ABNORMAL PREDICTION OF DENSE CROWD VIDEOS BY A PURPOSE–DRIVEN LATTICE BOLTZMANN MODEL

YIRAN XUE [a], PENG LIU [a], YE TAO [a,*], XIANGLONG TANG [a]

[a] School of Computer Science and Technology
Harbin Institute of Technology, Mailbox 352, 92 West Dazhi Street, Nan Gang District, Harbin 150001, China
e-mail: `xueyiran@hit.edu.cn`

In the field of intelligent crowd video analysis, the prediction of abnormal events in dense crowds is a well-known and challenging problem. By analysing crowd particle collisions and characteristics of individuals in a crowd to follow the general trend of motion, a purpose-driven lattice Boltzmann model (LBM) is proposed. The collision effect in the proposed method is measured according to the variation in crowd particle numbers in the image nodes; characteristics of the crowd following a general trend are incorporated by adjusting the particle directions. The model predicts dense crowd abnormal events in different intervals through iterations of simultaneous streaming and collision steps. Few initial frames of a video are needed to initialize the proposed model and no training procedure is required. Experimental results show that our purpose-driven LBM performs better than most state-of-the-art methods.

Keywords: video surveillance, crowd analysis, abnormal events, lattice Boltzmann model, purpose-driven strategy.

1. Introduction

During events that attract a large number of enthusiastic participants, such as sports events, festival celebrations, and religious activities, despite the implementation of various precautionary and security measures including the deployment of security personnel and installation of video surveillance, unexpected crowd disasters like deadly stampedes cannot be fully avoided. They seem to occur relatively frequently (Rodriguez *et al.*, 2011). Therefore, in recent years, detection of abnormal crowd events has been among hot topics in the field of computer vision research. Though most detection methods can identify abnormal or emergency crowd situations after a dangerous event has happened, few can predict them in advance.

In this paper, we propose a method of abnormal crowd prediction within some time intervals. The crowd individuals are considered to be made up of a considerable number of crowd particles. The proposed method modifies the lattice Boltzmann model (LBM) by adding a direction selection step to accurately represent the purpose-driven force of crowd particles. Velocity fields of the entire crowd can be predicted iteratively; these

employed the behaviour entropy feature as the criterion to identify abnormal crowd events.

1.1. Related work. In the field of computer vision, crowd analysis models can be divided into three classes: vision-based, physics-inspired, and physics-simulated approaches.

1.1.1. Vision-based approaches. Vision-based approaches use basic techniques in computer vision to analyse crowd events. Some of the important previous works on vision-based approaches are discussed below.

Sparse representation methods (Cong *et al.*, 2013; 2011) use sparse reconstruction cost (SRC) as detection criteria. Usually, the value of SRC in the case of abnormal events is larger than in normal cases. Though these methods are proved as having higher accuracy for anomaly detection, a considerably large number of normal samples are required as training data.

Another common vision-based approach uses the hidden Markov model (HMM), which can account for the inherently dynamic nature of observed features (Wang *et al.*, 2012; Mészáros *et al.*, 2014). In this model, one HMM is set up for all local areas so that it could work only for limited kinds of normal behaviour or specific crowded

---

*Corresponding author

scenes. HMMs are dependent on variant behaviour types, such as running or walking, i.e., when the type is changed, the model needs to be retrained.

Another vision-based model is the dynamic texture one, which is a spatio-temporal generative model for video patterns. It represents video sequences as observations via a linear dynamical system and exhibits spatio-temporal stationary properties (Raghavendra *et al.*, 2011a; Li *et al.*, 2014; Chetverikov and Péteri, 2005; Xu *et al.*, 2011). Recent research works (Chan and Vasconcelos, 2008) have shown that dynamic texture is more suitable for local unusual event detection in crowded scenes than optical flow. However, as in the case of sparse representation methods, this model also requires a large number of samples for normal pattern training.

Yu *et al.* (2016) proposed an effective automatic tracking algorithm of large-scale crowded targets; their technique involves clustering the targets into groups followed by three steps of group refinement: shrinking, growing, and merging. Then, a two-dimensional non-rigid thin-plate splines (TPSs) transform is performed to describe the mapping between the predictions and associated observations within the same group. The method performs well in tracking crowds from low-continuity videos and outperforms many baseline Kalman filters as well as multi-target tracking methods.

Alahi *et al.* (2014) introduced a large-scale dataset of 42 million trajectories extracted from real-world train stations. The behavioural signatures of neighbouring pedestrians were captured using a feature descriptor called the social affinity map (SAM). With an origin and destination (OD) prior, the destination of each individual is predicted.

Another method called particle advection is used to find particle trajectories in crowds. As described by Ali and Shah (2007), trajectories are extracted for crowd flow segmentation and stability analysis. By adding a chaotic model in the particle advection method, anomalies in complicated crowd scenes can be detected and localized (Wu *et al.*, 2010). After risk positions of a scene are located, the crowd behaviour at those positions can be identified using a dynamic system (Solmaz *et al.*, 2012). In the particle advection method, optical flow is used to drive the particle' motion while other attributes of high-density crowds, such as density and flow, are usually ignored. Tracking or trajectory based methods can locate the targets in scenes with a small number of pedestrians (Kowalski *et al.*, 2014; Dębski, 2014; 2016). However, tracking accuracies decrease with increasing crowds.

1.1.2. Physics-inspired approaches.
Several physics-inspired models have been proposed for crowd representation and abnormal event detection.

In the work of Cao *et al.* (2009), an energy model is presented, which estimates crowd kinetic energy and

motion directions based on optical flow techniques. Another energy model, proposed by Xiong *et al.* (2012; 2011), is based on crowd potential and kinetic energies, and is designed to detect two typical abnormal activities: gathering and running. Each individual should be extracted accurately when using these energy models. However, it is difficult to extract when the crowd density is extremely high.

The social force model (SFM), initially proposed by Mehran *et al.* (2009), is based on the assumption that the interaction force is a significant feature for analysing crowd behaviour. Raghavendra *et al.* (2011b; 2011a) introduced particle swarm optimization (PSO) into the SFM to optimize the interaction force. These methods are based on the general idea that people in high-density crowds always try to follow a general trend. They involve minimizing the interaction force using the PSO fitness function, aiming to push particles joining the crowd entity. A velocity-field-based SFM is proposed by Zhao *et al.* (2011). It provides better estimation of interactions using the collision probability in a dynamic crowd. The dense crowds characteristics of following a general trend and colliding with each other were separately considered by different models based on the social force. However, the two characteristics have never been considered simultaneously in one model.

Yuan *et al.* (2015) proposed a method named online anomaly detection for crowd via structure analysis (OADC-SA). The work set forth a structural context descriptor (SCD) to intuitively exploit the context clues between individuals, which originally introduced the potential energy function of particles' interforce. Then, a three-dimensional discrete cosine transform (DCT) is utilized to associate the targets in different frames. This method online detects abnormality through spatial-temporal analysis of SCD variation.

1.1.3. Physics-simulated approaches.
Certain crowd simulation methods exist that simulate crowd escape behaviour on the basis of mass and energy conservation. Lighthill, Witham and Richards (LWR) (Lee *et al.*, 2011) developed the LWR model, which is based on a fluid dynamics continuity equation. Using the finite element method, some researchers have proposed several improved LWR models (Al-nasur, 2006; Jiang *et al.*, 2010) as well; these describe crowds evolution using mass conservation and velocity-density equations. The disadvantages of the LWR and improved LWR models include a large scale of computation required.

The cellular automata model (Ahlquist and Breunig, 2012) considers an individual's socio-psychological and physical forces. In high-density scenarios, the movement of individuals is difficult to estimate because of inappropriate monitoring viewpoints and a high occlusion occurrence rate. Therefore, the cellular automata model

is not quite suitable for the description of a high-density crowd, as it is difficult to describe such crowds using this model.

Finally, it is important to note that the above methods focus on detection of abnormal events and not their prediction.

Johansson *et al.* (2008) proposed an abnormal prediction method based on crowd pressure obtained from the current velocity and density field. It is designed to calculate the degree of abnormality of events. The method provides short-term predictions; therefore, forecasting results are not quite useful.

In this paper, we propose a purpose-driven LBM without a training step. The LBM is used because its properties can reduce the amount of computation required (McNamara and Zanetti, 1988). Further, in our method, the collision process and crowds' activity of following a general trend are represented through a collision step and a purpose-driven streaming step, respectively. The proposed model can predict the scenes' next velocity field and identify an abnormal event with the help of the behaviour entropy model through iterations.

1.2. Similarities and differences between a gas and a dense crowd. Crowd state can roughly be considered a gas. An ideal gas is theoretically composed of many randomly moving point particles that do not interact with one another except when they collide elastically; during these collisions, these particles exchange energy and momentum. In the equilibrium state, particle velocities vary within a small range and the distribution of velocity is nearly unchanged (Mandl, 2008).

To ensure that the dense crowd is an equilibrium system, the distance between adjacent frames' velocity distribution is obtained through K–L divergence (Mathiassen and Skavhaug, 2002). The velocity field of each scene is estimated by the optical flow algorithm (Baker *et al.*, 2011). 38 dense crowd videos of the UCF crowd dataset (Ali and Shah, 2007) are chosen, from which 4 typical scenes and their K–L divergence curves are shown in Fig.1. Most scenes have a flat portion in the K–L distribution curves, which indicates that velocity distribution is nearly unchanged in those parts. This phenomenon shows that an equilibrium state also exists in dense crowds.

The definition of a crowd equilibrium state considered here is as follows: if a crowd velocity distribution tends to remain stable for a time period, the scene state is recognized to be in equilibrium state during that time. Similarities and differences between dense crowds and gases are listed in Table 1.

Gas particles are non-living, and therefore the factors that affect streaming gases primarily include the boundary, pressure, and temperature. Because gas particles have no individual purpose, all the states gas can assume are

considered normal and can be described; for example: turbulent flow, vortex, or laminar flow. On the other hand, humans are intelligent agents. Each individual has the ability to decide on his/her next movement. In addition, the movement is purpose-driven. Therefore, when crowd particle velocity distributions change rapidly, the scene becomes difficult to understand, and abnormal events may occur, especially in overcrowded places, where people are huddling or colliding with each other (Johansson *et al.*, 2008).

1.3. Overview of the proposed method. To predict crowd abnormal events, we modify the LBM by a adding a crowd purpose-driven strategy. The flowchart for our model is shown in Fig. 2.

The model iteratively performs the collision and streaming steps simultaneously. The purpose-driven LBM is added in the model's streaming step by adjusting the directions of particles inside each node. We denote by Δt the time required for one iteration. The proposed model predicts the number of crowd particles in the video during the time intervals $t + m\Delta t$, where $m = 1, \ldots, M$; M is the number of iterations. Then, the velocity field can be estimated by calculating the average velocity of each node, and this velocity field can be used to predict abnormal events using a behaviour entropy model.

1.4. Outline. The remainder of this paper is organized as follows. Section 2 contains the theoretical basis of the LBM. In Section 3, the purpose-driven LBM is introduced. Experiments and comparisons are presented in Section 4. The last section presents concludes the study.

2. Lattice Boltzmann model

The LBM was originally described for gases. Here a gas is divided into many small equal lattices that are usually called nodes (see Fig. 3(a)). The number of particles in a node changes because of collisions and streaming. To reduce the required computation, 9 pre-defined velocity vectors are used to represent all the possible velocities of particles in the node (see Fig. 3(b)).

It is assumed that the particles in the same direction have the same velocity (Wolf-Gladrow, 2000). The velocity of particles in the i-th direction is denoted by $\vec{e}_i$, $i = 0, 1, \ldots, 8$. We have

$$\vec{e}_i = \begin{cases} \vec{e}_0 = (0,0), \\ \vec{e}_{1,5} = (\pm c, 0), \\ \vec{e}_{3,7} = (0, \pm c), \\ \vec{e}_{2,4,6,8} = (\pm c, \pm c), \end{cases} \tag{1}$$

where $c = \Delta x / \Delta t$; Δx is the distance between the adjacent lattices in the horizontal direction, and Δt is the time interval considered for the model (Ahlquist and
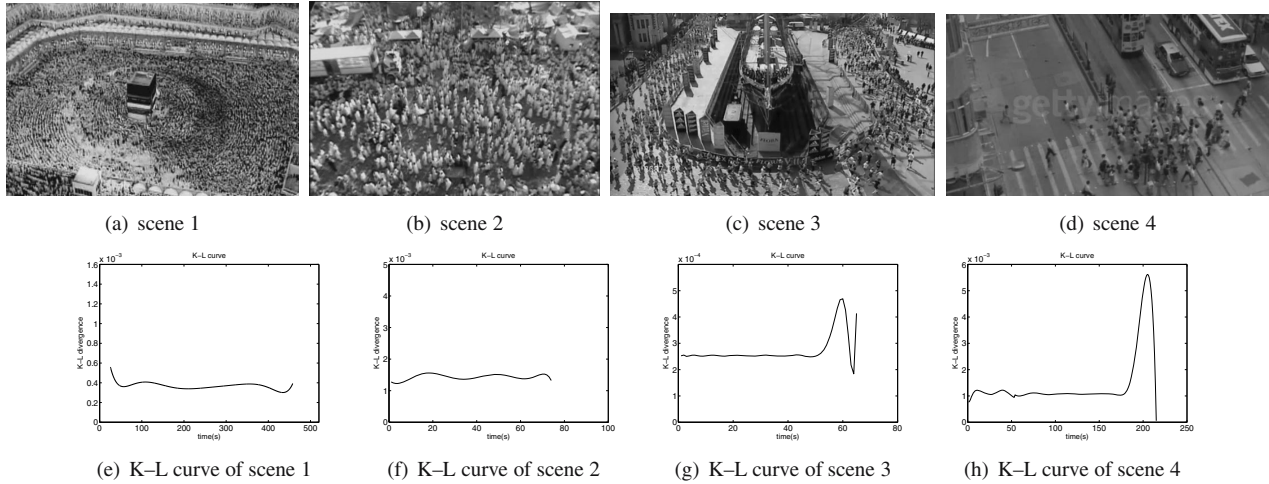
(a) scene 1 (b) scene 2 (c) scene 3 (d) scene 4

(e) K–L curve of scene 1 (f) K–L curve of scene 2 (g) K–L curve of scene 3 (h) K–L curve of scene 4

Fig. 1. K–L curves of four crowd scenes of the UCF crowd dataset.

Table 1. Comparison of dense crowds and gases.

		Dense crowds	Gases
Similarities		Equilibrium	Equilibrium
		Collisions between moving people	Collisions between streaming gas particles
		Moving	Streaming
Differences		Purposed-driven	Not purpose-driven
		Abnormal events can occur	No abnormal events occur
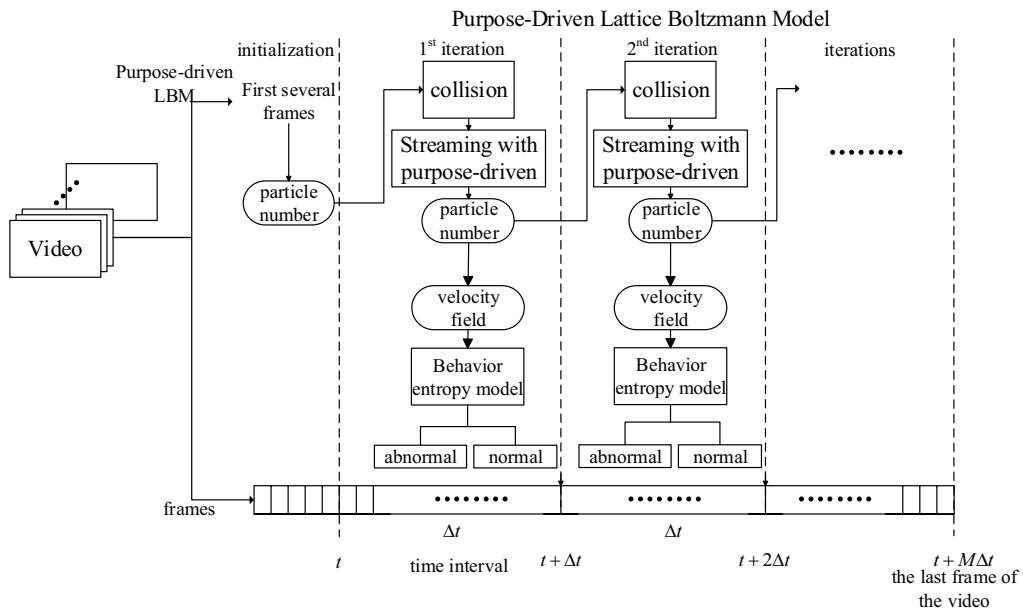
Fig. 2. Flowchart of the proposed method.

(a) fluid area   (b) nine directions inside a node
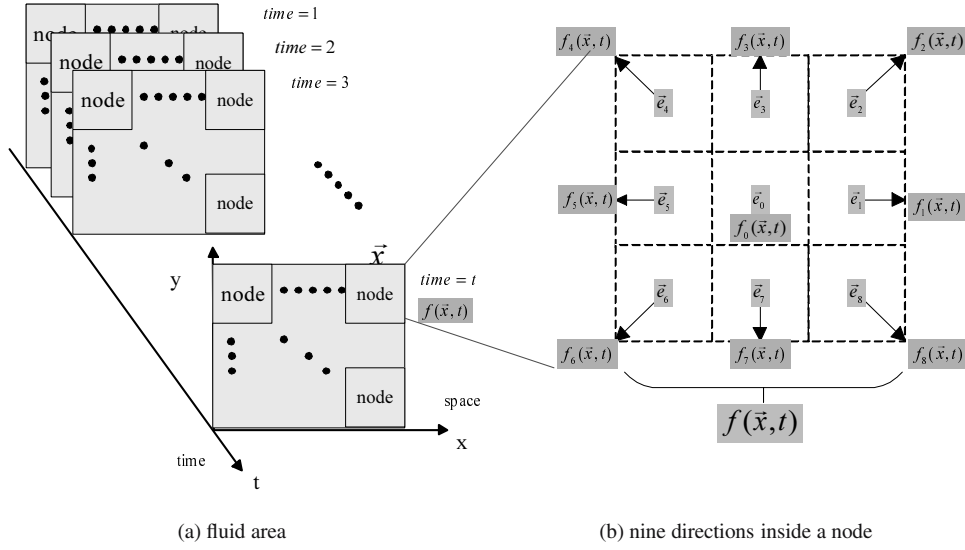
Fig. 3. Fluid area and particle directions of a node.

Breunig, 2012). If the width and height of the node are equal, then $\Delta y = \Delta x$. Here $f(\vec{x}, t)$ denotes the number of particles in the node in position $\vec{x}$ at time t (Fig. 3(a)). Further, inside a node, the number of particles in the i-direction is denoted as $f_i(\vec{x}, t)$ (Fig. 3(b)).

2.1. Collision step. The impact of a collision is measured by the varying numbers of particles within the nodes. Before a collision occurs, the number of particles in node $\vec{x}$ in the i-th direction at time t is determined by $f_i(\vec{x}, t)$. After a collision, the number of particles is determined by $f_i(\vec{x}, t + \Delta t)$, where Δt is the duration of the collision.

The collision can be formulated as follows:

$$f_i(\vec{x}, t + \Delta t) = f_i(\vec{x}, t) + \Omega_i(\vec{x}, t), \qquad (2)$$

where $\Omega_i(\vec{x}, t)$ is the variation in the number of particles caused by the collision. An approximate expression to estimate $\Omega_i(\vec{x}, t)$ was presented by Bhatnagar *et al.* (1954):

$$\Omega_i(\vec{x}, t) = -\frac{1}{\tau}[f_i(\vec{x}, t) - f_i^{eq}(\vec{x}, t)], \qquad (3)$$

where τ is a relaxation parameter, which is determined based on the collision times in the node; $f_i^{eq}(\vec{x}, t)$ is the number of particles in equilibrium state. In general, the Maxwell–Boltzmann distribution (Rowlinson, 2005) is used to represent the equilibrium distribution:

$$f_i^{eq}(\vec{x}, t) = \rho \frac{1}{2\pi RT} \exp\left(-\frac{(\vec{e}_i - \vec{u})^2}{2RT}\right), \qquad (4)$$

where ρ is the total number of particles in a node, which is also called the node density; $\vec{u}$ is the average velocity of

particles in the node. The two variables can be calculated as follows:

$$\rho = \sum_{i=0}^{8} f_i(\vec{x}, t), \quad \vec{u} = \frac{1}{\rho} \sum_{i=0}^{8} \vec{e}_i f_i(\vec{x}, t). \qquad (5)$$

The other two parameters, fluid constant R and temperature T, depend on the property of fluids. Thus, the collision function could be summarized as

$$f_i(\vec{x}, t + \Delta t) = f_i(\vec{x}, t) + \frac{1}{\tau}[f_i^{eq}(\vec{x}, t) - f_i(\vec{x}, t)]. \qquad (6)$$

2.2. Streaming step. In fluid dynamics, it is considered that particles simultaneously collide and stream with each other. After the collision step, the number of particles in node $\vec{x}$ in direction i is given by $f_i(\vec{x}, t + \Delta t)$.

These particles will continue to at move in the i-th direction and enter into node $\vec{x} + \vec{e}_i \Delta t$ in time Δt. Similarly, particles initially located in node $\vec{x} + \vec{e}_i \Delta t$ will continue to move in the i-th direction to node $\vec{x} + 2\vec{e}_i \Delta t$. Therefore, during the streaming step, all particles will be displaced by $\vec{e}_i \Delta t$ if they do not change their direction. This process of displacement is shown in Fig. 4.

As shown in Fig. 4, only particles in node $\vec{x}$ and traveling in the i-th direction can reach node $\vec{x} + \vec{e}_i \Delta t$ in the i-th direction. Therefore, we just need to assign the number of particles in node $\vec{x}$ in the i-th direction to node $\vec{x} + \vec{e}_i \Delta t$ in the i-th direction. Thus, the streaming process could be summarized as

$$f_i(\vec{x} + \vec{e}_i \Delta t, t + \Delta t) \Leftarrow f_i(\vec{x}, t + \Delta t), \qquad (7)$$

where $\Leftarrow$ indicates the update of the number of particles at the location of $\vec{x} + \vec{e}_i \Delta t$.
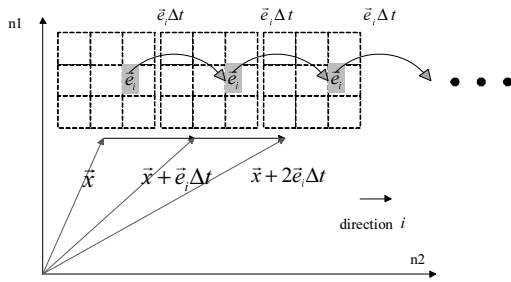
Fig. 4. LBM streaming step.

The LBM is essentially an iteration procedure involving a collision step and a streaming step. Through an expansion, the LBM model is proved to be a convergence system (Mandl, 2008).

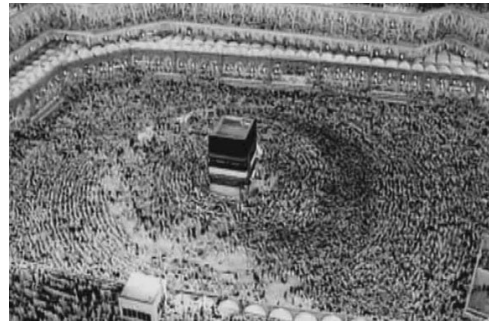## 3. Purpose-driven lattice Boltzmann model

As previously mentioned, the similarities and differences between dense crowds and gases are listed in Table 1. In order to use the LBM to evaluate dense crowds and determine state variations, the model is modified by adding a purpose-driven strategy. With the addition of this strategy, abnormal events can be predicted. In this paper, a model is proposed for crowd abnormal predictions assuming that the moving crowd is a type of fluid.

First, an image or a scene from a crowd video is divided into small lattices, with a size of $k \times k$; these lattices are also called nodes (see also Fig. 3). Then, the velocity field of the image is obtained using the optical flow. Similarly to the original LBM model, 9 velocity vectors with fixed directions are used to represent all the possible velocities of crowd particles in these nodes. Pixels with non-zero velocity are considered crowd particles. In addition, pixels with zero velocity are regarded as the scene background.
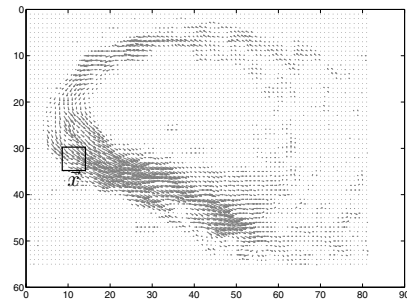
3.1. Model initialization.
The input for the model is the number of particles in each node. In this paper, $f_i(\vec{x}, t)$ is used to represent the input of the number of particles of node $\vec{x}$ in the i-th direction. It is obtained as follows.

A frame is illustrated in Fig. 5(a); the velocity field of this frame is extracted using the optical flow method (Baker *et al.*, 2011) (see Fig. 5(b)). A node is a small region of the velocity field (Fig. 5(c)). In the node, the direction and amplitude of the particle velocity are denoted as $\theta(x, y)$ and $r(x, y)$, respectively (Fig. 5(c)), where (x, y) is the position of a particle in the node, $0 < x < k$ and $0 < y < k$.
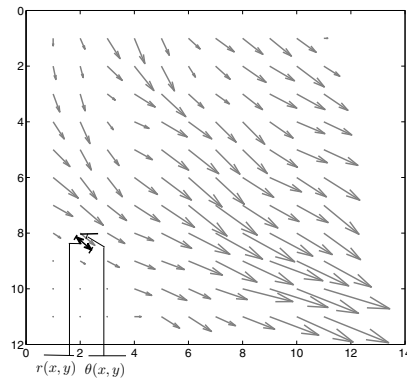
The directions of particles are normalized to one of 9

(a) frame of a crowd video sequence

(b) velocity field of the frame

(c) node in position $\vec{x}$

(d) number of particles in each direction of the node

Fig. 5. Initialization progress.

pre-defined directions according to

$$d(x,y) = \begin{cases} \text{round}\left(\frac{8\theta(x,y)}{2\pi}\right) \bmod 8 + 1 \\ \qquad \text{if} \quad r(x,y) \geq \beta, \\ 0 \qquad \text{if} \quad 0 < r(x,y) < \beta, \end{cases} \tag{8}$$

where $d(x,y)$ changes the direction of the particle into an integer in the range of $(0,8)$; β is the threshold that is used to determine whether a particle is regarded as stationary.

The nodes' initial number of particles in the i-th direction, $f_i(\vec{x},t)$, is obtained by calculating the number of particles in the i-th direction using (Fig. 5(d)):

$$f_i(\vec{x},t) = \text{card}(\{(x,y)|d(x,y)=i, x,y \leq k\}), \tag{9}$$

where $i = 0,1,\ldots,8$.

3.2. Crowd collision step. The crowd lattice Boltzmann model carries on the collision and streaming steps as a gas Boltzmann model.

In Eqn. (3), to estimate $\Omega_i(\vec{x},t)$, the number of particles in a node in an equilibrium state, i.e., $f_i^{eq}(\vec{x},t)$, is required. In Eqn. (4), R and T can be estimated using the fluid internal energy formula, which is

$$\frac{1}{2}\rho\vec{u}^2 = \frac{3}{2}RT, \tag{10}$$

where ρ is the fluid the particle density and $\vec{u}$ is particle average velocity. Because we have assumed that crowd particles satisfy the property of fluid in the given circumstances, we obtain

$$f_i^{eq}(\vec{x},t) = \frac{3}{2\pi\vec{u}^2}\exp\left(-\frac{3(\vec{e}_i-\vec{u})^2}{2\rho\vec{u}^2}\right). \tag{11}$$

3.3. Crowd streaming step. People try to choose the least-effort route to reach their goals when they walk in a crowd. In general, it takes less effort for people to follow the general trend of movement than to push their own way through a dense crowd (Silveira Jacques Jr. *et al.*, 2010; Still, 2000). In this work, the least-effort route is selected by a purpose-driven strategy.

Route selection involves the following two steps:
Step (i). Finding the main direction of each node.
The main direction of a node, denoted by α, is the direction that contains the largest number of particles. For each node, α can be calculated using

$$\alpha = \max_i\{f_i(\vec{x},t+\Delta t)|i=0,1,\ldots,8\}. \tag{12}$$

Step (ii). Adjusting the directions of particles that are not in the main direction or are stationary.
In our model, the particles that move with a uniform acceleration in each stream (see Fig. 6) have their offset updated by $\frac{1}{2}(\vec{e}_i+\vec{e}_\alpha)\Delta t$. Here, Δt is the time required
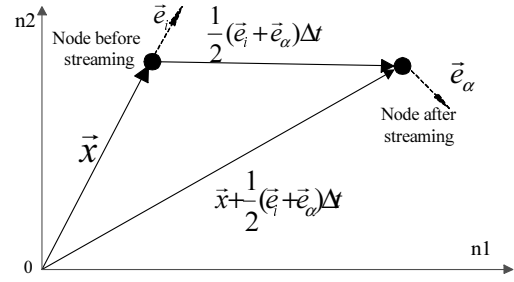

Fig. 6. Streaming step of a particle.

for each streaming. This time duration is equal to that of a collision, because the two steps occur simultaneously.

Before the stream stage, the number of particles in node $\vec{x}$ travelling in direction i is given by $f_i(\vec{x},t+\Delta t)$; however, during this stage, these particles will reach node $\vec{x}+\frac{1}{2}(\vec{e}_i+\vec{e}_\alpha)\Delta t$ in direction α. Therefore, the number of particles in a node is updated as follows:

$$f'_\alpha(\vec{x}+\frac{1}{2}(\vec{e}_i+\vec{e}_\alpha)\Delta t, t+\Delta t)$$
$$= f_i(\vec{x},t+\Delta t) + f_\alpha(\vec{x}+\frac{1}{2}(\vec{e}_i+\vec{e}_\alpha)\Delta t, t+\Delta t). \tag{13}$$

The algorithm of a purpose-driven LBM is summarized as Algorithm 1. At the end of each iteration, the algorithm estimates the number of particles in each node for each direction, such as $f'_\alpha(\vec{x},t+\Delta t)$. Then the corresponding velocity field, obtained using Eqn. (5), is provided as input for the behaviour entropy model to determine whether an abnormal event will occur.

3.4. Behaviour entropy model. It has been observed that abnormal situations in high-density scenarios are often caused by local events. In this work, the behaviour entropy model (BEM) used to identify local anomalous events. Behaviour entropy (BE) of node $\vec{x}$ is defined as

$$BE_{\vec{x}} = -P_{\vec{x}}\log_2 P_{\vec{x}}, \tag{14}$$

where $P_{\vec{x}}$ is a node's probability of keeping its former state, and can be calculated using the following equation:

$$P_{\vec{x}} = e^{-1} + \eta \cdot (1-e^{-1}), \quad 0 \leq \eta \leq 1, \tag{15}$$

where η is defined as

$$\eta = \frac{|\sum_{i=1}^{k_i}\vec{u}_i|}{\sum_{i=1}^{k_i}|\vec{u}_i|}, \quad |\vec{u}_i| \neq 0; \tag{16}$$

η describes the velocity factor among nodes in a small neighbourhood around node $\vec{x}$ and k_i is the number of

Algorithm 1. Purpose-driven lattice Boltzmann model.

Input: $f_i(\vec{x}, t)$ (obtained by Eqns. (8) and (9)), the number of nodes N

Output: $f'_\alpha(\vec{x}, t + \Delta t)$ of each iteration

  for iteration **do**

    Collision step:

    while $j = 1$ to N **do**

      estimate the density and the average velocity of the node: Eqn. (5)

      while $i = 0$ to 8 **do**

        estimate the equilibrium state: Eqn. (4)

        obtain the number of particles in the node's nine directions after collision: Eqn. (6)

      end while

    end while

    Streaming step:

    while $j = 1$ to N **do**

      find the main direction of the node: Eqn. (12)

      while $i = 0$ to 8 **do**

        obtain the number of particles in the node's nine directions after streaming: Eqn. (13)

      end while

    end while

    value assignment for the next iteration: $f_i(\vec{x}, t)^{[n+1]} = f'_\alpha(\vec{x}, t + \Delta t)^{[n]}$, n is the label of the current iteration

  end for

---

nodes in this neighbourhood. The value of η is equal to 1 if all nodes have the same velocity, in which case the value of $P_{\vec{x}}$ is equal to 1. When all nodes have equivalent values in any two opposite directions, then η equals zero and $P_{\vec{x}}$ becomes e^{-1}; the nodes' BE attains the largest value when this takes place. In fact, η represents velocity consistency of the scope.

The smaller the velocity consistence in a node, the larger the behaviour entropy. Abnormal events, such as gathering, running, and regressive walking, usually contain regions with small velocity consistence; therefore, the BE will rise when these events occur. Based on a BE threshold, we can consider nodes to be normal or abnormal.

4. Experiments and analysis

We apply our purpose-driven LBM to one public dataset: the UCF crowd dataset (Ali and Shah, 2007). It is collected from the web (Getty Images, BBC Motion Gallery, YouTube, Thought Equity) and primarily designed for crowd behaviour recognition. In this work, we obtain ground truth labels manually. The dataset has many dense crowd samples, which are appropriate for testing the proposed model.

To verify the methods' performance for abnormal crowd event prediction, first we conducted accuracy testing for velocity fields. Then, abnormal prediction experiments were carried out. The number of input frames is discussed in Section 4.3.

Experiments were conducted on a computer with a 2 GB RAM and 2.6 GHz CPU. The average computation time was 7.8×10^{-3} s/iteration.

4.1. Predicted and real values for the velocity field.
Experiments using the LBM (Ahlquist and Breunig, 2012) and the purpose-driven LBM are conducted for each of five typical dense crowd scenes: blocking, lane, bottleneck, ring/arch, and fountainhead. These typical dense crowd scenes were proposed by Solmaz *et al.* (2012). Figure 7 shows the corresponding crowd video frame images chosen from the UCF dataset. Figure 8 shows the velocity fields obtained using the optical flow method. The velocity fields predicted by the purpose-driven LBM are shown in Fig. 9 and those of the LBM in Fig. 10. As presented in Figs. 7–10, the velocity fields predicted by the purpose-driven LBM are closer to the original velocity fields, whereas those predicted by the LBM tend to distribute crowd particles in a relatively larger area.

To measure the similarity between the velocity fields predicted by our purpose-driven LBM and the original velocity of these videos, structural similarity index measurement (SSIM), which is often used to measure the similarity of video frames (Wang *et al.*, 2004), was adopted. If two frames are exactly similar, then the value of SSIM is 1. The results obtained using the LBM were considered a basis for comparison showing the better results obtained using our method. In this work, each predicted velocity field is made up of two component matrices: a horizontal component matrix u and a vertical component matrix v. The structural similarity indices of the two matrices in each scene are presented in Figs. 11 and 12, respectively.

As shown in the figures, the structural similarity index curve for the purpose-driven LBM is always closer to 1 than that of the LBM. This indicates that the velocity field of the purpose-driven LBM is closer to the ground truth. The purpose-driven model tends to guide more crowd particles following the general trend of the scene. Moreover, its streaming step could effectively avoid the diffusibility of the LBM.

The fact that the lines in Figs. 11 and 12 tend to be stable with time indicates that the purpose-driven LBM maintains the convergence of the LBM perfectly.

4.2. Abnormal event prediction.
Several selected state-of-the-art methods were involved in the following contrast experiments. However, due to the lack of prediction methods at present, most of the contrast
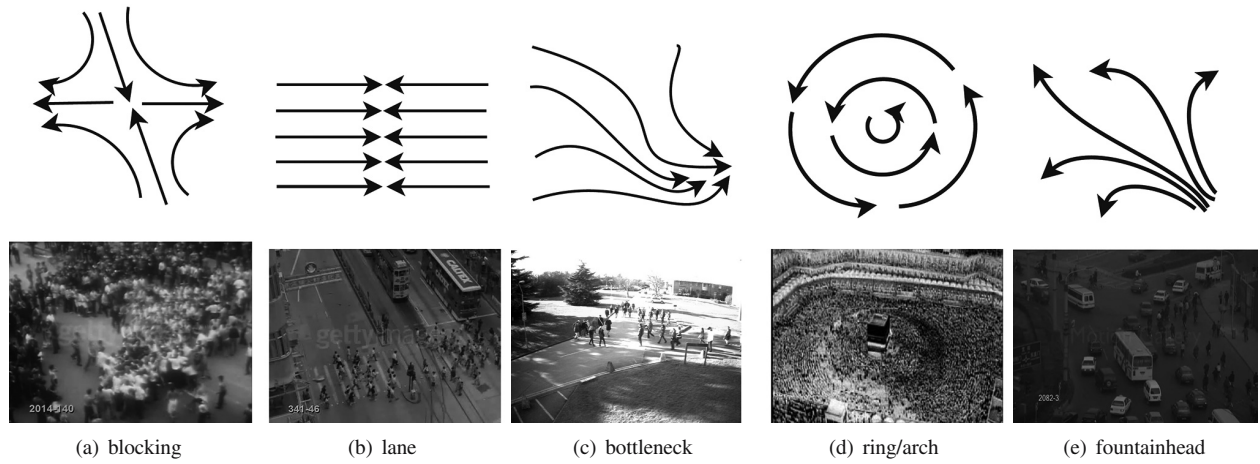
(a) blocking (b) lane (c) bottleneck (d) ring/arch (e) fountainhead

Fig. 7. Five typical high-density crowd scenes.

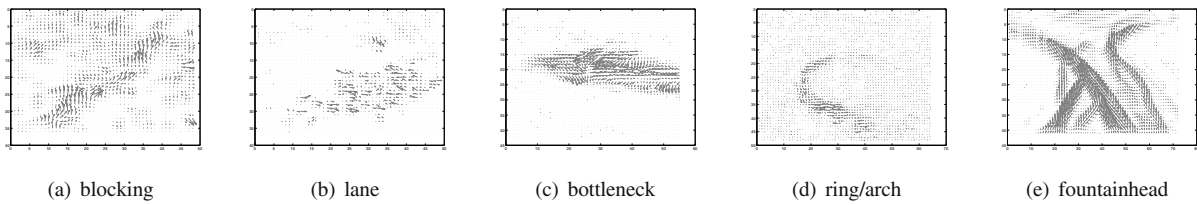(a) blocking (b) lane (c) bottleneck (d) ring/arch (e) fountainhead

Fig. 8. Velocity fields of original video sequences: $t = 20$ s (a), $t = 30$ s (b), $t = 200$ s (c), $t = 90$ s (d), $t = 180$ s (e).

(a) blocking (b) lane (c) bottleneck (d) ring/arch (e) fountainhead
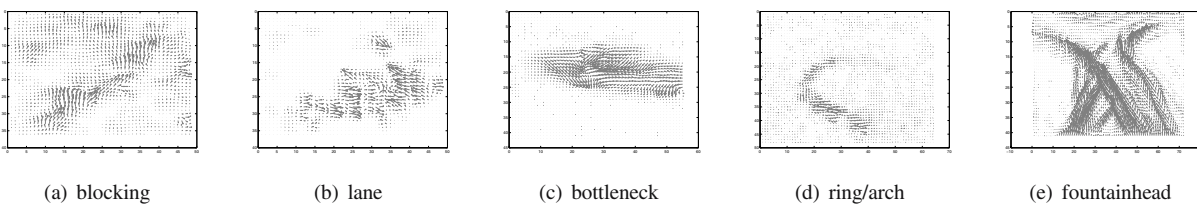
Fig. 9. Velocity fields obtained by the purpose-driven LBM: $t = 20$ s (a), $t = 30$ s (b), $t = 200$s (c), $t = 90$ s (d), $t = 180$ s (e).

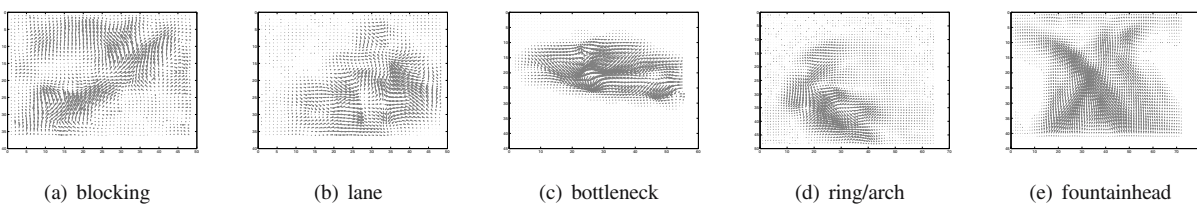(a) blocking (b) lane (c) bottleneck (d) ring/arch (e) fountainhead

Fig. 10. Velocity fields obtained by the LBM: $t = 20$ s (a), $t = 30$ s (b), $t = 200$ s (c), $t = 90$s (d), $t = 180$ s (e).

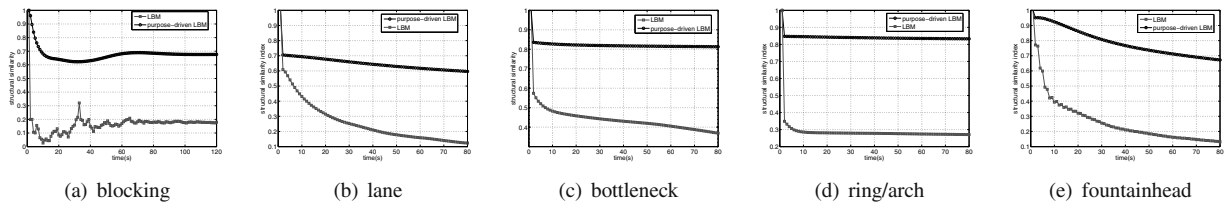Fig. 11. Structural similarity index curves of the u component in predicted velocity fields.
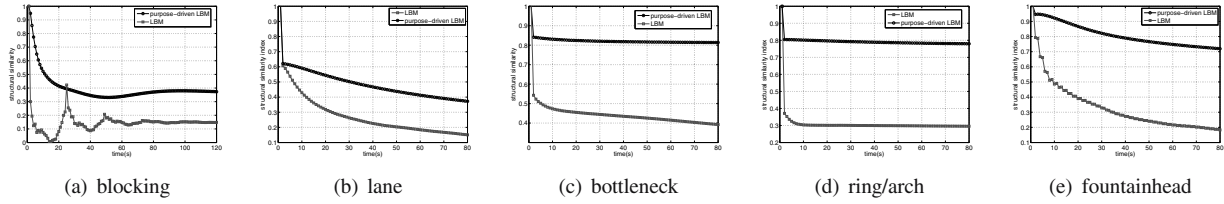
(a) blocking (b) lane (c) bottleneck (d) ring/arch (e) fountainhead

Fig. 12. Structural similarity index curves of the v component in predicted velocity fields.

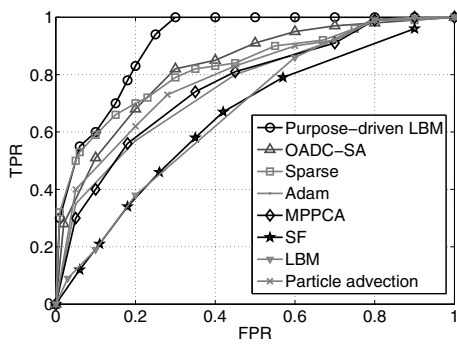(a) blocking (b) lane (c) bottleneck (d) ring/arch (e) fountainhead

Fig. 13. ROC curves of different methods in abnormal event detection.

methods involved are for detecting abnormal events.

We compare the purpose-driven LBM with the sparse (Cong *et al*., 2013), MPPCA (Wu *et al*., 2010) and Adam (Adam *et al*., 2008) approaches, the social force (SF) (Raghavendra *et al*., 2011b), the particle advection method (Ali and Shah, 2007), the LBM (Wolf-Gladrow, 2000), and OADC-SA (Yuan *et al*., 2015).

Figure 13 shows the ROC curves of the methods. As can be seen, the ROC curve of the purpose-driven LBM is higher than that of the others.

Table 2 lists the area under the ROC curve (AUC), the equal error rate (EER), and the rate of detection (RD) for all the evaluated methods. In general, the larger the AUC of a method, the higher its accuracy in detecting abnormal events. The AUC of the purpose-driven LBM is also higher than that of the others.

The sparse method (Cong *et al*., 2013) requires many normal crowd samples to train a dictionary. However, some scenes of the dataset do not contain sufficient normal samples. Thus, with an incomplete dictionary, the sparse

method cannot perform well.

Further, the particle advection method (Ali and Shah, 2007) does not perform well, either. Although optical flow fields preserve particle velocity, the method considers the particle movement directly without any further processing. Therefore, it cannot characterize the physical properties of a crowd, such as following a general trend or adjusting the movement direction momentarily.

The MPPCA (Wu *et al*., 2010) and Adam (Adam *et al*., 2008) methods use dynamic texture to represent video sequences and exhibit spatio-temporal properties. Dynamic texture is suitable for local unusual event detection in crowded scenes, but does not perform well in the case of global events.

OADC-SA (Yuan *et al*., 2015) performs well when the scene is not too crowded. The first step in OADC-SA is extracting each target of the crowd using the state-of-the-art pedestrian detection algorithm proposed by Dollár *et al*. (2014). When the scene is overcrowded like in the case of some scenes in the UCF dataset, the targets cannot be accurately extracted. Thus, the algorithm cannot perform well in overcrowded scenes.

Furthermore, the diffusibility in the case of the LBM (Wolf-Gladrow, 2000) renders it an inefficient method; this problem is overcome by the purpose-driven LBM by adjusting particles into nine main directions. This is one of significant improvements over the LBM that we made in our purpose-driven LBM.

The purpose-driven LBM does not need advance training. After initializing the model with the average velocity field obtained from the first five frames of the test samples, an anomaly object can be located beforehand. These objects are not in normal size and the velocities might lead to a larger behaviour entropy. The purpose-driven strategy, which considers the particle's predisposition of following the general

Table 2. Comparison of the proposed method with state-of-the-art methods for detection of abnormal events in the UCF dataset.

Method	AUC	EER (%)	RD (%)
Purpose-driven LBM	0.96	20	46
OADC-SA (Yuan *et al.*, 2015)	0.87	25	38
Sparse (Cong *et al.*, 2013)	0.88	25	40
MPPCA (Wu *et al.*, 2010)	0.81	40	18
Adam (Adam *et al.*, 2008)	0.80	38	24
SF (Raghavendra *et al.*, 2011b)	0.78	31	21
LBM (Wolf-Gladrow, 2000)	0.75	30	28
Particle advection (Ali and Shah, 2007)	0.85	42	18

trend, makes the particles of the objects region gather together gradually over iterations. Thus, abnormal objects are captured/relatively easier than in the case of other methods.

4.3. Number of input frames. The input of the purpose-driven LBM is the velocity field of the first few frames. It is observed that a small number of frames leads to a high error rate; however, it cannot be concluded that a large number of input frames is better. This work chooses different input frame numbers to find a suitable option. Fifteen dense crowd videos from the UCF dataset are chosen to estimate the error rate when different number of frames are the input into the purpose-driven LBM.

As shown in Fig. 14, the error rate decreases with increasing the number of input frames; however, it stops decreasing when the frame number is larger than five. Therefore, the number of input frames is set to five. The tick marks of the graph's horizontal axis indicate the names of videos of the UCF dataset (see Fig. 14).

4.4. Behaviour entropy curve. Abnormal events are found by inputting predicted velocity fields into the behaviour entropy model. Such events lead to high behaviour entropy values, whereas normal events do not. The experimental results are presented in Fig.15

As shown in the figure, the behaviour entropy curve has three parts. In part A, the predicted velocities' behaviour entropy is considerably small, and therefore the scene is judged to be normal. Part A is when two groups of people are walking across the street. The behaviour entropy of part B grows when the two groups of people are predicted to meet in the middle of the street. Thus, part B is judged to be abnormal; in the real scene, a large number of people gather in a small region of the street, and the crowd disaster is most likely to occur at this moment. In part C, the behaviour entropy again decreases to a normal level; in the real scene, the two groups of people have passed the dangerous area.

5. Conclusion

Our proposed method predicts abnormal events from crowd scenes using a purpose-driven LBM. In our method, the average velocity field of input frames is transformed into a number of particles in all nodes and directions. The direction lattice, which contains the greatest number of particles, is considered the main direction of that node. Adjusting another in terms of the main direction leads to particles moving in particular positions instead of spreading to all directions. The purpose-driven LBM can thus overcome the diffusivity of the original LBM, and provide better predictions for the velocity field in each time interval and better detection accuracy for local abnormal events. The proposed method can be used to accurately predict and detect both global and local level abnormal events. Our experiments to test the detection accuracy of the proposed method prove that it is better than that of many of the state-of-the-art methods that have been previously proposed.

References

Adam, A., Rivlin, E., Shimshoni, I. and Reinitz, D. (2008). Robust real-time unusual event detection using multiple fixed-location monitors, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **30**(3): 555–560.

Ahlquist, J.S. and Breunig, C. (2012). Model-based clustering and typologies in the social sciences, *Political Analysis* **20**(1): 92–112.

Al-nasur, S.J. (2006). *New Models for Crowd Dynamics and Control*, PhD thesis, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Alahi, A., Ramanathan, V. and Fei-Fei, L. (2014). Socially-aware large-scale crowd forecasting, *2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA*, pp. 2211–2218.
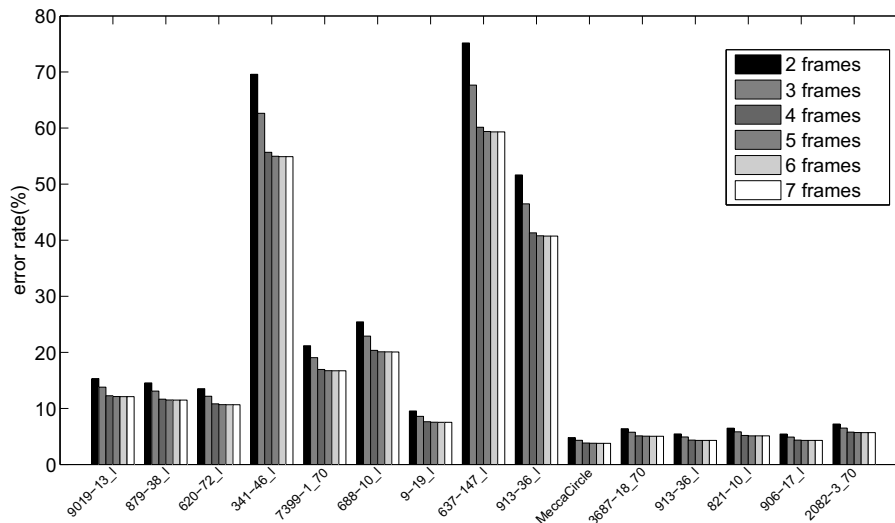
Fig. 14. Error rate of different numbers of input frames. The videos 9019-13_l, 879-38_l, 620-72_l, 341-46_l, 7399-1_70, 688-10_l, 9-19_l, 637-147_l, 913-36_l, MeccaCircle, 3687-18_70, 913-36_l, 821-10_l, 906-17_l, and 2082-3_70 of the UCF dataset are chosen in this experiment.

Ali, S. and Shah, M. (2007). A Lagrangian particle dynamics approach for crowd flow segmentation and stability analysis, *2007 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Minneapolis, MN, USA*, pp. 1–6.

Baker, S., Scharstein, D., Lewis, J.P., Roth, S., Black, M.J. and Szeliski, R. (2011). A database and evaluation methodology for optical flow, *International Journal of Computer Vision* **92**(1): 1–31.

Bhatnagar, P.L., Gross, E.P. and Krook, M. (1954). A model for collision processes in gases. I: Small amplitude processes in charged and neutral one-component systems, *Physical Review* **94**(3): 511–525.

Cao, T., Wu, X., Guo, J., Yu, S. and Xu, Y. (2009). Abnormal crowd motion analysis, *2009 IEEE International Conference on Robotics and Biomimetics (ROBIO), Guilin, China*, pp. 1709–1714.

Chan, A.B. and Vasconcelos, N. (2008). Modeling, clustering, and segmenting video with mixtures of dynamic textures, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **30**(5): 909–926.

Chetverikov, D. and Péteri, R. (2005). A brief survey of dynamic texture description and recognition, *in* M. Kurzyński *et al.* (Eds.), *Computer Recognition Systems*, Advances in Soft Computing, Vol. 30, Springer, Berlin/Heidelberg, pp. 17–26.

Cong, Y., Yuan, J. and Liu, J. (2011). Sparse reconstruction cost for abnormal event detection, *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Colorado Springs, CO, USA*, pp. 3449–3456.

Cong, Y., Yuan, J. and Liu, J. (2013). Abnormal event detection in crowded scenes using sparse representation, *Pattern Recognition* **46**(7): 1851–1864.

Dębski, R. (2014). High-performance simulation-based algorithms for an alpine ski racer's trajectory optimization in heterogeneous computer systems, *International Journal of Applied Mathematics and Computer Science* **24**(3): 551–566, DOI: 10.2478/amcs-2014-0040.

Dębski, R. (2016). An adaptive multi-spline refinement algorithm in simulation based sailboat trajectory optimization using onboard multi-core computer systems, *International Journal of Applied Mathematics and Computer Science* **26**(2): 351–365, DOI: 10.1515/amcs-2016-0025.

Dollár, P., Appel, R., Belongie, S. and Perona, P. (2014). Fast feature pyramids for object detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **36**(8): 1532–1545.

Jiang, Y.-Q., Zhang, P., Wong, S.C. and Liu, R.-X. (2010). A higher-order macroscopic model for pedestrian flows, *Physica A: Statistical Mechanics and Its Applications* **389**(21): 4623–4635.

Johansson, A., Helbing, D., Al-Abideen, H.Z. and Al-Bosta, S. (2008). From crowd dynamics to crowd safety: A video-based analysis, *Advances in Complex Systems* **11**(04): 497–527.

Kowalski, M., Kaczmarek, P., Kabaciński, R., Matuszczak, M., Tranbowicz, K. and Sobkowiak, R. (2014). A simultaneous localization and tracking method for a worm tracking system, *International Journal of Applied Mathematics and Computer Science* **24**(3): 599–609, DOI: 10.2478/amcs-2014-0043.

Lee, B.H., Koo, Y.-H., Oh, J.Y., Cheon, J.S., Tahk, Y.-W. and Sohn, D.-S. (2011). Fuel performance code cosmos for analysis of LWR UO2 and MOX fuel, *Nuclear Engineering and Technology* **43**(6): 499–508.
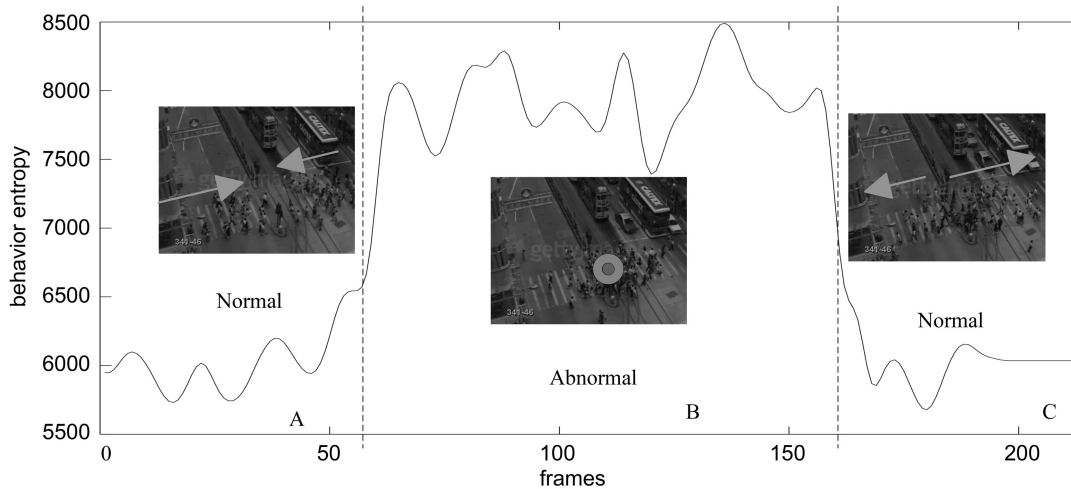
Fig. 15. Behaviour entropy of predicted velocity fields.

Li, W., Mahadevan, V. and Vasconcelos, N. (2014). Anomaly detection and localization in crowded scenes, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **36**(1): 18–32.

Mandl, F. (2008). *Statistical Physics, 2nd Edition*, Manchester Physics, Hoboken, NJ.

Mathiassen, J.R. and Skavhaug, A. (2002). Texture similarity measure using Kullback–Leibler divergence between gamma distributions, *ECCV 2002: 7th European Conference on Computer Vision, Copenhagen, Denmark*, Part III, pp. 133–147.

McNamara, G.R. and Zanetti, G. (1988). Use of the Boltzmann equation to simulate lattice-gas automata, *Physical Review Letters* **61**(20): 2332.

Mehran, R., Oyama, A. and Shah, M. (2009). Abnormal crowd behavior detection using social force model, *2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Miami, FL, USA*, pp. 935–942.

Mészáros, A., Papp, J. and Telek, M. (2014). Fitting traffic traces with discrete canonical phase type distributions and Markov arrival processes, *International Journal of Applied Mathematics and Computer Science* **24**(3): 453–470, DOI: 10.2478/amcs-2014-0034.

Raghavendra, R., Bue, A.D., Cristani, M. and Murino, V. (2011a). Optimizing interaction force for global anomaly detection in crowded scenes, *2011 IEEE International Conference on Computer Vision Workshops, Barcelona, Spain*, pp. 136–143.

Raghavendra, R., Bue, A., Cristani, M. and Murino, V. (2011b). Abnormal crowd behavior detection by social force optimization, *in* A.A. Salah and B. Lepri (Eds.), *Human Behavior Understanding: Second International Workshop, HBU 2011*, Springer, Berlin/Heidelberg, pp. 134–145.

Rodriguez, M., Laptev, I., Sivic, J. and Audibert, J.Y. (2011). Density-aware person detection and tracking in crowds, *2011 International Conference on Computer Vision, Barcelona, Spain*, pp. 2423–2430.

Rowlinson, J.S. (2005). The Maxwell–Boltzmann distribution, *Molecular Physics* **103**(21–23): 2821–2828.

Silveira Jacques Jr., J.C., Raupp Musse, S. and Rosito Jung, C. (2010). Crowd analysis using computer vision techniques, *IEEE Signal Processing Magazine* **27**(5): 66–77.

Solmaz, B., Moore, B.E. and Shah, M. (2012). Identifying behaviors in crowd scenes using stability analysis for dynamical systems, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **34**(10): 2064–2070.

Still, G.K. (2000). *Crowd Dynamics*, PhD thesis, University of Warwick, Coventry.

Wang, B., Ye, M., Li, X., Zhao, F. and Ding, J. (2012). Abnormal crowd behavior detection using high-frequency and spatio-temporal features, *Machine Vision and Applications* **23**(3): 501–511.

Wang, Z., Bovik, A.C., Sheikh, H.R. and Simoncelli, E.P. (2004). Image quality assessment: From error visibility to structural similarity, *IEEE Transactions on Image Processing* **13**(4): 600–612.

Wolf-Gladrow, D. (2000). *Lattice-Gas Cellular Automata and Lattice Boltzmann Models—An Introduction*, Lecture Notes in Mathematics, Vol. 1725, Springer, Berlin.

Wu, S., Moore, B.E. and Shah, M. (2010). Chaotic invariants of lagrangian particle trajectories for anomaly detection in crowded scenes, *2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, CA, USA*, pp. 2054–2060.

Xiong, G., Cheng, J., Wu, X., Chen, Y.-L., Ou, Y. and Xu, Y. (2012). An energy model approach to people counting for abnormal crowd behavior detection, *Neurocomputing* **83**: 121–135.

Xiong, G., Wu, X., Chen, Y.L. and Ou, Y. (2011). Abnormal crowd behavior detection based on the energy model, *2011 IEEE International Conference on Information and Automation (ICIA), Shenzhen, China*, pp. 495–500.

Xu, J., Denman, S., Sridharan, S., Fookes, C. and Rana, R. (2011). Dynamic texture reconstruction from sparse codes for unusual event detection in crowded scenes, *Proceedings of the 2011 Joint ACM Workshop on Modeling and Representing Events, Scottsdale, AZ, USA*, pp. 25–30.

Yu, H., Zhou, Y., Simmons, J., Przybyla, C., Lin, Y., Fan, X., Mi, Y. and Wang, S. (2016). Groupwise tracking of crowded similar-appearance targets from low-continuity image sequences, *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA*, pp. 952–960.

Yuan, Y., Fang, J. and Wang, Q. (2015). Online anomaly detection in crowd scenes via structure analysis, *IEEE Transactions on Cybernetics* **45**(3): 548–561.

Zhao, J., Xu, Y., Yang, X. and Yan, Q. (2011). Crowd instability analysis using velocity-field based social force model, *2011 IEEE Conference on Visual Communications and Image Processing (VCIP), Tainan, China*, pp. 1–4.

Yiran Xue is a PhD candidate at the School of Computer Science and Technology, Harbin Institute of Technology, China. His research focuses on abnormal movement prediction in dense crowd scenes.

Peng Liu is an associate professor at the School of Computer Science and Technology, Harbin Institute of Technology, China. He obtained his doctoral degree in microelectronics and solid state electronics in 2007. His research interests cover image processing, video processing, pattern recognition and design of VLSI circuits.

Ye Tao is a PhD candidate at the School of Computer Science and Technology, Harbin Institute of Technology, China. His research focuses on abnormal motion detection.

Xianglong Tang is a professor at the School of Computer Science and Technology, Harbin Institute of Technology. He received his doctoral degree in computer application technology in 1995. His research interests cover pattern recognition, aerospace image processing, medical image processing and machine learning.