University, Dorset, England, UK

e-mail: [zsabeur@bournemouth.ac.uk](mailto:zsabeur@bournemouth.ac.uk)

one will require the support of computer machine intelligence for focusing only on those events where specifically detected behaviour requires attention from security practitioners and first responders. It is in this perspective that we specialised over the years, and since the launch of the international eVACUATE research project [1], on the automated detection of individuals or groups of humans and unusual behaviour in public spaces using artificial intelligence (AI). Specifically, we characterised crowd as a dynamic system whose thermodynamic-related parameters can be directly or indirectly measured using sensory observations including vision-based features. These are then used for detecting and understanding human behaviour while adapting statistical mechanics principles which relate to the state of dynamic order or disorder of the crowd system at multiple spatial scales. This was the foundation of our research work over the years, which is ongoing, regarding human behaviour understanding using AI.

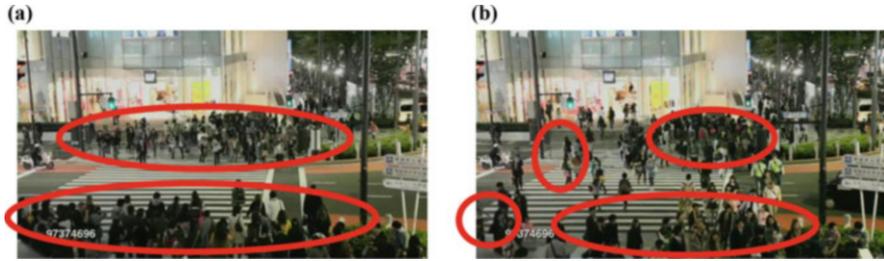
## ***1.1 Definition of Crowd***

The definition of crowd, for understanding crowd behaviour, requires deep insights into the important selected features representing crowds. These should be viewed in context of a complex dynamic system. A crowd ‘system’ can be considered as a collection of loosely coordinated individuals who may share a common and temporarily bound interest. This covers spectators and people moving.

From a practical sense, there are some nuances to the simple definition presented above. Figure 1 shows two images of a pedestrian crossing at two different times. Initially there are two distinct crowds which each desire to cross to the other side of the road. This is shown by the two large red ellipses in Fig. 1a. After the situation has evolved, there are a number of different possibilities as shown in the red ellipses in Fig. 3b. These two figures show that the notion of crowd refers to a dynamically changing system which may potentially undergo some phase transitions, in this case crowd splitting or merging through time. The intention of each individual in the crowd is unpredictable instantaneously, but it could become understood, therefore possibly predictable over the time, as it is shown in Fig. 1.

### **1.1.1 Crowd Multiple Scales**

There is also the notion of scales which needs to be considered when one tries to understand crowd behaviour. This is indeed needed to study dynamically complex systems. Namely, and for the case of crowd systems, the observed collective crowd behaviour is related to the inner dynamics of each behaviour of individuals within the crowd. It also includes the entailed learning processes between individuals and their abilities to share a collective intelligent crowd behaviour.



**Fig. 1** Pedestrian crossing showing dynamic evolution of crowds at different times (a) and (b)



**Fig. 2** Microscale view of crowd with single individuals

Crowd modelling challenges and the interpretation of empirical observation data go hand in hand with the multi-scaling perspectives. In fact, the methods used for understanding or modelling crowd behaviour employ multiple-scale perspectives, in order to generate suitable mathematical structure describing crowd dynamics at each individual spatial scale. The specific scales at which the crowd can be defined are as follows:

*Microscale* Crowd structure and behaviour is identified by analysing each composite individual behaviour in the crowd. The state of each of individual behaviour is computed using features such as position in space and velocity. These are understood as time-dependent variables (see Fig. 2 for illustration).

*Mesoscale* Crowd structure and behaviour is understood through the dynamics of distinct patterns representing clusters of individuals who may share similar behaviour. Namely, this is represented by spatial cluster or group positions, collective velocities and kinetic energies. In this case the crowd system is represented statistically through a distribution function over the mesoscale-based states. This is illustrated in Fig. 3.

*Macroscale* Crowd structure and behaviour is assimilated as a continuum (or blob), where its dynamic state is described by average quantities such as density, central position, velocity and kinetic energy. These features are time- and space-dependent variables. They are statistical averages of the microscale states of individuals. This is illustrated with two distinct crowds as shown in Fig. 4.



**Fig. 3** Mesoscale view of a crowd with a group of individuals sharing behaviour



**Fig. 4** Macroscale view of a crowd with two distinct crowds (blobs)

### 1.1.2 Crowd Behaviour Considerations

Crowd behaviour detection requires the measurement and understanding of the dynamic motion or mechanics of the crowd at multiple scales. This may be highly dependent on the unfolding of contextual events and the nature of the recipient spatial environment in which crowd evolves with time. Using a pedestrian crossing as an example (see Fig. 4), the expected behaviour of the crowd is that they intend to cross safely to the other side of the road. In such context of spatial environment, human behaviour may be considered unusual if the crowd at any considered macro-, meso- or microscale deviates from what is expected. Thus, with this example, unusual behaviour could be the whole crowd trying to cross with the presence of cars. Equally, groups within the crowd may want to do such thing or indeed a single individual. But note that the so-called unusual behaviour may sometimes not be ‘compromising to safety’, but it is just unexpected to occur given the context of rules concerning pedestrian crossings. In much of the literature, unusualness is also defined as a statistical deviation from what is happening overall (the so-called ‘usual’ behaviour).

Below is a set of considered definitions of pedestrian behaviour with typically observed features for measurements and understanding:

- *Pedestrians taking detours or moving in a different walking direction to the main crowd, with the intention of taking the fastest route, than that of the crowd, in*

*order to reach their specific desired destination. This is also not the shortest route spatially.*

- *Pedestrians in usual circumstances keep to individual optimal speeds, the value of which is normally distributed around a mean of 1.34 m/s.*
- *Pedestrians, when they can, usually keep a certain distance from one another, as well as from pavements, walls and other obstacles. This distance gets smaller, with increasing pedestrians speed and/or density.*
- *Pedestrians' speeds can considerably increase in context of a perceived situation which may lead to compromising their safety and security. Their individual motions will appear random to almost unpredictable.*
- *At sufficiently high crowd densities, the motion of pedestrians is observed to be similar to that of fluid flows. In this case, concepts and associated features of fluid flow turbulent diffusion and advection, i.e. dispersion, can be adopted.*
- *Crowd behaviour primarily including motions is not due to immediate neighbours' interactions but often distant ones too. This is often caused by so-called behaviour propagation in the crowd.*

The above-mentioned crowd behaviours are viewed under the framework of a complex dynamic system, while simulation models use specific features to reproduce them realistically. In other words, the following capabilities of human behaviour in context of crowd systems should be reached while they are embedded in performing numerical models:

- **Ability to express a strategy:** Humans are capable of developing specific strategies related to their organisational ability depending on their own state and on that of the entities in their immediate vicinity. These can be expressed without the application of any principle imposed by the outer environment.
- **Heterogeneity:** Crowds, irrespective of their types, can be assumed to be heterogeneously distributed. This includes, in addition to different walking capabilities, the possible presence of leaders and the individual level of experience or prior knowledge.
- **Interactions:** These can involve individuals within the crowd connected to their immediate neighbours but also distant ones. In fact, crowd systems can be assumed to communicate at various spatial scales and may possibly choose different interaction paths, depending on the circumstances and spatial boundary conditions in which they could be.

Recall that there are conditions of natural groupings or clustering in crowds. This was illustrated in Fig. 3, when we introduced crowd mesoscale behaviour earlier. In this case, one also introduced the concepts of 'seed' and seed behaviour understanding. The seed may be considered at the leading individual in a grouping or cluster at mesoscale. The seed's behaviour is basically the closest to the group aggregate. Thus, the seed can be considered to influence the rest of the group and beyond. It is understood that a seed may be the originator or the source of behaviour which not only has the potential to lead a cluster and influence its motion, therefore

behaviour, but also the triggering mechanism for the propagation of such behaviour across the crowd system in the macroscale.

## 2 Crowd Behaviour Detection and Modelling

### 2.1 *From Motions to Behaviour Understanding*

As was discussed earlier, crowds can be viewed as complex systems, where their behaviour is determined by the inner dynamics of the system, including those respective to the macro-, meso- and microscales. To infer a full description of the system's behaviour from these dynamics remains a challenging task. It is also worth noting that such description of system dynamics is often not fully achieved, due to limited sensor observations for measurements and/or the constraints due to regulations for openly experimenting on the spatial environment with crowds of interest. Further, the full analysis of different scales and various types of dynamics within the crowd system will require the use of different mathematical models. In an attempt to handle the complexities and ambiguities of the realm of crowd behaviour detection, research efforts which deal with the problem of crowd behaviour often settle for answering specific questions about crowd behaviour instead of offering a full description of crowd behaviour as a complete theory. Thus, some specific questions on crowd behaviour are considered in this chapter. These are listed below:

- *Is the crowd behaving in an unusual manner?*
- *Is the crowd showing signs of specific behaviours (for example, panic)?*
- *Is the crowd changing behaviour due to the actions of an individual or group of individuals?*
- *Is this change in behaviour propagating in the crowd?*

### 2.2 *Measurement of Crowd*

The question of 'how to measure a crowd' can have many different answers. The answer may depend on the type of information required and the level of granularity of interest which needs to be adopted. Here, the inner workings and state of a whole crowd is investigated. As a result, a set of features are defined for the crowd. These features are chosen with the aim of characterising the state and type of crowds in terms of human behaviour. Nevertheless, we will assume that the crowd is homogeneous in type. In this, while the micro-level motions within the crowd are observed and measured, the defined properties and features describe the overall crowd as one type of crowd. An example of such homogeneous crowd type can be a competitive marathon, where the crowd is composed of single individuals who all

share the same goal. The homogeneity assumption will also hold for cases such as a shopping mall wherein there may be small groups of people as well as individuals while one assumes that they have the same goal of shopping in such environment. The method proposed here is motivated by physical analogies of thermodynamics and statistical mechanics, where the macroscopic properties of matter are derived from microscopic properties and states of the underlying molecular systems.

### 2.2.1 Crowd Analogies to Physical Systems

Various physical analogies and modelling approaches have been used in crowd and traffic modelling. A physical model requires a hypothetical structure of controlled parameters which need to be fine-tuned for simulating crowd dynamics that is in accord with experimental observations. In this section, some of the more popular modelling analogies in this domain are reviewed and evaluated. These include (a) *cellular automata*, (b) *social force model* and (c) *molecular fluid dynamics*.

- (a) *Cellular automata (CA)* has been used to simulate crowd dynamics in situations such as evacuation [2–5]. In this, CAs evaluate the feasibility of different evacuation scenarios. It has also been shown that CA can simulate certain effects such as line formation in the crowd [6]. However, CA does not aim to capture all the microscopic dynamics but only that which is necessary to derive a specific macro effect.
- (b) *Social force model* is another popular method for crowd simulation [7, 8]. It has also been used to detect points with high social friction within the crowd [3, 9]. In particular, Helbing et al. noted [9]:

The motion of pedestrians can be described as if they would be subject to ‘social forces’. These ‘forces’ are not directly exerted by the pedestrians’ personal environment, but are a measure for the internal motivations of the individuals to perform certain actions (movements).

In essence, the social force model is based on a simple model wherein the individuals move according to their goals and environmental constraints. It is assumed that each individual in the crowd has a desired direction and velocity while seeking to keep a social distance from other members of the crowd as well as avoiding hitting walls. To calculate the social force model, an estimate of the individual goals is required. Other methods have been proposed to estimate the individual desired directions and velocities in a crowd [3, 9]. A bag-of-words method is used to select features from within the social force fields in consecutive frames. These bag-of-words features are subjected to further learning of unusualness detection in crowds using latent Dirichlet allocation (LDA) [9].

- (c) *Molecular fluid dynamics* has also been investigated for modelling pedestrian motions. Henderson was the first to propose a gas-kinetic model for pedestrian flows [10]. Using this basis of a Boltzmann-like gas-kinetic model, Helbing [11, 12] developed a special theory for pedestrians, distinguishing between different

groups within the crowd with different types of motions and goals. Moore et al. [13] argue against the gas-kinetic-based modelling of crowd for high-density crowds and note that for a high-density crowd, the behaviour appears to be liquid-like with interaction forces dominating the motion of pedestrians.

We propose to use analogies from thermodynamics and statistical mechanics for describing the state of crowds. While thermodynamics is concerned with heat and temperature and their relationship to energy and work at molecular matter levels, our major interest here is to derive macroscopic properties of crowds from statistical mechanics, in terms of the microscopic constituents of crowds. It is this conceptual link between the microscopic constituents of matter and its macroscopic properties that we needed to borrow while adapting thermodynamics and statistical mechanics principles to derive macroscopic features of crowds from their microscopic constituents. These are individuals within the crowd.

Following the above, we set up holistic features in a way that would enable us to describe and differentiate between different kinds of crowds and also different states of a crowd. As will be discussed in the next section, three parameters are postulated. These are *structure*, *energy* and *translation*. In this case, any crowd system type can be projected onto a point within the structure-energy-translation dimensional space. The aim is to achieve a good separation between different types of crowd in this three-dimensional space. These parameters are then used in a contextual crowd behavioural model which models the normative behaviour of crowds within well-defined situations. If there are discrepancies between the expected and the perceived behaviours, the behaviour is deemed to be unusual.

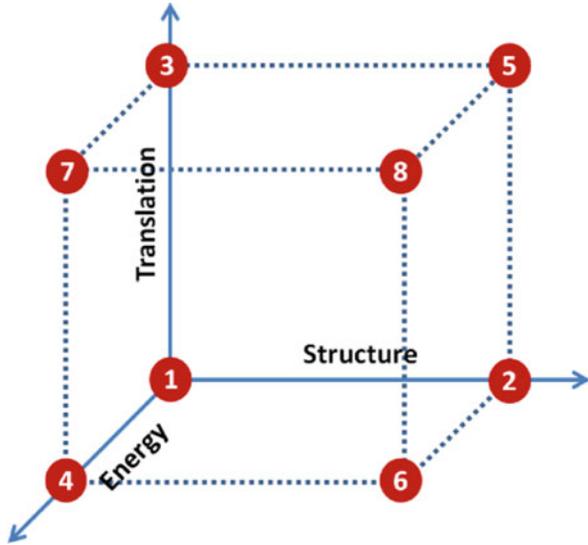
### 2.2.2 Crowd Representation with Its Holistic Features

We assume that a force keeps the members of the crowd together. The strength of connections between the members of the crowd 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 of the crowd simply refers to *energy*. We also consider that the crowd moves in space, while we refer to as *translation*. Figure 5 shows a representation of the *structure-energy-translation* crowd space.

Table 1 also illustrates a set of hypothetical examples of various types of crowds, while Fig. 5 shows where these reside in the *structure-energy-translation* crowd space.

Regarding the values of the structure parameter, it is worthy to note that a high structure score may denote one of the two underlying reasons: (i) a high social interaction between the members of the crowd (the members of the crowd maintain a pattern within a crowd), or (ii) a high value for structure may also be the result of an enforced structure by the environment (barriers, passages and doorways are examples of elements which can impose environmental structure). The structure

**Fig. 5** Crowd space with hypothetical examples

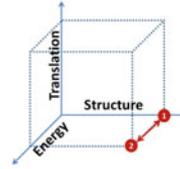


**Table 1** Hypothetical examples of crowd states

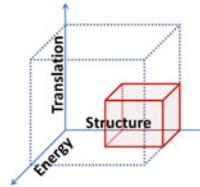
1	A number of individuals walking slowly in different directions
2	Very unmotivated crowd at a less excited football game
3	Shopping centre at closing time
4	A group of panicking people locked in a room
5	People on an escalator
6	Crowd at a football match celebrating a winning goal
7	People escaping towards an exit
8	A bull run (when the bull arrives)

parameter only evaluates the level of structure in the crowd and does not differentiate between the pedestrian-imposed environment structures.

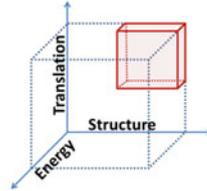
As shown in Fig. 5, a crowd may reside in any location in the *structure-energy-translation* crowd space. However, for any given situation, there would be an expectation of where the crowd should be, while a divergence from this expected, or desired, position may be a cause for alarm. Figure 6 shows sub-spaces of ‘usual’, or expected, crowd behaviour under various contextual situations and crowd types. By mapping the crowd into the *structure-energy-translation* crowd space and learning the limits of ‘usual’ behaviour, a crowd with unusual behaviour can be defined as



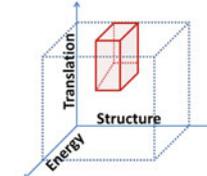
(a) Spectators at a stadium while a football match is in progress. State 1 shows the low energy state when the crowd is motionless, while state 2 represents a high energy state. For example when the crowd is celebrating a winning goal.



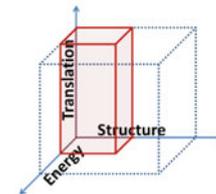
(b) Spectators at a stadium (before and after a football match). Here the crowd has a non-zero translation.



(c) Walking crowd on an escalator with high Translation and Structure, but low Energy.



(d) Walking crowd on stairs with lower Structure, as each individual is moving with respective own speed, leading to higher Energies, when compared to (c).



(e) Crowd at an airport main entrance hall. Low Structure is observed, with fluctuating values for Translation and low to medium Energy levels.

**Fig. 6** Various ‘usual’ behaviours in the *structure-energy-translation* crowd space. (a) Spectators at a stadium while a football match is in progress. State 1 shows the low energy state when the

a crowd which does not fall within the limits of perceived ‘normality’ or strictly speaking ‘usualness’.

### 2.2.3 Approach

As mentioned before, we draw analogies from thermodynamics and statistical mechanics principles. The concept of crowd *energy* refers to its *internal energy*, while *translation* relates to crowd flow velocities. These velocities may be derived at various scales. Namely, at micro-, meso- or macroscales. As for the concept of crowd *structure*, it relates to the *entropy* of the states of a molecular system.

### 2.2.4 Translation Through Flow

As noted earlier, crowd flow can be derived at different scales. The most interesting of which is the one at mesoscale, as it concerns the flow of sub-groups within a crowd. Here, the term flow is used interchangeably with the term flow velocity. In fluid dynamics, flow velocity,  $v$ , is defined as

$$v = \frac{\dot{m}}{\rho \cdot A} \quad (1)$$

where  $\dot{m}$  denotes the fluid mass flow,  $\rho$  is its density and  $A$  is the flow cross-sectional area. With consideration of a sub-group within the crowd, its density  $\rho$  can be computed using the entire volume occupied by the sub-group. The number of individuals crossing cross-sectional planes of the crowd flow can be counted to find the mass flow  $\dot{m}$ .

In some circumstances a sub-group within the crowd can be represented by a ‘Gaussian blob’, of which its speed and direction can be denoted by the mean speed and direction of its constituents. This will be referred to as *translation*. Hence, *translation* is known as the measurement of how an entire sub-group (or whole group) travels in space, while flow measures the rate at which a mass of fluid crosses a plane.

---

← **Fig. 6** (continued) crowd is motionless, while state 2 represents a high energy state. For example, when the crowd is celebrating a winning goal. **(b)** Spectators at a stadium (before and after a football match). Here the crowd has a non-zero translation. **(c)** Walking crowd on an escalator with high translation and structure but low energy. **(d)** Walking crowd on stairs with lower structure, as each individual is moving with respective own speed, leading to higher energies, when compared to **(c)**. **(e)** Crowd at an airport main entrance hall. Low structure is observed, with fluctuating values for translation and low to medium energy levels

### 2.2.5 Internal Kinetic Energy

The internal energy  $U$  of a crowd as a thermodynamic system can be used as a measure of how excited a crowd can be. It is defined as follows:

$$U = U_{\text{kinetic}} + U_{\text{potential}} \quad (2)$$

$U_{\text{kinetic}}$  and  $U_{\text{potential}}$  represent the *kinetic* and *potential* energies, respectively.

The kinetic energy  $U_{\text{kinetic}}$  is defined as follows:

$$U_{\text{kinetic}} = \frac{1}{2}mv^2 \quad (3)$$

where  $m$  and  $v$  represent the mass and internal flow velocity of a given sub-group within the crowd system.

As for the potential energy  $U_{\text{potential}}$ , its calculation is substantially complex to derive, particularly in context of crowd system thermodynamics. Its specifically relates to molecular systems which undergo thermodynamic phase transitions, where it is paramount to compute their potential energy.

However, some important pedestrian modelling approaches took inspiration from molecular systems theories with thermodynamic phase transitions [11, 12]. These postulate crowd potential energy as the ‘common sense’ of tasks pedestrians would take for reaching their expected destination. Nevertheless, such approach is not yet practical for us to implement in our experiments on crowd behaviour understanding. Therefore, we have not considered it in this study.

In the next section, we will particularly discuss the notion of entropy as an analogy to crowd structure. The computation of crowd *structure* while using analogous methods for calculating entropy is discussed with results presented.

### 2.2.6 Structure Through Entropy

Although initially defined within thermodynamics, the concept of entropy was generalised using Maxwell-Boltzmann classical statistical mechanics theory [14]. For example, entropy,  $S$ , is simply a measure of disorder in a molecular system:

$$S = -k_B \sum_i p_i \ln p_i \quad (4)$$

where for a classic molecular system with a discrete set of microstates,  $p_i$  is the probability of occurrence for microstate  $i$ .  $k_B$  is the Boltzmann constant.

The same concept of entropy was also translated in 1948, under Shannon’s information theory in computer science and informatics [15]. Entropy, mostly denoted by  $H$ , is in this case a measure of uncertainties in random variables in data communication systems:

$$H = -\sum_i p_i \log_b p_i \tag{5}$$

Entropy is defined at a macroscopic level, where a given macroscopic state can have varying microscopic statistical realisations. The initial definition of entropy in statistical mechanics,  $S = k_B \ln W$ , connects entropy directly to the number of microstates,  $W$ , which corresponds to the macroscopic state of the given system.

Considering the states of matter, in classical terms which are solid, liquid or gas, the levels of entropy for these states can be intuitively understood. In a solid state, molecules oscillate in a vicinity of a fixed location, and the entropy is low. In a liquid state, molecules move freely but keep distances from one another, while the entropy is intermediate in values. However, in a gas state, molecules move randomly anywhere, while entropy increases to a higher level. Entropy, here, increases across these three matter states, with growing uncertainties on the location the constituting molecules of matter.

As noted above, *entropy* is really a measure of disorder, while in this case *structure* can be canonically described as a measure of *order*. For a normalised entropy in the range of [0, 1], structure and entropy are complementary and add up to unity. One of the challenges in evaluating the value of *structure* using the concept of *entropy* is that for each crowd example, only a sub-set of all possible microstates represents that macro-state is observed. Therefore, it is not possible to count the number of microstates or calculate their probabilities directly. An extra step is required to infer a model or a description for all the possible microstates using the observed microstates. Figure 7 shows a diagram of the required steps for calculating the entropy of crowd using the observed set of microstates.

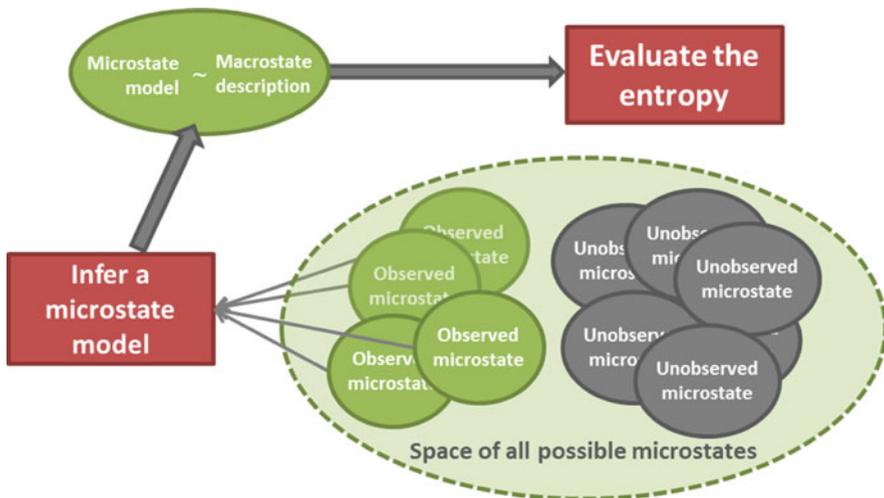


Fig. 7 Evaluating entropy through the observed microstates of a crowd

**Table 2** Micro-space modelling parameters

Notation	Definition
$N_f$	Number of frames in the time window which is analysed
$N_p$	Average number of individuals in $N_f$ frames
$N_l$	Number of spatial bins
$l_i$	$i$ th spatial bin
$Y$	A discrete random variable defined on discrete sample space $S_d$
$S_d$	A numerical and discrete sample space of densities at a location
$(f_Y)_i$	Probability mass function for $X$ at bin $l_i$

**Fig. 8** Crowd density map

### 2.2.7 Calculating Entropy

Before discussing the model, it is useful to define our notations used in the calculation of *entropy*. These notations are listed in Table 2. Also, Fig. 8 illustrates the values for crowd density at the centre of each spatial bin.

### 2.2.8 Approach 1: Preserving the Density Pattern

The joint entropy of a population of  $N_p$  individuals scattered in  $N_l$  locations with probability mass function  $(f_Y)_i$  is described as follows:

$$H(X_1, \dots, X_{N_p}) = - \sum_{x_1 \in \mathcal{L}_X} \dots \sum_{x_{N_p} \in \mathcal{L}_X} P(x_1, \dots, x_{N_p}) \log [P(x_1, \dots, x_{N_p})] \quad (6)$$

where  $X_k$  is a triple  $(x_k, \mathcal{L}_X, \mathcal{P}_{X_k})$  and 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  $\mathcal{P}_{X_k} = \{p_{k,1}, p_{k,2}, \dots, p_{k,N_l}\}$ , with  $P(x_k = l_i) = p_{k,i}$ . Here  $\mathcal{P}_{X_k}$  and the joint probabilities,  $P(x_1, \dots, x_{N_p})$ , are unknown. The joint probabilities can be calculated using the probability mass functions  $(f_Y)_i$ . However, the computation cost is in the order of  $O(N_l^{N_p})$ . More efficient algorithms can reduce this computation cost. However, we argue against the validity of this approach, since it is prone to over-fitting the model to the sample set of observed microstates. Relaxing some of the conditions in this model may be favourable.

### 2.2.9 Approach 2: Preserving the Density Pattern with Independent Pedestrians

One of the conditions which can be relaxed in the first approach is the assumption of dependence between the positions of pedestrians. In the example below, it will be shown that although there is a reason to believe that these positions are in fact dependent, sufficient information is not available to understand their dependencies accurately and in an unbiased manner.

In support of the dependency argument, let us consider that people in a crowd system tend to keep distances from each other, known as personal space. Also depending on the relationships between the pedestrians, they may tend to further avoid other pedestrians or groupings. From a different point of view, consider the following example: A number of clusters of pedestrians are observed in different locations. There may be different causes for this effect. *Hypothesis A*: There might be some relationship between members of the crowd (the second person goes to the place where the first person randomly selected). In this, even if the initial selection for the person 1 was fully random with equal chances, due to the high correlation between the first and second persons, what is observed is an environment where a certain location seems very popular. However, equally probable is that the location itself is indeed popular and people cluster there for that reason (this will be called *Hypothesis B*). The point is that sufficient information is not given in favour of either *Hypothesis A* or *B* in the above example.

We propose then that when analysing crowd formation through a few correlated frames, the simpler model which can exhibit similar outcomes is more viable. In this model, the locations of pedestrians are considered to be independent. We hypothesise that a pattern is formed in the crowd if each individual is bounded by the same pattern. Also, when taking this approach, the calculation of *entropy* simplifies significantly.

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

$$P(x_j = l_i) = \frac{n_{i,j}}{N_f} \quad (7)$$

Given that the location of individuals is considered as independent and that there is no differentiation between individuals, the probability of any individual selecting bin  $l_i$  is the same as any other. Thus, an estimate of the probability of selecting bin  $l_i$ ,  $P(x = l_i)$ , can be given by

$$P(x = l_i) = \frac{\sum_{k=1}^{N_p} P(x_k = l_i)}{N_p} = \frac{\sum_{k=1}^{N_p} \frac{n_{i,k}}{N_f}}{N_p} = \frac{\sum_{k=1}^{N_p} n_{i,k}}{N_f N_p} = \frac{n_i}{N_f N_p} \quad (8)$$

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 of one another, the joint entropy of the crowd,  $H(X_1, \dots, X_{N_p})$ , simplifies to the following:

$$H(X_1, \dots, X_{N_p}) = \sum_{k=1}^{N_p} H(X_k) \quad (9)$$

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

Thus:

$$H(X_1) = H(X_2) = \dots = H(X_{N_p}), \quad (10)$$

$$H(X_1, \dots, X_{N_p}) = N_p H(X) \quad (11)$$

where  $X$  is a triple  $(x, \mathcal{L}_X, \mathcal{P}_X)$ , and 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. (5). The crowd entropy in Eq. (8) can be computed in linear time.

## 2.2.10 Pre-processing

Three pre-processing stages are required before the entropy can be computed. These are specified as follows:

### 2.2.11 Real-World Pedestrian Locations

The locations of pedestrians in an image have been subjected to projective transform. The real-world positions can be retrieved using the camera calibration matrix and head-height plane homography transforms.

### 2.2.12 Internal Position Estimation

The internal position of each pedestrian within the crowd,  $x_i$ , is also 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 internal position in a time step  $dt$  can be calculated as

$$dx_i = dx_o - v_f dt \quad (12)$$

where  $v_f = \frac{\dot{m}}{\rho \cdot A}$ ,  $\dot{m}$  is the estimated mass flow,  $\rho$  is the mass density and  $A$  is the area. For a calibrated footage and given the density maps,  $A$  and  $\rho$  can be calculated. Given the tracks of pedestrians, the vertical and horizontal mass flows are estimated at two vertical and horizontal surface planes through the mid-point of the crowd's spatial space.

### 2.2.13 Internal Position Density Map

Once the internal positions of individuals are known, an internal density map can be created. Note that the size of the density map bins,  $w_{\text{bin}}$ , is a significant parameter in the calculation of entropy. In this, a too large bin will mask the very information that entropy is aiming to extract, while a too small bin will be prone to noise.

In addition to the above, entropy normalisation under the concept of specific entropy needs to be computed as follows:

### 2.2.14 Normalising Entropy

As well as the level of disorder in the crowd, the value of the crowd entropy depends on:

1. The number of individuals in the crowd
2. The extent of the crowd spatial area

### 2.2.15 Specific Entropy

Specific entropy is the entropy per unit of mass. Let each individual to have a unit of mass; the specific entropy,  $H_k$ , will be the entropy of one individual in this crowd:

$$H_k = H(X) \quad (13)$$

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

### 2.2.16 Specific Entropy per Unit of Area

Entropy is maximised if  $\mathcal{P}_X$  is uniform:

$$H(X) \leq \log |\mathcal{L}_X| \text{ with equality iff } \forall i \in \{1, \dots, N_l\} p_i = \frac{1}{|\mathcal{L}_X|} = \frac{1}{N_l}$$

It can be seen that the maximum value of entropy increases with the increase in the number of spatial bins,  $N_l$ . We borrow a concept from information theory called *redundancy*. Redundancy is a measure for 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:

$$R(X) = 1 - \frac{H(X)}{\log |\mathcal{A}_X|} \quad (14)$$

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

$$h_k = \frac{H_k}{\log N_l} \quad (15)$$

As noted, entropy is a measure of disorder, while structure can be described as a measure of order. For a normalised entropy in the range of  $[0, 1]$ , structure and entropy are complementary and add up to one. Let  $s_k$  be the normalised structure:

$$s_k = 1 - h_k \quad (16)$$

## 3 Experimental Results

Three crowd examples have been used in the experiments [16]. Experiment A shows a crowd of pedestrians climbing down a staircase. This motion of crowd is clearly unidirectional. This example depicts an indoor scene with artificial lighting, and the crowd is viewed from an oblique-frontal view. Similar crowds may be observed at a metro station or a stadium. Figure 9 shows one frame example of this crowd. This figure also shows three calibration planes. In this, the orange plane is the reference plane drawn manually. The blue plane and the yellow plane are the ground-level and the head-level planes, respectively, projected back to the image plane after calibration. The red circles show the position of the pedestrians' heads on the head-level plane. For this experiment, people's heads are labelled manually. Figure 10 shows the second crowd (Experiment B) which focuses on pedestrians on an escalator which is on the left-hand side of the same video footage. Here the pedestrians are standing still while the escalator carries them upwards. Finally, Fig. 11 (Experiment C) shows a larger crowd of people in an open indoor space with



**Fig. 9** Crowd on stairs (Experiment A)



**Fig. 10** Crowd on an escalator (Experiment B)

many pedestrians moving in different directions. This type of crowds may be found within airports or shopping malls and so forth. Following from the examples in Fig. 6, it is expected that (i) the crowd in Experiment B (Fig. 10) has the largest structure, since people are standing still; (ii) the crowd in Experiment A (Fig. 9) has a smaller structure than the crowd in Experiment B, but still larger than the crowd in Experiment C (Fig.11); and (iii) the smallest structure is envisaged for crowd in Experiment C.

Figure 12 shows the overall structure results from experiments A, B and C. The experiments were carried out for varying time window sizes ( $w_{tw}$ ) and spatial bin widths ( $w_{bin}$ ). Figure 12a shows the results, where a time window size of 5 s is used. In this, the values of structure are as expected:

$$s_k (X_{exp_B}) > s_k (X_{exp_A}) > s_k (X_{exp_C})$$

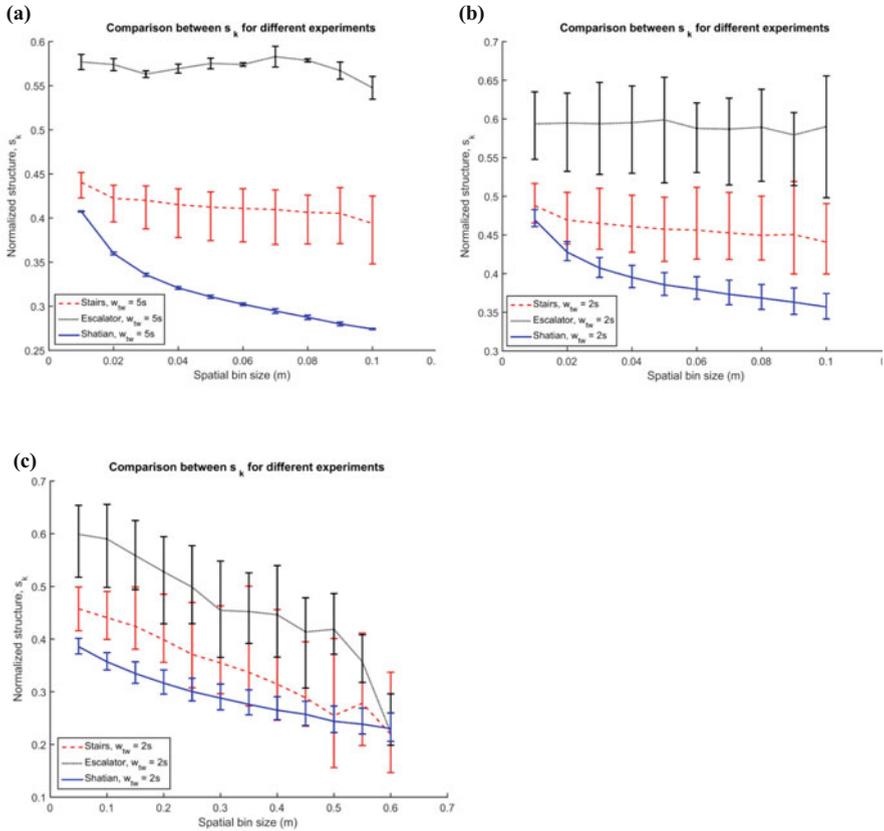


**Fig. 11** Crowd in an open space (Experiment C)

Figure 12b shows the structure values for the same range of spatial bins, but the time window size has been reduced to 2 s. One can see that the order of structure values is still as expected. Nevertheless, while the separation between the various crowds is mostly achieved, the uncertainty on the value of structure increased considerably. Figure 12b, c also demonstrates the effects of spatial bin-size variations. The spatial bins in the range of  $0.01 \text{ m} \leq w_{\text{bin}} \leq 0.6 \text{ m}$  with a time window size of 2 s are investigated in these two graphs. One notes that the smallest bin size does not offer a good separation between crowds, while at the largest bin size of 0.6 m, all crowds show the same structure values. The best separation is achieved for bin sizes between 0.04 m and 0.1 m. Although as mentioned a time window of 5 s offers a much better separation, 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's field of view. As a consequence, the results for Experiment B when analysed with a 5-second time window may be considered as less reliable.

## 4 Ongoing Research and Future Perspectives

In our subsequent works, which we conducted recently, we have also looked into other possibilities which may provide an estimate for the structure of the crowd and compare the respective performances of these approaches with our proposed method over a larger set of crowd conditions. One of the methods which we have examined is that of Zhou et al. using the concept of 'crowd collectiveness' [17, 18]. With it we have been able to track individuals and groups in crowd-associated confined spaces



**Fig. 12** Experiments with normalised specific entropy. **(a)** Experiments with a 5-second time windows ( $w_{TW} = 5s$ ). **(b)** Experiments with a 2-second time windows ( $w_{TW} = 2s$ ). **(c)** Experiments with larger spatial bins ( $0.5 m \leq w_{bin} \leq 0.6 m$ )

such as stadium arenas and in context of the event expected activities [19–21]. The unusualness of groups’ behaviour is detected accordingly to provide an operational approach to security practitioners to respond in a scalable way. In this experiment, groups panic at a segment of the stadium arena and run towards the pitch. This is detected critically in time and produces an alert for security to focus on leading such type of distressed crowd to safety, as shown in Fig. 13.

We have further investigated on the actual mechanics of behaviour propagation in a crowd in recent years, particularly on behaviour which originates from a so-called seed which represents an individual or indeed a group of people behaving within the crowd. This is indeed of great importance to understand and capture trends in it so that we could develop a forecasting capability of behaviour. Although this research work is at its early stages while it is being conducted in the most recently launched S4AllCities research project [22, 23], it is showing us some

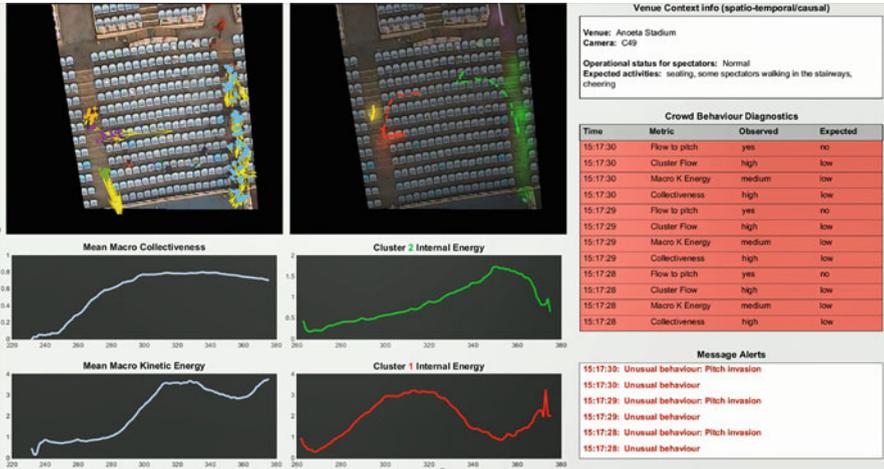


Fig. 13 Group panic behaviour detection in a stadium arena

encouraging findings. Namely, we are obtaining stable tracking of trajectories of individuals as well as groups, where we could derive their parametric functions. These will lead us onto developing data-driven models which will predict trends of such trajectories in future time frames. These of course will be derived with growing computed uncertainties downstream in time and space. We are therefore planning to computationally correcting these trends once new observation measurements are obtained in time in order to reduce and control such uncertainties, leading to a much improved learning process for understanding intentional behaviour in the near future.

**Acknowledgements** The authors are very grateful to all of the eVACUATE research partners, particularly Drs. Zlatko Zlatev, Gianluca Correndo and Lee Middleton, for their collaboration, while we initiated the research work at the University of Southampton. We are also thankful to Professor Nicola Bellomo for his important research discussions on the fundamental modelling of crowd when we initiated this research under the eVACUATE project in 2013. Our research work was partly supported by the European Commission, initially under H2020 Grant Number 313361 in the eVACUATE project (2013–2017) and more recently under H2020 Grant Number 883522 of the ongoing S4AllCities project (2020–2023).

## References

1. Evacuate (2013–2017). A holistic, scenario independent, situation awareness and guidance system for sustaining the active evacuation route for large crowds. H2020 project ID number 313361. <http://www.evacuate.eu/>
2. Parisi, D. R. and Dorso, C. O. (2005). *Microscopic dynamics of pedestrian evacuation*. Physica A, Vol. 354, pp. 606–618.

3. Helbing, D. and Molnar, P. (1995). *Social force model for pedestrian dynamics*. Physical Review E, Vol. 51, 5, pp. 4282–4286
4. Burstedde, C., et al., et al. (2001). *Simulation of pedestrian dynamics using a two-dimensional cellular automaton*. Physica A, Vol. 295, pp. 507–525.
5. Kirchner, A., Nishinari, K. and Schadschneider, A. (2003). *Friction effects and clogging in a cellular automaton model for pedestrian dynamics*. Physical Review E, Vol. 67(5), pp. 056122–056122
6. Heneina, C. M. and White, T. (2007). *Macroscopic effects of microscopic forces between agents in crowd models*. Physica, Vol. 373, pp. 694–712.
7. Guo, R.Y. and Huang, H.J. (2008). *A mobile lattice gas model for simulating pedestrian evacuation*. Physica A, Vol. 387, pp. 580–586.
8. Weng, W. G., et al., et al. (2006). *Cellular automaton simulation of pedestrian counter flow with different walk velocities*. Physical Review E, Vol. 74(3), pp. 036102–036102
9. Helbing, D., Farkas, I. and Vicsek, T. (2000). *Simulating dynamical features of escape panic*. Nature, Vol. 407, pp. 487–490.
10. Mehran, R., Oyama, A. and Shah, M. (2009). *Abnormal Crowd Behavior Detection using Social Force Model*. IEEE Conference on Computer Vision and Pattern recognition (CVPR), 2009, pp. 935–942
11. Henderson, L. F. (1971). *The Statistics of Crowd Fluids*. Nature, Vol. 229, pp. 381 - 383.
12. Helbing, D. (1992). *A Fluid-Dynamic Model for the Movement of Pedestrians*. Complex Systems, Vol. 6, pp. 391–415.
13. Moore, B. E., et al., et al. (2011). *Visual Crowd Surveillance Through a Hydrodynamics Lens*. Communications of the ACM, Vol. 54, pp. 64–73.
14. Tolman, R. C. (1979). *The principles of statistical mechanics*. Courier Corporation.
15. C.E. Shannon. (1948). “A Mathematical Theory of Communication”, *Bell System Technical J.*, vol. 27, pp. 379–423, July and Oct. 1948.
16. Middleton, L., Zlatev, Z., Sabeur, Z.A, Arbab-Zavar, B., Makantasis, K., Karatzalos, K., Christophoulou, I. and Bellomo, N. (2015). D3.4: Multi-scale behaviour recognition technical specification v2.0. Evacuate project deliverable. H2020 project ID number 313361. <http://www.evacuate.eu/>
17. Zhou, B., Tang, X., & Wang, X. (2012). Coherent filtering: Detecting coherent motions from crowd clutters. In *European Conference on Computer Vision*, pp. 857–871. Springer, Berlin, Heidelberg.
18. Zhou, B., Tang, X., & Wang, X. (2013). Measuring crowd collectiveness. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3049–3056.
19. Correndo, G., Arbab-Zavar, B., Zlatev, Z., Sabeur, Z.A. (2015). 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.
20. Sabeur, Z.A., Doulamis, N., Middleton, L., Arbab-Zavar, B., Correndo, G., Amditis, A. (2015). 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.
21. Arbab-Zavar, B. and Sabeur, Z.A. (2020). Multi-scale crowd feature detection using vision sensing and statistical mechanics principles. *Machine Vision and Applications*, 31(4), 1–16.
22. S4AllCities (2020–2023). Safe and Secure Smart Spaces for all Cities H2020 project ID number 883522. <https://www.s4allcities.eu/project>.
23. Sabeur Z., Angelopoulos C.M., Collick L., Chechina N., Cetinkaya D., Bruno A. (2021) Advanced Cyber and Physical Situation Awareness in Urban Smart Spaces. In: Ayaz H., Asgher U., Paletta L. (eds) *Advances in Neuroergonomics and Cognitive Engineering*. AHFE 2021. Lecture Notes in Networks and Systems, vol 259. pp. 428–441. Springer, Cham. [https://doi.org/10.1007/978-3-030-80285-1\\_50](https://doi.org/10.1007/978-3-030-80285-1_50)