A hybrid social influence model for pedestrian motion segmentation

*Habib Ullah · Mohib Ullah · Muhammad Uzair

Abstract We propose a hybrid social influence model (HSIM) which is a novel and automatic method for pedestrian motion segmentation. One of the major attractions of the HSIM is its capability to handle motion segmentation when the pedestrian flow is randomly distributed. In the proposed HSIM, we first extract the motion information from the input video through particle initialization and optical flow. The particles are then examined to keep only the significant and nonstationary particles. To detect consistent segments, we adopt the communal model (CM) that models the influence of particles on each other. The CM infers influence from uncorrelated behaviors among particles and models the effect that particle interactions have on the spread of social behaviors. Finally, the detected segments are refined to eliminate the effects of oversegmentation. We perform extensive experiments on four benchmark datasets and compare the results with two baseline and four state-of-the-art motion segmentation methods. Our results show that HSIM achieves superior pedestrian motion segmentation and outperforms the compared methods in terms of both Jaccard Similarity Metric (JSM) and F-score.

Keywords Optical flow · Social force model · Motion segmentation · Pedestrian dynamics

*Habib Ullah, Corresponding author,
University of Hail, Hail, Saudi Arabia
E-mail: h.ullah@uoh.edu.sa

Mohib Ullah
Norwegian University of Science and
Technology, Gjovik, Norway
E-mail: mohib.ullah@ntnu.no

Muhammad Uzair
COMSATS Institute of Science and
Technology, Wah Cantt, Pakistan
E-mail: uzair@ciitwah.edu.pk

1 Introduction

Pedestrian flows representing movements of group of individuals are pervasive in many real-world environments. Segmenting motion in the pedestrian flows is an important and challenging computer vision task [2][43][14][39][5] that can effectively decompose the pedestrian motion into meaningful segments. The segments can then be exploited for the detection and recognition of various anomalous events in a video surveillance system.

The distribution of pedestrian represents segments of the pedestrian flows located in different places of the scene. When the density of the pedestrian changes over time, the distribution changes accordingly. Therefore, the distribution and density of pedestrians affect its coherency that represents the interconnection among different segments. A consistent distribution over time presents a coherent motion of pedestrians. The distribution changes with the changing density flowing in several directions. Previous approaches [2][25][40][43] only address the crowed motion segmentation. These approaches are based on the assumption that the degree of movement remains substantially consistent within a global area. This means that considering various densities of crowd flows, the distribution remains the same over time. However, this assumption may not be valid in many real-world pedestrian scenarios where pedestrians are randomly distributed over time. For example, random distribution of groups of pedestrians on a crossroad can be observed during office hours. However, the distribution changes in later hours.

To address these challenges of varying pedestrian distributions over time, we propose a novel social influence based HSIM method for pedestrian motion segmentation. The main advantage of our HSIM method is its capability to infer interactions and dynamics for

motion segmentation when the structure of the pedestrian flow is randomly distributed. Our HSIM is based on the observation that the change in the social influence of one segment can affect the motion pattern of another segment. In this way, it models the effect of change in the distribution of the whole system. Each particle represents the position of a pixel. We argue that, if we understand exactly how particles interact; how one particle influences another; and how they are situated relative to one another; then we can model the motion of the randomly distributed pedestrians as a whole.

A pedestrian flow is a collection of spatially proximate individuals who also exhibits various dynamic properties as a whole. A pedestrian flow may have a non-uniform distribution when people are separated into groups with different social relationships and walk in unrestricted areas. Groups are likely to exhibit an increased level of similarity represented by our proposed social influence model. Our model formulates the behavior of the particles instead of individual pedestrians to deal with the situations the particles are confronted with in a social system. Our model exploits social influences for localized constituents in order to estimate recurrent structures in the frames, with the important distinction that such constituents are not expected to fully contain a person. Our model does not treat each person individually. Instead we consider particles and there social influences on the neighboring particles. Our model considers orientation as a prior, meaning that a set of particles are considered, only if their relative offset in terms of direction of motion appears in a predefined range. A set of particles on an individual pedestrian or a group of pedestrians with consistent direction of motion would construct a segment. In this case, our proposed distribution independent social force model (DI-SFM) does not need to distinguish between intra-person and inter-person scenarios. This is why the DI-SFM deals more effectively with these scenarios.

We regard each moving object as part of the pedestrian flow and static objects or static groups of people as a background. In our method, we assume that there is no camera motion. Our proposed HSIM method is illustrated in Fig. 1. Firstly, we initialize the particles on the video frame. The particle initialization involves the identification of the pixel locations to be tracked. A set of particles represent a moving object in the video as can be seen in the third column of Fig. 1. Each object is represented by arbitrary number of particles. The number of particles associated with each object depends on the location of the object with reference to the camera and its overlapping with other objects in the video. To track particles, we explore Lucas-Kanade optical flow [49] to extract the motion information from the video on frame-by-frame basis. Secondly, we propose the distribution independent social force model (DI-SFM). Inspired by Yu et al.'s work [48] we only retain the particles that contribute to pedestrian motion segmentation. It means that the DI-SFM removes the noisy particles associated with non-motion regions. However, different from Yu et al.'s work [48], our DI-SFM also takes into account the random distribution of the pedestrian flow. For this purpose, we exploit the squeezing and spreading of the neighboring particles by modeling personal and repulsive forces as these forces take into account the physiological states of pedestrians naturally. In the third step, we compute the orientation information of the particles to perform pedestrian motion segmentation. For consistent pedestrian motion segmentation, we propose the communal model (CM) inspired by the work of Pan et al.[30]. Finally, to refine the segmentation output, an unsupervised graph-based dissimilarity measure (GBDM) approach [11] driven by low computation complexity is exploited to group regions that are coherent both in appearance and motion. Thus, our overall HSIM framework presents a consistent pedestrian motion segmentation.

The main contributions of this paper are:

1. We propose a novel and efficient algorithm called HSIM for pedestrian motion segmentation. One of the major attractions of our proposed method is its capability to deal with the randomly distributed pedestrian flows. Our HSIM method consists of DI-SFM and CM models. The DI-SFM removes particles related to non-motion regions and the CM segments the pedestrian motion consistently.

2. Our proposed method is driven by low computational overheads compared to the state-of-the-art methods [2][14][39][27][17][37].

3. We extensively evaluate the proposed method on four standard datasets and compared to two baseline and four state-of-the-art methods. Our results show that the proposed method significantly outperforms all the reference methods. Preliminary results of our proposed work on a few video sequences were presented in [42][34] where we only classified the crowd behaviors [42] and tracked the entities [34].

To evaluate our proposed HSIM, we perform extensive experiments on four benchmark datasets and compare the results with two baseline and four state-of-the-art motion segmentation methods including the robust trajectory clustering (RTC) [39], the spatio-temporal segmentation (STS) [27], the Lagrangian particle dynamics (LPD) [2], the multicue-based segmentation (MBS) [14],
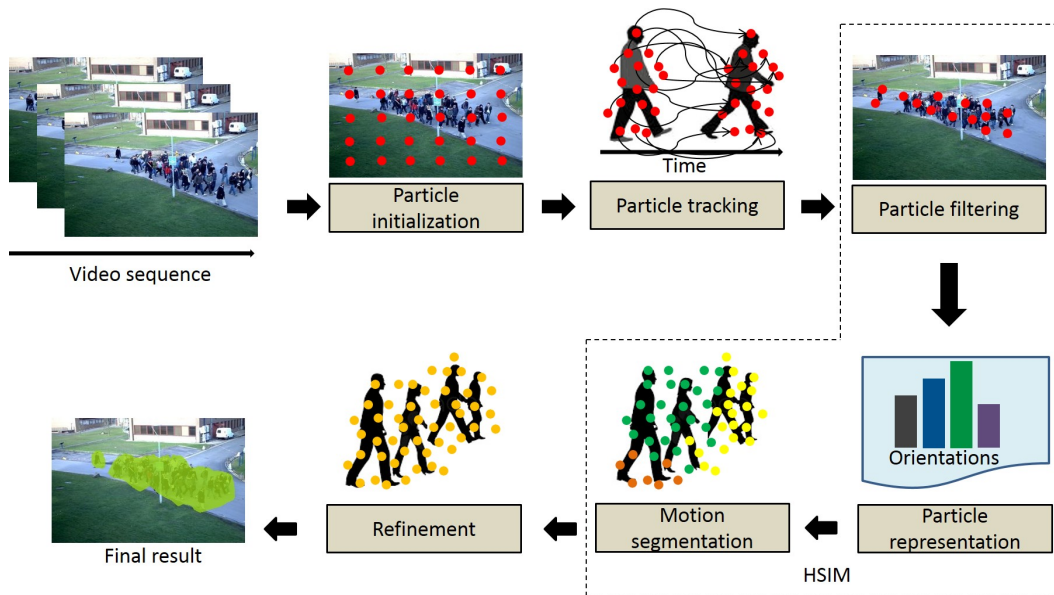
Fig. 1 Proposed HSIM method for pedestrian motion segmentation. Each particle depicts the position of a pixel. For the sake of visualization only a limited set of particles are superimposed on the frame. The particle filtering stage retains the most significant and nonstationary particles associated with the pedestrian flow. These particles along with their representations are used to perform pedestrian motion segmentation in the next stages.

the granular crowd segmentation (GCS) [17], and collective transition (CT) [37]. Our experimental results show that the HSIM achieves superior pedestrian motion segmentation. For quantitative analysis, we compute both the Jaccard Similarity Metric (JSM) [38] and the F-score [17]. Our proposed HSIM outperforms all the reference methods in both qualitative and quantitative evaluations.

The rest of the paper is organized as follows. In Section 2, an overview of related work is presented. The pedestrian motion segmentation is presented in Section 3. Experimental results on the benchmark datasets are shown in Section 4 and the conclusion is presented in Section 5.

2 Related work

The Pedestrian motion segmentation and anomaly detection methods are related with each other. Therefore, we divide them into three broad categories. This categorization is based on the density of flows under consideration. Methods targeting upto five individuals are discussed under the category of low-level density. Methods targeting more than five and less than 15 individuals are grouped under the term mid-level density. Similarly, methods targeting more than 15 individuals are discussed under the category of grand-level density.

In the category of low-level density, Ledig et al. [21] introduce a patch-based segmentation method. They find patch correspondences and the associated patch displacements, which allow the estimation of segmentation. They also assess both the agreement of the segmentation boundary and the conservation of the segmentation shape. Nguyen et al. [29] propose a consensus model for motion segmentation in dynamic scenes. For this purpose, they combine several unsupervised methods to deal with the label correspondence problem. Bosch et al. [6] perform video object segmentation using both color and motion models. These models are combined into a cost function that encodes the likelihood of a pixel to either belonging to the object or not. Fu et al. [13] propose the contour-guided color palette (CCP) to perform preliminary segmentation. The CCP integrates contour and color cues. Final segmentation results are obtained by using a combination of techniques such as leakage avoidance, fake boundary removal, and merging small regions. Seyedhosseini et al. [35] propose contextual hierarchical model (CHM), which learns contextual information in a hierarchical framework for image segmentation. For this purpose, at each level of the hierarchy, a classifier is trained based on downsampled input images and outputs of previous levels. The proposed CHM then incorporates the resulting multi-resolution contextual information into a classifier to segment the input image at original resolution. Luo et al. [23] propose an adaptive manifold denoising approach for both

rigid and non-rigid objects segmentation. An adaptive kernel space is used in which two feature trajectories are mapped into the same point if they belong to the same rigid object. Then an embedded manifold denoising approach is employed with the adaptive kernel to segment the motion of rigid and non-rigid objects. Narayana et al. [28] exploit the property that under camera translation, optical flow orientations are independent of object depth. Then a probabilistic model is proposed that estimates the number of observed independent motions and results in a labeling that is consistent with real-world motion in the scene.

In mid-level density category, Milan et al. [26] obtain video segmentation by exploiting low level image information and associates every super pixel to a specific target or classifies it as background. Hou et al. [14] propose a multicue based segmentation (MBS) approach to formulate the segmentation process by grouping the feature points with a human model. The approach uses a block-based implicit shape model to collect patches from a human being and assess the possibility of their occurrence in each part of a body. The combination of appearance cues with coherent motion of the feature points in each individual is explored. Rodger et al. [33] present a segmentation technique based on initial super pixels representation. Tracks of people are used as Bayesian prior by assuming that human activity tends to be quite structured. Mukherjee et al. [27] use a Gaussian mixture model (GMM) to segment each frame while a recursive filtering updates the parameters of the GMM. The method propagates Gaussian clusters through each new frame, update the variance recursively, and create or remove clusters as necessary. Huerta et al. [15] fuse the knowledge from the color, edge and intensity to perform object segmentation. Li et al. [22] use a spatio-temporal model (STM) to analyze motion trajectories of multiple objects to extract consistent segments. Khan et al. [16] identify locations where pedestrians appear as sources and the locations where they disappear as sinks. Chan et al. [9] use a spatio-temporal generative model namely the mixtures of dynamic textures (MDT) for segmenting the videos.

In the grand-level density category for motion segmentation, Biswas et al. [5] perform motion segmentation using super-pixels in H.264 compressed videos. Different segments are determined by measuring the confidence score of the multi-scale super-pixel boundaries. Ali et al. [2] propose a Lagrangian particle dynamics (LPD) approach to segment the flow. For this purpose, the finite time lyapunov exponent [36] is exploited to define the boundaries of different segments. Kok et al. [17] propose the granular computing to aggregate similar pixels into atomic structure granules for

motion segmentation. The structure granules are used to isolate the motion and background regions. Wu et al. [45] introduce local-translation domain model (LTD) that explores the scattered motion field (SMF) for motion segmentation. According to the SMF, the Gâteaux derivative is used to segment the boundaries of homogeneous flows. Kumar et al. [18] track objects in crowd using an objective function which considers both colour and texture property of videos. Shao et al. [37] propose a collective transition (CT) prior for group segregation in crowded scenes. Wang et al. [43] segment the motion using the thermal diffusion process (TDP). The TDP investigates motion correlation among particles to identify different flows. Sharma et al. [38] segment motion by extracting the temporal evolution of dense trajectories. The trajectories are grouped based on the information of shape, location, and density.

Anomalous events can be detected after segmenting the scenes into meaningful parts. It is worth noticing that anomaly detection methods are mostly categorized in grand-density flow analysis. Leach et al. [20] detect subtle anomalies based on both social and scene contexts. Carrera et al. [7] detect anomalies using the sparse representation of image patches. Xu et al. [47] detect anomalous events based on a hierarchical activity-pattern discovery framework. The proposed framework is an unsupervised method for constructing normal activity patterns at different levels. Cheng et al. [10] propose a one-class support vector machine with Bayesian derivation to detect unusual events. For this purpose, an event representation scheme is introduced that refers to a time series of spatial windows in proximity. This scheme captures the unstructured property of an event. Wu et al. [46] propose a Bayesian framework for escape panic detection by modeling the motion in both the presence and absence of escape events. Biswas et al. [4] propose a real time anomaly detection for H.264 compressed video streams. For this purpose, preencoded motion vectors (MVs) are utilized assuming that MVs have distinct characteristics during anomaly than usual. Thida et al. [41] detect and localize abnormal events based on a spatio-temporal Laplacian eigen map (STLE). The STLE is learned from the spatial and temporal variations of local motions. Baxter et al. [3] propose behavior-based people tracking approach using priors mediated by head-pose. Leach et al. [19] detect social groups in surveillance videos. For this purpose, they combine gazing direction of pedestrians with visual attention feature. Our proposed HSIM method is distribution independent. Therefore, it is capable to cope with various densities of pedestrians. Moreover, our HSIM is computationally more efficient than the previous methods [39][27][2][14][17][37].

3 Proposed hybrid social influence model (HSIM) for pedestrian motion segmentation

In this section, we describe in detail each stage of our proposed HSIM method for pedestrian motion segmentation. The proposed method consists of four main stages namely motion information extraction, retaining the motile particles, motion segmentation, and finally the refinement of the segmentation map.

3.1 Motion information extraction

We begin by initializing a set of particles over the video frame. The particle initialization represents the localization of pixels positions uniformly spread over the video. We then track these particles using the Lucas-Kanade optical flow [49] technique. Although different optical flow techniques can be used in this stage but we choose the Lucas-Kanade optical flow [49] technique due to its simplicity and low computational burden. Since we are interested only in the particles associated with the moving pedestrians instead of static regions, we propose an approach for retaining the motile particles in the next stage.

3.2 Retaining the motile particles

In our proposed HSIM method, we consider pedestrian flow as a social system where the movement of one particle influences the neighboring particles. Therefore, we model the motion patterns of the pedestrian scene through the density independent social force model (DI-SFM). The DI-SFM retains particles that contribute to the motion segmentation. For this purpose, the DI-SFM removes particles associated to regions not exhibiting significant motion. The main attractions of the DI-SFM are bipartite. Firstly, the DI-SFM models the motion of particles based on social forces which exist in real-world scenarios. Secondly, the DI-SFM takes into account the random distribution of the pedestrian flow. The social forces are a measure representing the motivation of the individual particle to carry out certain movements, and consider the influence of the other particles surrounding it. Rinke et al. [32] describe that the modeling of social forces is based on Newton's second law (NSL) of motion. According to the NSL, the state of an object will change in the presence of an external force. In shared space environments the movements of particles are not regulated by themselves, but is the result of spontaneous interaction among them, who exchange the priority according to social rules such as considering neighboring particles personal spaces. When the

structure of the pedestrian flow is randomly distributed, the movement of one particle influences the neighboring particles. Thus change in the velocities occurs that changes the state of the particles. Therefore, the DI-SFM considers the NSL to model the velocity of each particle q with mass m_q according to Eq. (1),

$$|D_q^t| = m_q \frac{dv_q}{dt} = |\Lambda_q^{t,a}| \tag{1}$$

where $|D_q^t|$ is the magnitude of the DI-SFM. m_q is the mass and v_q is the current velocity of the particle q. $\Lambda_q^{t,a}$ is the acceleration force which has two parts, the personal force $\Lambda_q^{t,p}$ and the repulsive force $\Lambda_q^{t,r}$, as formulated in Eq. (2).

$$\Lambda_q^{t,a} = \Lambda_q^{t,p} + \Lambda_q^{t,r} \tag{2}$$

The acceleration force $\Lambda_q^{t,a}$, the personal force $\Lambda_q^{t,p}$, and the repulsive force $\Lambda_q^{t,r}$ are real-valued numbers and our DI-SFM model treats them independent of the directions. The personal force $\Lambda_q^{t,p}$ represents the attempt of a particle to seek a particular goal and destination. For this purpose, the personal force $\Lambda_q^{t,p}$ takes into account the actual velocity v_q^t and the desired velocity $v_q^{t,d}$ of the particle q. If a particle q is moving, it will walk into the desired direction with a certain desired speed. A deviation of the actual velocity v_q^t from the desired velocity $v_q^{t,d}$ due to the deceleration procedure or avoidance actions leads to a tendency to approach the velocity $(v_q^{t,d} - v_q^t)$ within a certain relaxation time ϵ. Therefore, the personal force $\Lambda_q^{t,p}$ can be formulated in Eq. (3),

$$\Lambda_q^{t,p} = \frac{1}{\epsilon}(v_q^{t,d} - v_q^t) \tag{3}$$

where ϵ is called the relaxation/time parameter. The personal force $\Lambda_q^{t,p}$ is based on the current velocity v_q^t and the desired velocity $v_q^{t,d}$ of a particle q. The current velocity v_q^t represents the actual velocity of the particle which is computed as Euclidean distance between initial and current locations in the consecutive frames at time $t-1$ and t. The desired velocity $v_q^{t,d}$ represents the velocity of the particle under the tempt of the neighboring particles in a consistent situation. Such situation means there is no squeezing and spreading of the neighboring particles. The desired velocity $v_q^{t,d}$ is calculated in a spatial window of size $w \times w$. An example of calculation of these velocities is shown in Fig. 2. In the window, the central particle q considers its neighboring particles c_j in the same video frame for computation. However, when the distribution changes, the velocity of
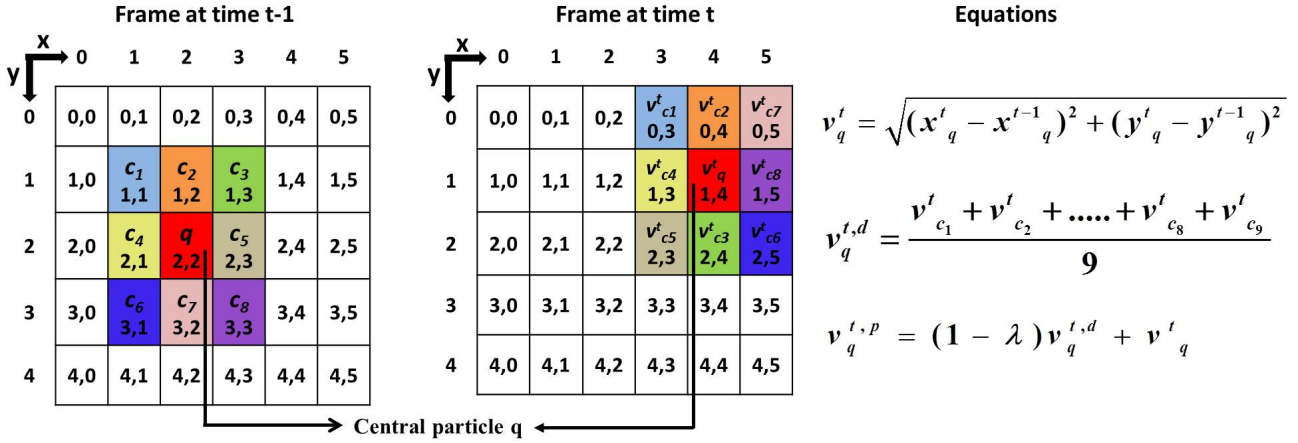
Fig. 2 Particle motion in consecutive frames. The particle q annotated in red is the central particle. The particles annotated in other colors are the neighboring particles. The calculation of current velocity v_q^t, the desired velocity $v_q^{t,d}$, and the personal desired velocity $v_q^{t,p}$ is presented.

the particle q also changes either due to the squeezing or spreading of the neighboring particles c_j. In such irregular flow of the particles, the velocity of the particle q is called personal desired velocity $v_q^{t,p}$ (as shown in Fig. 2) formulated in Eq. (4),

$$v_q^{t,p} = (1 - \lambda)v_q^{t,d} + v_q^t \qquad (4)$$

where λ is a panic parameter. If a particle q exhibits an individualistic action then λ decreases. Consequently, the desired velocity $v_q^{t,d}$ is replaced with the personal desired velocity $v_q^{t,p}$ in the formulation of the personal force $\Lambda_q^{t,p}$ as given in Eq. (5).

$$\Lambda_q^{t,p} = \frac{1}{\epsilon}(v_q^{t,p} - v_q^t) \qquad (5)$$

Modeling the personal force through Eq. (5) is one of the novelty of our proposed HSIM method. The relaxation/time parameter ϵ should not be set to $\epsilon = 0$ since the personal desired force will be undeterminable. We have provided the details of the dependency of our method on ϵ in the parameter analysis Section (Sec. 4.1). A significant difference between the values of $v_q^{t,p}$ and v_q^t presents the strong influence of the neighboring particles on the central particle q. Therefore, the personal force $\Lambda_q^{t,p}$ will bring greater contribution to the DI-SFM in term of the influence of the neighboring particles on the central particle. Without the use of Eq. (5), the DI-SFM model may not take into account the effect of squeezing and spreading of the neighboring particles.

It is worth mentioning that Eq. (3) and Eq. (5) are different from each other. The modeling of Eq. (3)

considers the desired velocity $v_q^{t,d}$. The particle q exhibits the desired velocity $v_q^{t,d}$ under the influence of the neighboring particles in a consistent situation where no squeezing and spreading occurs. In this situation, the desired velocity $v_q^{t,d}$ is calculated over a spatial window of size $w \times w$. However, the modeling of Eq. (5) considers the personal desired velocity $v_q^{t,p}$. When the distribution of the particles changes, the velocity of the particle q also changes either due to the squeezing or spreading of the neighboring particles. To cope with this irregular flow, the personal desired velocity $v_q^{t,p}$ of the particle q is modeled according to Eq. (4).

The repulsive force $\Lambda_q^{t,r}$, in Eq. (2), represents both the attempt of particle q to keep a certain safety distance from other particles, and the desire to gain more space in congested situations. For this purpose, the repulsive fore $\Lambda_q^{t,r}$ depends on the actual velocity v_q^t and the average distance Ω_q^t of a particle q from its neighboring particles. If the particle q is moving with higher actual velocity, the chance of getting an impact from the neighboring particle is higher. Therefore, with the increase of actual velocity v_q^t, the repulsive force $\Lambda_q^{t,r}$ increases. The motion of the particle q is influence by other particles. In particular, the particle q keeps a certain distance from other particles that depends on the particle distribution and desired speed. The personal space of the particle q represented by Ω_q^t plays an essential role. A particle normally feels increasingly uncomfortable the closer it gets to a neighboring particle, who may react in an unpredictable way. This results in repulsive force $\Lambda_q^{t,r}$ which can be modeled as an exponential function of the average distance Ω_q^t of particle q. If the distance of the particle q increases from the

neighboring particles, the repulsive force $\Lambda_q^{t,r}$ decreases and vice versa. Therefore, the repulsive force $\Lambda_q^{t,r}$ is formulated in Eq. (6).

$$\Lambda_q^{t,r} = v_q^t exp(\frac{-\Omega_q^t}{K_0} + \frac{K_1}{\Omega_q^t}) \qquad (6)$$

Modeling the repulsive force through Eq. (6) is another novelty of the proposed HSIM. In Eq. (6), Ω_q^t represents the average distance of a particle q from its neighboring particles over a spatial window of size $w \times w$ in the current frame at time t. Both constants K_0 and K_1 are used for tuning to determine the effect of the neighboring particles in term of the average distance Ω_q^t on the model. The average distance Ω_q^t is present in Eq. (6) in both positive $+\Omega_q^t$ and negative $-\Omega_q^t$ formats to help the tuning process. For example, if $K_0 = 1$, $K_1 = 0.3$, and $\Omega_q^t = 2$. The result of the exponential function is 0.15. Similarly, if $K_0 = 0.4$, $K_1 = 1$, and $\Omega_q^t = 2$. The result of the exponential function is 0.01. It means the output value is in smaller range to meet the condition of the DI-SFM magnitude $|D_q^t|$ (explained in the next paragraph). Therefore, it helps to prevent the propagation of the motion of the irrelevant particles in our HSIM model. It is reasonable to model particles such that they keep small distances from the neighboring particles, to which they are related or attracted to, and keep far distances from the particles associated with non-motion regions. Therefore, when Ω_q^t is very small, particles are squeezed and the repulsive force will increase, reflecting the strong reactions of those located in areas of high interactions. We put the formulation of personal force $\Lambda_q^{t,p}$ from eq. (5) and the formulation of repulsive force $\Lambda_q^{t,r}$ from Eq. (6) into Eq. (1). Overall, the DI-SFM can be summarized as in Eq. (7), where ϵ, K_0 and K_1 are constants.

$$|D_q^t| = |\frac{1}{\epsilon}(v_q^{t,d} - v_q^t) + v_q^t exp[\frac{-\Omega_q^t}{K_0} + \frac{K_1}{\Omega_q^t}]| \qquad (7)$$

All particles are of the same sizes, therefore, we set mass $m_q = 1$ in Eq. (7). Our paper is related to pedestrian motion segmentation where no particle gets significance or priority over any other particle. Therefore, setting mass $m_q = 1$ is consistent with the model. Some particles can be assigned different masses for different problems such as anomaly detection and panic detection. The magnitude $|D_q^t|$ in Eq. (7) is used as a threshold to remove particles associated with static regions. We calculate $|D_q^t|$ for each particle in a video frame. If for a particle q, the magnitude $|D_q^t|$ meets the condition $0 < |D_q^t| \leq 1$, then the particle q is retained. This is and advantage of the DI-SFM that defines the bounds for
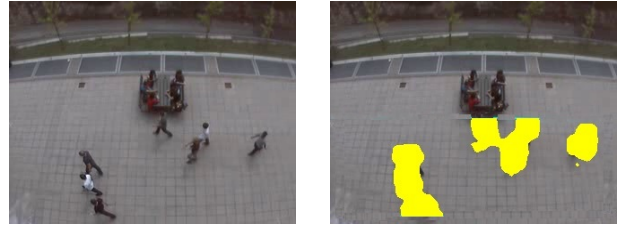
Fig. 3 Retaining the motile particles with distribution independent social force model (DI-SFM). The original frame from a video sequence (left column); the potential particles are annotated in yellow (right column).

the threshold for the particles of interest. Starting from the top-left corner of a video frame, each particle under observation is a central particle. We calculate $|D_q^t|$ for the particle using Eq. (7) and determine if it is to be retained or removed. Fig. 3 depicts a frame from a video sequence (left column) where the retained particles of interest are annotated in yellow (right column).

3.3 Motion segmentation

Our proposed HSIM is based on the communal model that is based on the work of Pan et al. [30]. It has been used in many other fields including voting models in physics [8], cascade models in epidemiology [44], attitude influence in psychology [12], and information exchange models in economics [1]. The analysis and understanding of who influences whom in the pedestrian flow can be complicated because internal states of individuals are often not observable. Therefore, it is difficult to infer influence from correlated behaviors between particles. The internal states represent the intentions of the individuals to get closer or move away from the neighboring pedestrians. Additionally, the context of specific interactions can change the effect one particle has on another. Our proposed HSIM presents better understanding of the interactions in the randomly distributed pedestrian flows. Our model uncovers shifts in the way states may be transmitted between particles at different points in time. It is built on the concept that the state of a particle is influenced by the states of its neighbors. Each particle in the distributed pedestrians has a defined strength of influence over every other particle. Thus, each relationship can be weighted according to this strength considering the fact that pedestrian flow is randomly distributed.

Our proposed HSIM model utilizes interaction influence to estimate how much the state of one particle affects the state of another particle. Additionally, it infers how influence changes over time. Therefore, the

segmentation of pedestrian flow can be naturally modeled using the HSIM. For this purpose, we compute the average distance among the particles and their density as two representative features to infer the interaction among particles. For both features, orientation is used as a prior, meaning that a set of particles are considered, only if their relative offset in terms of direction of motion appears in a predefined range. Without loss of generality, we select four different directions quantized with a step size of 90 degrees. We consider particles that are associated with the pedestrian flow. Then we evaluate their interactions through the HSIM model. Thus it allows modeling the pedestrian dynamics, adaptively.

According to the HSIM, the state h_{q,c_j}^t of the particle q surrounded by the neighboring particles c_j is derived from consecutive temporal observations at time t and $t-1$. We model the influence among particles by computing the influence matrix U_{q,c_j}. For this purpose, we consider a spatio-temporal window of size $C \times C$ where $C^2 - 1$ is the total number of the neighboring particles. For the spatio-temporal window $C \times C$, we use only consecutive frames at time $t-1$ and t. In our work, particles are generated through the Good-Features-To-Track algorithm, and tracked by the Lucas-Kanade optical flow [49]. The influence matrix U_{q,c_j} is computed at discrete steps at the end of each tracking period. In order to model the U_{q,c_j}, we assume that each particle relates with the others according to a Gaussian distribution. Therefore, the influence of the neighboring particles c_j on the central particle q is computed as in Eq. (8) and Eq. (9),

$$U_{q,c_j}^G = \frac{1}{\sigma_1 \sqrt{2\pi}} e^{\frac{\sum_{j \in \{1...C^2-1\} \wedge c_j \neq q} d(q(t),c_j(t-1))}{2\sigma_1^2}} \tag{8}$$

$$U_{q,c_j}^A = \frac{1}{\sigma_2 \sqrt{2\pi}} e^{\frac{\sum_{j \in \{1...C^2-1\} \wedge c_j \neq q} d(q(t),c_j(t-1))}{2\sigma_2^2}} \tag{9}$$

where d is the distance between the central particle q and a neighboring particle c_j in consecutive frames at time t and $t-1$, respectively. The distance is calculated as $d = |x_q^t - x_{c_j}^t| + |y_q^t - y_{c_j}^t|$. Both σ_1^2 and σ_2^2 are the variances calculated from the motion information of the central particle and the neighboring particles in the spatio-temporal window $C \times C$. In Eq. (8), σ_1^2 is calculated considering both the velocities and accelerations of the central particle and the neighboring particles. In Eq. (9), σ_2^2 is calculated using only velocities of the central particle and the neighboring particles. Both U_{q,c_j}^G and U_{q,c_j}^G are scalar quantities. Particles are then classified in two states (h), namely grouped (G) or alone (A) according to the Eq. (10) and Eq. (11).

$$S\left(h_{q,c_j}^t \Rightarrow G\right) = Int(G/h^{t-1}) \times \frac{U_{q,c_j}^G}{C^2} \tag{10}$$

$$S\left(h_{q,c_j}^t \Rightarrow A\right) = Int(A/h^{t-1}) \times \frac{U_{q,c_j}^A}{C^2} \tag{11}$$

where $Int(G/h^{t-1})$ and $Int(A/h^{t-1})$ represent the interaction elements. We model them using the matrices E_{q,c_j} and F_{q,c_j} both of size $C \times C$, so as $Int(G/h^{t-1}) = det(E_{q,c_j})$ and $Int(A/h^{t-1}) = det(F_{q,c_j})$. $det(.)$ represents the determinant of a matrix. For both matrices, the size $C \times C$ is the same as the size of spatio-temporal window in the previous Section. The matrix E_{q,c_j} consists of both the velocities and pressures in the form of pairs for q and c_j as shown in Eq. (12).

$$E_{q,c_j} = \begin{pmatrix} \frac{v_{q,c_1}^{t,t-1} + p_{q,c_1}^{t,t-1}}{2} & \frac{v_{q,c_1}^{t,t-1} + p_{q,c_1}^{t,t-1}}{2} & \frac{v_{q,c_1}^{t,t-1} + p_{q,c_1}^{t,t-1}}{2} \\ \frac{v_{q,c_1}^{t,t-1} + p_{q,c_1}^{t,t-1}}{2} & \frac{v_{q,q}^{t,t-1} + p_{q,q}^{t,t-1}}{2} & \frac{v_{q,c_1}^{t,t-1} + p_{q,c_1}^{t,t-1}}{2} \\ \frac{v_{q,c_1}^{t,t-1} + p_{q,c_1}^{t,t-1}}{2} & \frac{v_{q,c_1}^{t,t-1} + p_{q,c_1}^{t,t-1}}{2} & \frac{v_{q,c_1}^{t,t-1} + p_{q,c_1}^{t,t-1}}{2} \end{pmatrix} \tag{12}$$

where $v_{q,c_1}^{t,t-1} = |v_q^t - v_{c_1}^{t-1}|$ and $p_{q,c_1}^{t,t-1} = |p_q^t - p_{c_1}^{t-1}|$. v_q^t is the velocity of the central particle q at time t and $v_{c_1}^{t-1}$ is the velocity of the neighboring particle c_1 at time $t-1$. p_q^t is the pressure of the central particle q at time t and $p_{c_1}^{t-1}$ is the pressure of the neighboring particle c_1 at time $t-1$. The pressure p_q^t is computed according to $p_q^t = \frac{f_q^t}{A}$ using the force f_q^t and the area A. The force is $f_q^t = m \times g$ where m is the mass and g is the acceleration. The mass is set to $m = 1$ for consistency with the other equations in the model and the acceleration is calculated from the difference of velocities. We calculate the area A from the size of the spatio-temporal window according to $A = C^2$.

The matrix F_{q,c_j} consists of only velocity information as shown in Eq. (13).

$$F_{q,c_j} = \begin{pmatrix} v_{q,c_1}^{t,t-1} & v_{q,c_1}^{t,t-1} & v_{q,c_1}^{t,t-1} \\ v_{q,c_1}^{t,t-1} & v_{q,q}^{t,t-1} & v_{q,c_1}^{t,t-1} \\ v_{q,c_1}^{t,t-1} & v_{q,c_1}^{t,t-1} & v_{q,c_1}^{t,t-1} \end{pmatrix} \tag{13}$$
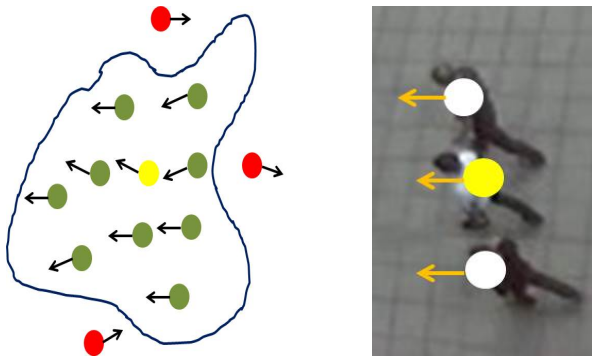
Fig. 4 Particle grouping. A synthetic example of particle grouping is presented in the left column. In the right column, particles annotated in white constitute group with the reference particle annotated in yellow (only one particle per moving object is displayed for the purpose of visualization).

High score S will dictate the state of the central particle q in the current time window. The grouped particles G constitute one segment R. The alone particles A are left to make group with other neighboring particles. This concept presents motion segmentation of particles considering both simple and complex distributed situations that can be modeled using grouped Eq. (10) and alone Eq. (11). These equations formulate the behavior of the particle to deal with the situations it is confronted with in a social system. It is therefore possible to put the behavior of a particle into the equations of motion classified as grouped and alone. According to these equations, the temporal changes of a particle surrounded by other particles are described by motion information in spatio-temporal window $C \times C$. Clearly, this information must represent the effect of the environment (e.g., other particles) on the behavior of the described particle. Therefore, the change in motion is exerted by the environment on a particle's body. This change describes the concrete motivation of the described particle to act. In the case of particle behavior this motivation evokes the physical production of an acceleration or a deceleration as a reaction to the perceived information that the particle obtains about its environment. In summary, one can say that a particle acts as if it would be subject to external forces of the surroundings. For the purpose of visualization, in Fig. 4, a set of synthetic particles are presented. Each moving object in Fig. 4 (right column) is represented by a single particle for the purpose of visualization. However, each object is represented by arbitrary number of particles in general. In the Fig. 4 (left column), a central particle, annotated in yellow, is grouped with the neighboring particles annotated in green. Three particles, annotated in red, are not included in the same group since their orientations do not conform to the

orientation of the central particle. Moreover, a central particle annotated in yellow and neighboring particles annotated in white are shown in Fig. 4 (right column).

We analyze the particles one by one starting from the top left corner of a frame. We observe each particle at a given time as the central particle. We then calculate $S\left(h_{q,c_j}^t \Rightarrow G\right)$ and $S\left(h_{q,c_j}^t \Rightarrow A\right)$ for each central particle. If the magnitude of $S\left(h_{q,c_j}^t \Rightarrow G\right)$ is great than the magnitude of $S\left(h_{q,c_j}^t \Rightarrow A\right)$ that means the central particle constructs a group or a coherent segment R. If the magnitude of $S\left(h_{q,c_j}^t \Rightarrow G\right)$ is less than the magnitude of $S\left(h_{q,c_j}^t \Rightarrow A\right)$, then the central particle is either part of the background or it could be the neighboring particle of any other central particle. Moreover, once a set of particles are considered to be the neighboring particles of a central particle, they are not analyzed again to affiliate them to any other central particle.

The repose with which our HSIM method can deal with the varying pedestrians distributions, as well as an automatically adaptive nature, are unmatched in other methods [2][14][39][27][17][37]. Our proposed HSIM method is based on the social influences that depend on the interaction of each particle with all its neighbors. In such interaction, however, each particle can have different number of neighbors. Due to this property, our HSIM method can easily deal with complex physical layout of pedestrian flows and various regions of the scenes that are completely irregular in terms of distributions.

To eliminate the undesirable effects of varying distributions, the key is to analyze the interactions among particles to represent coherent segments in the pedestrian flows. Our HSIM method considers the social influences in the pedestrian scenes from particle level to group level. This is to mitigate the effects of the changing distributions for an effective pedestrian motion segmentation. The utilization of our modeling counteracts the difficulty to group particles in the scenes due to the variation in the pedestrian distributions over time. For this purpose, the evolution of boundaries of different consistent segments is inferred from the interaction elements of our model. Our HSIM model acquires optimized boundary adherence of coherent segments in pedestrian flows due to the modeling of the social interactions. Since our proposed HSIM is leveraging on the influence matrix and the social interactions, conformation to varying distributions is achieved.

| Input					Output				
7	7	7	7	7	7	7	7	7	7
7	6	6	6	7	7	7	7	7	7
2	2	6	6	6	3	3	7	7	7
3	3	4	4	4	3	3	5	5	5
3	3	5	5	5	3	3	5	5	5

Fig. 5 Example of the working of GDBM: In the output, R_6 is merged with R_7, R_4 is merged with R_5, and R_2 is merged with R_3.

3.4 Refinement of segmentation map

The output from the communal model in the previous Section consists of pedestrian motion segmentation. To further refine the segmentation and to obtain a regularized representation of the motion field, we use a procedure based on the GBDM. The distinguishing feature of the GBDM is in the adoption of spanning tree T for accurate segmentation in local regions with complex motion patterns. In fact, to refine the motion segmentation, different methods can be used. However, due to low computation complexity, we explore the GBDM [11]. Additionally, the GBDM [11] is an unsupervised method that takes into account the segmentation in terms of both appearance and motion.

We illustrate an example to show the segment merging ability of the GBDM in Fig. 5. According to the GBDM, a pedestrian flow can be represented as a set of nodes of a graph, where each node corresponds to a particle of the video frame. We created the input and fed it to the GDBM that in turn merges different regions. The input and output are considered in terms of pixel values. The region with pixel values equal to 7 is represented by R_7. Therefore, regions R_2, R_3, R_4, R_5, R_6, and R_7 are shown in the input. In the output, R_6 is merged with R_7, R_4 is merged with R_5, and R_2 is merged with R_3. The GDBM is an effective way to fuse similar motion regions, thus limiting the effect of unrefined segmentation in pedestrian flows.

The GBDM transforms a video into a 3D lattice that generates a minimum spanning tree $T(V, E)$ for hierarchical representation as shown in Fig. 6. In the tree T, a set Γ is considered where $\bigwedge(\Gamma)$ represents the set of all subsets of Γ. A partition ρ of Γ is a subset $\rho \subseteq \bigwedge(\Gamma)$ if the intersection of any two distinct elements of ρ is empty and if the union of all elements in ρ is equal to Γ. In the tree $T(V, E)$, V is the set of all vertices and E is the set of all edges. Each element of V is a vertex that represents a particle and each element of E is an edge that represents a weight ω. Let R_1 and R_2 be two segments of a partition ρ. We represent the vertices by α_i and β_j that correspond to particle P_i
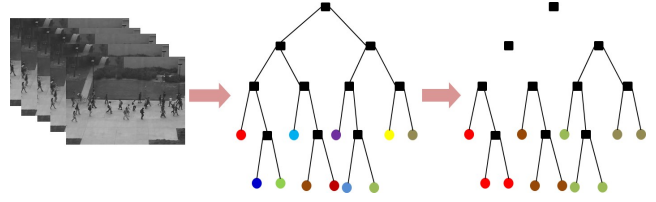
Fig. 6 Graph-based dissimilarity measure (GBDM). A tree has been constructed and the identification of different segments is made from the tree.

and P_j, respectively. The edge set E is constructed by connecting pairs of particles that are neighbors in an 8-connected manner. The GBDM evaluates two segments R_1 and R_2 in the tree T and find out if they are merged or not. For this purpose, we calculate the weight ω of all edges as intensity difference of particles in Eq. (14),

$$\omega((\alpha_i, \beta_j)) = |I_{P_i} - I_{P_j}| \tag{14}$$

where I_{P_i} and I_{P_j} are the intensities of particle P_i and particle P_j, respectively. We convert the color intensities into grey scale intensities to find Eq. (14). According to Eq. (14), the edges between two vertices in the same segment should have relatively low weights, and edges between vertices in different segments should have higher weights.

The GBDM measures the evidence for a boundary between two segments by comparing intensity differences across the boundary, and the intensity differences between neighboring particles within each segment. To further elaborate, two measures are considered to know whether two segments must be merged. The first measure is called the difference $D(R_1, R_2)$ and the second measure is called the minimal difference $D_M(R_1, R_2)$ between two neighboring segments R_1 and R_2. Therefore, the region merging predicate, formulated in Eq. (15), is calculated for merging two adjacent segments,

$$P = \begin{cases} true & \text{if} \quad D(R_1, R_2) \leq D_M(R_1, R_2) \\ false & \text{otherwise} \end{cases} \tag{15}$$

The difference $D(R_1, R_2)$ between two neighboring segments R_1 and R_2 is calculated according to Eq. (16),

$$D(R_1, R_2) = \min_{\alpha_i \in R_1, \beta_j \in R_2} \{w(\alpha_i, \beta_j)\} \tag{16}$$

$D(R_1, R_2)$ is the smallest edge weight among all the edges that link R_1 to R_2. In Eq. (15), $D_M(R_1, R_2)$ is the minimal difference formulated in Eq. (17),

$$D_M(R_1, R_2) = min\{D_I(R_1) + \gamma(R_1), D_I(R_2) + \gamma(R_2)\}$$

$$(17)$$

where γ is a function that controls the extent of the difference between two segments R_1 and R_2. $D_I(R_1)$ and $D_I(R_2)$ are the internal differences of segment R_1 and segment R_2, respectively. The internal difference of segment R_1 is calculated as in Eq. (18),

$$D_I(R_1) = \max_{\alpha_i, \alpha_j \in R_1, i \neq j} \{w(\alpha_i, \alpha_j)\} \qquad (18)$$

The internal difference $D_I(R_1)$ is the highest edge weight among all the edges linking two vertices of R_1 in the tree T. If the difference is greater than their internal differences there is an evidence of a boundary between them. For small segments, $D_I(.)$ is not a good measure of the local characteristics of the given information. In the extreme case, when $|R_1| = 1$, we have $D_I(R_1) = 0$. $|R_1|$ represents the size of segment R_1 in terms of total number of particles. The γ function in Eq. (17) is formulated in Eq. (19),

$$\gamma(R_1) = \frac{z}{|R_1|} \qquad (19)$$

where z is a parameter considering the size of the segments to be merged. For example, a larger z would merge smaller segments. The merging criterion of two segments R_1 and R_2 is formulated in Eq. (20)

$$D(R_1, R_2) \leq min\{D_I(R_1) + \frac{z}{|R_1|}, D_I(R_2) + \frac{z}{|R_2|}\} \quad (20)$$

The merging criterion depends on the parameter z. To take this dependency into account, the scale $H_{R_2}(R_1)$ of R_1 relative to R_2 is formulated in Eq. (21).

$$H_{R_2}(R_1) = (D(R_1, R_2) - D_I(R_1))|R_1| \qquad (21)$$

The scale $H_{R_2}(R_1)$ is a measure based on the difference $D(R_1, R_2)$ between R_1 and R_2, on the internal difference $D_I(R_1)$ of R_1, and on the size $|R_1|$ of R_1. Furthermore, the scale H(R_1,R_2) based on both $H_{R_2}(R_1)$ and $H_{R_1}(R_2)$ is formulated in Eq. (22):

$$H(R_1, R_2) = max(H_{R_2}(R_1), H_{R_1}(R_2)) \qquad (22)$$

Therefore, Eq. (20) is reformulated according to Eq. (23) for segment merging.

$$z \geq H(R_1, R_2) \qquad (23)$$

Thus, two adjacent segments R_1 and R_2 must be merged at scale z if their dissimilarity measure is smaller than or equal to z. The overall procedure of the proposed HSIM method is summarized in Algorithm 1.

Algorithm 1: Proposed HSIM Algorithm

Input : Video sequence
Parameters : $K_0, K_1, \epsilon, z, \lambda$
Set $L = \{l_n\}_{n=1}^N$ of N
initialized points over the starting frame.
**Output:** Regions of consistent motion.
foreach *TrackingInterval=t* **do**
  Track the points in L by KL tracker
  foreach $l_n \in L$ **do**
    $\Lambda_q^{t,p} \leftarrow \frac{1}{\epsilon}(v_q^{t,p} - v_q^t)$ (5)
    $\Lambda_q^{t,r} \leftarrow v_q^t exp(\frac{-\Omega_q^t}{K_0} + \frac{K_1}{\Omega_q^t})$ (6)
    $|D_q^t| \leftarrow \frac{1}{\epsilon}(v_q^{t,d} - v_q^t) + v_q^t exp[\frac{-\Omega_q^t}{K_0} + \frac{K_1}{\Omega_q^t}]$ (7)
  end
  1 Remove static particle using the threshold $|D_q^t|$
  2 Calculate influence matrices U_{q,c_j}^G and
    U_{q,c_j}^A using Eq. (8) and Eq. (9), respectively.
  3 Calculate interaction elements $Int(G/h^{t-1})$ and
    $Int(A/h^{t-1})$.
  4 **if** $S\left(h_{q,c_j}^t \Rightarrow G\right) > S\left(h_{q,c_j}^t \Rightarrow A\right))$ **then**
    $l_n \in$ coherent segment R
    else
    | $l_n \in individual$
    end
  end

  5 Refinement using GBDM: **if** $z >= H(R_1, R_2)$ **then**
  | merge R_1 and R_2
  end
end

---

4 Experiments

We evaluate the performance of our proposed HSIM method for pedestrian motion segmentation on four standard benchmark datasets. These include the publicly available benchmark PETS2009 [31], UCSD [24], USC [50], and CUHK [37] datasets. The PETS2009 dataset consists of various densities and distributions of pedestrian sequences. We used both L1 and L2 subsets from PETS2009 dataset. The UCSD dataset represents pedestrian videos captured by a fixed camera overlooking pedestrian walkways. We consider both Ped1 and Ped2 subsets from UCSD dataset that consist of low-density to high-density pedestrian sequences. The resolutions of Ped1 and Ped2 are 158x238 and 240x360, respectively. The USC dataset consists of a single outdoor scene that is captured from a camera with a 40 degrees tilt angle. There are total 900 frames where the frame size is 360x240. The CUHK dataset consists of crowd videos with different densities and perspective scales. This dataset is collected considering various environments, e.g. streets, shopping malls, airports, and parks. We compare the performance of our proposed HSIM method with two baseline approaches and four closely related state-of-the-art approaches: the robust trajec-
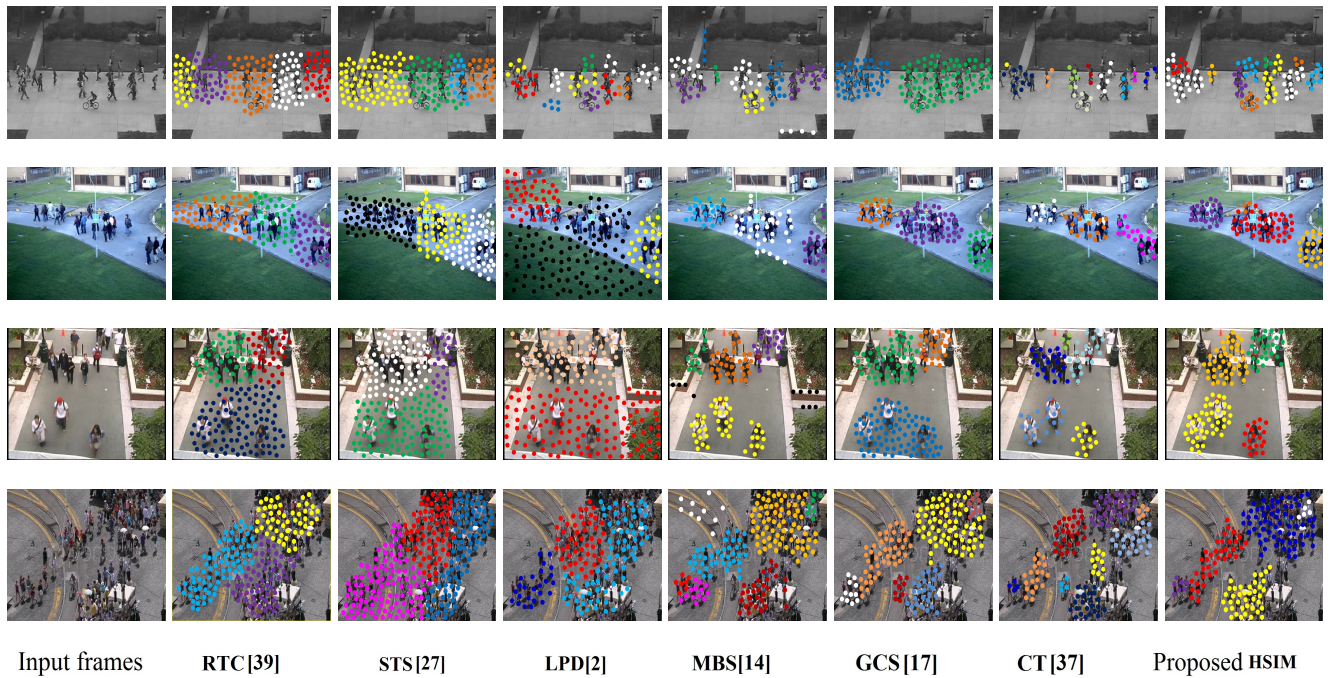
Fig. 7 Qualitative comparison. The proposed HSIM method is compared with two baseline: RTC [39] and STS [27], and four state-of-the-art: LPD [2], MBS [14], GCS [17], and CT [37] motion segmentation methods. First column represents input frames from UCSD, PETS2009, UCS, and CUHK datasets, respectively. The proposed HSIM presents more accurate segmentation results. Moreover, our method segments the pedestrian flows consistently according to the physical layouts.

tory clustering (RTC) [39], the spatio-temporal segmentation (STS) [27], the Lagrangian particle dynamics (LPD) [2], the multicue-based segmentation (MBS) [14], the granular crowd segmentation (GCS) [17], and the collective transition (CT) [37]. For pixel-level quantitative evaluation of motion segmentation, the Jaccard Similarity Metric (JSM) and the F-score (FS) are calculated to measure the overall accuracy according to the TCA approach [38] and GCS approach [17], respectively.

The qualitative evaluation of the reference methods and our proposed method is presented in Fig. 7. Results for one sequence from each dataset are shown. The first, second, third, and last rows present results for the reference methods and our method for the UCSD, PETS2009, USC, and CUHK datasets, respectively. For the purpose of visualization, the frames are superimposed by color segmentation results. First column depicts the sample frames taken from the original video sequences, while second to last columns show the results obtained using the RTC [39], the STS [27], the LPD [2], the MBS [14], the GCS [17], the CT [37], and the proposed HSIM method, respectively. Our proposed HSIM method segments the pedestrian motion accurately in the last column. Our method copes with the random pedestrian distribution due to the modeling of the social influences. The RTC [39], the STS [27],

and the LPD [2] perform inconsistent motion segmentation. The RTC [39] and the LPD [2] approaches are based on a temporal smoothness constraint for motion segmentation. Therefore, both approaches resulted into improper segmentation due to the varying distributions of the pedestrians. The boundaries of segments are discontinuous when the distribution of the pedestrians is changing over time. In the STS [27] approach, the recursive filtering cannot update the parameters of the Gaussian mixture model adaptively to cope with the problem of random pedestrian distribution. Therefore, the STS [27] includes static regions as part of the segmentation map. Thus the overall segmentation is very noisy as can be seen in the second column. The MBS [14] and the CT [37] cannot segregate pedestrian groups into consistent segments due to its inability to take into account the dynamic nature of the underlying motion. In case of randomly distributed flows, it is important to associate homogeneous segments properly. The GCS [17] uses correlation among pixels to outline the boundaries between moving pedestrians and background regions. However, the correlation among pixels cannot be retained due inter-occlusions and dynamic scene structure when the distribution changes over time. It can be seen in the first row and sixth column where dynamically different pedestrian segments are merged together.

Table 1 Quantitative analysis. The average Jaccard Similarity Metric (JSM) for the reference methods and our proposed HSIM method for the four datasets, UCSD, PETS2009, USC, and CUHK are presented.

Datasets	RTC[39]	STS[27]	LPD[2]	MBS[14]	GCS[17]	CT[37]	Prop.
UCSD	54.31	46.72	49.53	75.60	69.58	75.23	79.64
PETS	47.82	44.38	48.41	69.32	72.66	71.49	75.10
USC	43.69	37.91	40.99	71.53	76.87	75.67	82.23
CUHK	48.62	43.48	51.68	73.98	78.93	79.13	81.81
Average	48.66	43.12	47.65	72.61	74.51	75.38	79.69

Table 2 Quantitative analysis. The average F-scores for the reference methods and our proposed HSIM method for the four datasets, UCSD, PETS2009, USC, and CUHK are presented.

Datasets	RTC[39]	STS[27]	LPD[2]	MBS[14]	GCS[17]	CT[37]	Prop.
UCSD	0.32	0.27	0.29	0.45	0.43	0.53	0.58
PETS	0.30	0.26	0.29	0.39	0.41	0.42	0.55
USC	0.33	0.27	0.28	0.49	0.51	0.51	0.61
CUHK	0.41	0.35	0.43	0.50	0.55	0.56	0.58
Average	0.34	0.29	0.32	0.46	0.47	0.51	0.58

Additionally, we provide multi-frames qualitative evaluation of the reference methods and our proposed method in Fig. 8. Results of initial, middle, and last frames of a video sequence from each dataset are shown. For example, the first, second, and third rows present results for the reference methods and our method for a video sequence from USCD dataset considering initial, middle, and last frames. First column depicts the sample frames taken from the original video sequences, while second to last columns show the results obtained using the RTC [39], the STS [27], the LPD [2], the MBS [14], the GCS [17], the CT [37], and the proposed HSIM method, respectively. The initial frames for all the datasets presenting no segmentation show that the segmentation process is not started yet. Our proposed HSIM method segments the pedestrian motion accurately in the last column. The RTC [39], the STS [27], and the LPD [2] segment the pedestrian motion improperly as can be seen in the second, third, and fourth columns. The results of the MBS [14] method, the GCS [17] method, and the CT [37] method presented in the fifth, sixth, and seventh columns show that these methods cannot perform motion segmentation accurately since they cannot handle the changing physical layouts of the pedestrian flows.

For quantitative evaluation, we computed the average JSM and F-scores for the four datasets. The results are shown in Table 1 and Table 2. We outperform all the reference methods due to strong capability of HSIM to

segment pedestrian motion. Thanks also to the DI-SFM for removing the static particles. Our method results in greater JSM and F-score for the four datasets as shown in Table 1 and Table 2, respectively. All the reference methods ignore change in the pedestrian distribution that holds only if the flow is coherent over time. As the pedestrian distribution changes, the performance declines. Such changes result into isolated regions, thus declining the overall performance. Moreover, the reference methods cannot aggregate structurally similar pixels into meaningful segments. Detecting and partitioning localized pedestrian segments is a very challenging problem.

4.1 Parameters Analysis

To examine the robust performance of our proposed HSIM method for the four datasets, we consider 24 different parameter configurations reported in Table 3. For this purpose, we use different spatial windows (of size $w \times w$) and spatio-temporal windows ($C \times C$), DI-SFM magnitudes ($|D|$), DI-SFM parameters (K_0, K_1, ϵ, and λ), and the GBDM parameter (z). For each experiment, the same sizes of the windows are considered for the sake of consistency. Six different sizes of windows and values for other parameters are taken into account. These other parameters ($|D|$, K_0, K_1, ϵ, and λ) can be varied in the range from 0.1 to 1. The DBDM parameter (z) is varied in the range from 5 to 30. It is worth noticing that each parameter is independent of each other.

The variations in the performances with respect to 24 different configurations are reported in Fig. 9. In the left and right columns of Fig. 9, gradual improvement in the performances in terms of both average JSM and average F-score can be noticed from configuration 1 to 4 for all the four datasets. However, the performances decline from configuration 4 to 7. The change in the performances (both average JSM and average F-score) can be noticed consistently for all the four datasets from configuration 7 to 9 (improved), 9 to 13 (declined), 13 to 16 (improved), 16 to 19 (declined), 19 to 22 (improved), and 22 to 24 (declined). Changing the window size from 3x3 to 9x9, the parameter z from 5 to 20, and the parameter λ from 1 to 0.5 improve the performance significantly. Therefore, it is worth to consider more particles in the surroundings. The first configuration shows poor performance since it considers the least particles in the surroundings and the value of λ is 1 that means the personal desired velocity $v_q^{t,p}$ in Eq. (4) is equal to the current velocity v_q^t only, thus it ignores the neighboring particles. Moreover, changing

| Input frames | RTC[39] | STS[27] | LPD[2] | MBS[14] | GCS[17] | CT[37] | Proposed HSIM |

Fig. 8 Qualitative comparison. The proposed HSIM method is compared with two baseline: RTC [39] and STS [27], and four state-of-the-art: LPD [2], MBS [14], GCS [17], and CT [37] motion segmentation methods. The top three rows represent frames from UCSD dataset, the next three rows represent frames from PETS2009 dataset, the second last three rows represent frames from USC dataset, and the bottom three rows represent frames from CUHK dataset. The last column shows that our proposed HSIM method performs accurate segmentation.
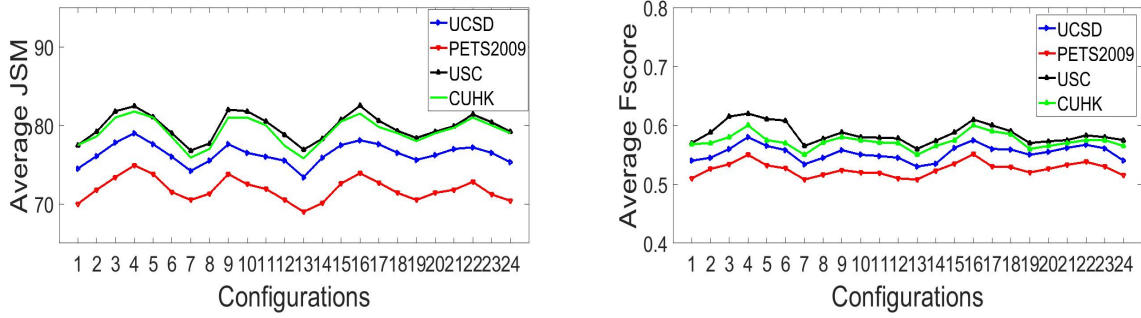
Fig. 9 Configuration set. The average JSM (left column) and average F-score (right column) are presented for the four datasets based on 24 configurations. Consistent variations in both graphs can be observed. The performance of our proposed HSIM method does not change by changing other parameters except the sizes of the windows and the parameter z.

Table 3 Parameter configurations. For parameters analysis for our proposed HSIM method, 24 different configurations are listed based on windows, DI-SFM magnitude ($|D|$), DI-SFM parameters (K_0, K_1, ϵ, λ), and the GBDM parameter (z). The parameters ($w \times w$) and ($C \times C$) represent the spatial and spatio-temporal windows, respectively.

Config	$w \times w$	$C \times C$	$|D|$	K_0	K_1	ϵ	z	λ
1	3x3	3x3	.1	.1	.1	.1	5	1
2	5x5	5x5	.3	.3	.3	.3	10	.9
3	7x7	7x7	.5	.5	.5	.5	15	.7
4	9x9	9x9	.7	.7	.7	.7	20	.5
5	11x11	11x11	.9	.9	.9	.9	25	.3
6	13x13	13x13	1	1	1	1	30	.1
7	13x13	13x13	.1	.1	.1	.1	5	1
8	11x11	11x11	.3	.3	.3	.3	10	.9
9	9x9	9x9	.5	.5	.5	.5	15	.7
10	7x7	7x7	.7	.7	.7	.7	20	.5
11	5x5	5x5	.9	.9	.9	.9	25	.3
12	3x3	3x3	1	1	1	1	30	.1
13	3x3	3x3	1	1	1	1	5	1
14	5x5	5x5	.9	.9	.9	.9	10	.9
15	7x7	7x7	.7	.7	.7	.7	15	.7
16	9x9	9x9	.5	.5	.5	.5	20	.5
17	11x11	11x11	.3	.3	.3	.3	25	.3
18	13x13	13x13	.1	.1	.1	.1	30	.1
19	3x3	3x3	.1	.1	.1	.1	30	.1
20	5x5	5x5	.3	.3	.3	.3	25	.3
21	7x7	7x7	.5	.5	.5	.5	20	.5
22	9x9	9x9	.7	.7	.7	.7	15	.7
23	11x11	11x11	.9	.9	.9	.9	10	.9
24	13x13	13x13	1	1	1	1	5	1

Table 4 Computational complexity. The computational complexities of the reference methods and our proposed HSIM method are presented in term of average number of seconds required to execute a frame.

Reference and proposed methods	Average execution time for one frame
RTC [39]	15
STS [27]	13
LPD [2]	21
MBS [14]	18
GCS [17]	09
CT [37]	10
Proposed HSIM	04

cantly by changing other parameters except the window size, the parameter z, and the parameter λ.

4.2 Computational overheads

To calculate the computational complexity, a 16GB RAM computer with a 3.5 GHz CPU is used to perform the experiments. Further reduction in the computational complexities is possible since these implementations are not optimized. In Table 4, we provide the computational complexities of our HSIM method, two baseline methods, and four state-of-the-art methods. These complexities are presented in term of average number of seconds per frame over all the datasets. The computational overheads of the RTC [39], the STS [27], the LPD [2], the MBS [14], the GCS [17], and the CT [37] requiring 15, 13, 21, 18, 09, and 10 seconds on average for processing a frame are very high. Our proposed HSIM requires only 04 seconds on average to execute a frame. Thus, our HSIM method shows significant improvement over the reference methods in term of computational overheads.

the window size from 9x9 to 13x13 declines the performance. It means that particles located far away do not contribute to the segmentation process. Therefore, the performance of our method does not change signifi-

4.3 Weaknesses of our proposed HSIM method

Our proposed HSIM method is targeted at motion segmentation considering pedestrian scene as a whole. Therefore, it does not render fine segmentation for the individual objects. The significance of the individual object does not prevail in pedestrian flows where the holistic characteristics of the scene are the accumulated result of the individualistic properties. The segmented information is used to detect various types of anomalies for a surveillance system. However, our method is not tested for pedestrian anomaly detection. Our method presents improvement in the performance in term of computational complexity. However, further improvement is needed to make it applicable for real-time applications.

5 Conclusion

We proposed an automatic method for pedestrian motion segmentation. We demonstrated the robustness of our proposed method using video sequences from four benchmark datasets: UCSD, PETS2009, USC, and CUHK. We compared the results of our proposed HSIM with two baseline and four closely related state-of-the-art methods. Experimental results demonstrated that the proposed HSIM method outperform the compared methods in term of both accuracy and computational complexity.

In our future work, we are interested in strategies for merging anomaly detection with our motion segmentation information. Camera motion is another important factor that will be considered for improved robustness. Furthermore, new features will be investigated to further reduce the computational complexity.

There is no conflict of interest.

References

1. Acemoglu, D., Ozdaglar, A., ParandehGheibi, A.: Spread of (mis) information in social networks. Games and Economic Behavior **70**(2), 194–227 (2010)
2. Ali, S., Shah, M.: A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. In: International conference on computer vision and pattern recognition, IEEE CVPR, pp. 1–6 (2007)
3. Baxter, R.H., Leach, M.J., Mukherjee, S.S., Robertson, N.M.: An adaptive motion model for person tracking with instantaneous head-pose features. Signal processing letters, IEEE SPL **22**(5), 578–582 (2015)
4. Biswas, S., Babu, R.V.: Anomaly detection in compressed h. 264/avc video. Multimedia tools and applications, Springer MTA **74**(24), 11,099–11,115 (2015)
5. Biswas, S., Praveen, R.G., Babu, R.V.: Super-pixel based crowd flow segmentation in h. 264 compressed videos. In: International conference on image processing, IEEE ICIP, pp. 2319–2323 (2014)
6. Bosch, M., Li, G., Wang, K.: A two-stage video object segmentation using motion and color information. In: International conference on image processing, IEEE ICIP, pp. 3916–3920 (2015)
7. Carrera, D., Boracchi, G., Foi, A., Wohlberg, B.: Scale-invariant anomaly detection with multiscale group-sparse models. In: International conference on image processing, IEEE ICIP, pp. 3892–3896 (2016)
8. Castellano, C., Fortunato, S., Loreto, V.: Statistical physics of social dynamics. Reviews of modern physics **81**(2), 591 (2009)
9. Chan, A.B., Vasconcelos, N.: Modeling, clustering, and segmenting video with mixtures of dynamic textures. Transactions on pattern analysis and machine intelligence, IEEE PAMI **30**(5), 909–926 (2008)
10. Cheng, K.W., Chen, Y.T., Fang, W.H.: An efficient subsequence search for video anomaly detection and localization. Multimedia tools and applications, Springer MTA pp. 1–22 (2015)
11. De Souza, K.J.F., de Albuquerque Araújo, A., do Patrocínio, Z.K., Guimarães, S.J.F.: Graph-based hierarchical video segmentation based on a simple dissimilarity measure. Pattern recognition letters, Elsevier PRL **47**, 85–92 (2014)
12. Friedkin, N.E.: The attitude-behavior linkage in behavioral cascades. Social Psychology Quarterly **73**(2), 196–213 (2010)
13. Fu, X., Wang, C.Y., Chen, C., Wang, C., Jay Kuo, C.C.: Robust image segmentation using contour-guided color palettes. In: International conference on computer vision, IEEE ICCV, pp. 1618–1625 (2015)
14. Hou, Y.L., Pang, G.K.: Multicue-based crowd segmentation using appearance and motion. Transactions on systems, man, and cybernetics: systems, IEEE TSMC **43**(2), 356–369 (2013)
15. Huerta, I., Amato, A., Roca, X., Gonzàlez, J.: Exploiting multiple cues in motion segmentation based on background subtraction. Journal of neurocomputing, Elsevier NC **100**, 183–196 (2013)
16. Khan, S.D., Bandini, S., Basalamah, S., Vizzari, G.: Analyzing crowd behavior in naturalistic conditions: Identifying sources and sinks and characterizing main flows. Journal of neurocomputing, Elsevier NC **177**, 543–563 (2016)
17. Kok, V.J., Chan, C.S.: Grcs: Granular computing-based crowd segmentation. Transactions on cybernetics, IEEE TC (2016) (2016)
18. Kumar, M., Bhatnagar, C.: Zero-stopping constraint-based hybrid tracking model for dynamic and high-dense crowd videos. The Imaging Science Journal, Taylor & Francis ISJ **65**(2), 75–86 (2017)
19. Leach, M.J., Baxter, R., Robertson, N.M., Sparks, E.P.: Detecting social groups in crowded surveillance videos using visual attention. In: International conference on computer vision and pattern recognition, IEEE CVPR Workshops, pp. 461–467 (2014)
20. Leach, M.J., Sparks, E.P., Robertson, N.M.: Contextual anomaly detection in crowded surveillance scenes. Pattern recognition letters, Elsevier PRL **44**, 71–79 (2014)
21. Ledig, C., Shi, W., Bai, W., Rueckert, D.: Patch-based evaluation of image segmentation. In: International conference on computer vision and pattern recognition, IEEE CVPR, pp. 3065–3072 (2014)
22. Li, R., Chellappa, R.: Group motion segmentation using a spatio-temporal driving force model. In: International conference on computer vision and pattern recognition, IEEE CVPR, pp. 2038–2045 (2010)

23. Luo, D., Huang, H.: Video motion segmentation using new adaptive manifold denoising model. In: International conference on computer vision and pattern recognition, IEEE ICCV, pp. 65–72 (2014)
24. Mahadevan, V., Li, W., Bhalodia, V., Vasconcelos, N.: Anomaly detection in crowded scenes. In: IEEE CVPR, pp. 1975–1981 (2010)
25. Mehran, R., Moore, B.E., Shah, M.: A streakline representation of flow in crowded scenes. In: European conference on computer vision, Springer ECCV, pp. 439–452 ((2010))
26. Milan, A., Leal-Taixé, L., Schindler, K., Reid, I.: Joint tracking and segmentation of multiple targets. In: International conference on computer vision and pattern recognition, IEEE CVPR, pp. 5397–5406 (2015)
27. Mukherjee, D., Wu, Q.J.: Streaming spatio-temporal video segmentation using gaussian mixture model. In: International conference on image processing, IEEE ICIP, pp. 4388–4392 (2014)
28. Narayana, M., Hanson, A., Learned-Miller, E.: Coherent motion segmentation in moving camera videos using optical flow orientations. In: International conference on computer vision, IEEE ICCV, pp. 1577–1584 (2013)
29. Nguyen, T.M., Wu, Q.J.: A consensus model for motion segmentation in dynamic scenes. Transactions on circuit and system for video technology, IEEE CSVT (2015) (2015)
30. Pan, W., Dong, W., Cebrian, M., Kim, T., Fowler, J.H., Pentland, A.S.: Modeling dynamical influence in human interaction: Using data to make better inferences about influence within social systems. Signal Processing Magazine, IEEE SPM **29**(2), 77–86 (2012)
31. PETS dataset: http://www.cvg.rdg.ac.uk/pets2009/a.html. In: PETS2009 (2009)
32. Rinke, N., Schiermeyer, C., Pascucci, F., Berkhahn, V., Friedrich, B.: A multi-layer social force approach to model interactions in shared spaces using collision prediction. Transportation Research Procedia, Elsevier TRP **25**, 1249–1267 (2017)
33. Rodger, I., Connor, B., Robertson, N.M.: Recovering background regions in videos of cluttered urban scenes. In: International conference on image processing, IEEE ICIP, pp. 4097–4101 (2015)
34. Rota, P., Ullah, H., Conci, N., Sebe, N., De Natale, F.G.: Particles cross-influence for entity grouping. In: European signal processing conference, IEEE EUSIPCO, pp. 1–5 (2013)
35. Seyedhosseini, M., Tasdizen, T.: Semantic image segmentation with contextual hierarchical models. Transactions on pattern analysis and machine intelligence, IEEE PAMI **38**(5), 951–964 (2016)
36. Shadden, S., Lekien, F., Marsden, J.: Definition and properties of lagrangian coherent structures from finite-time lyapunov exponents in two-dimensional aperiodic flows. Physica D: Nonlinear Phenomena **212**(3), 271–304 (2005)
37. Shao, J., Change Loy, C., Wang, X.: Scene-independent group profiling in crowd. In: International conference on computer vision and pattern recognition, IEEE CVPR, pp. 2219–2226 (2014)
38. Sharma, R., Guha, T.: A trajectory clustering approach to crowd flow segmentation in videos. In: International conference on image processing, IEEE ICIP, pp. 1200–1204 (2016)
39. Shi, F., Zhou, Z., Xiao, J., Wu, W.: Robust trajectory clustering for motion segmentation. In: International conference on computer vision, IEEE ICCV, pp. 3088–3095 (2013)
40. Solmaz, B., Moore, B.E., Shah, M.: Identifying behaviors in crowd scenes using stability analysis for dynamical systems. Transactions on pattern analysis and machine intelligence, IEEE PAMI **34**(10), 2064–2070 (2012)
41. Thida, M., Eng, H.L., Remagnino, P.: Laplacian eigenmap with temporal constraints for local abnormality detection in crowded scenes. Transactions on cybernetics, IEEE C **43**(6), 2147–2156 (2013)
42. Ullah, M., Ullah, H., Conci, N., De Natale, F.G.: Crowd behavior identification. In: International conference on image processing, IEEE ICIP (2016)
43. Wang, W., Lin, W., Chen, Y., Wu, J., Wang, J., Sheng, B.: Finding coherent motions and semantic regions in crowd scenes: A diffusion and clustering approach. In: European Conference on Computer Vision–ECCV 2014, pp. 756–771. Springer (2014)
44. Watts, D.J., Dodds, P.S.: Influentials, networks, and public opinion formation. Journal of consumer research **34**(4), 441–458 (2007)
45. Wu, S., San Wong, H.: Crowd motion partitioning in a scattered motion field. Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE SMC **42**(5), 1443–1454 (2012)
46. Wu, S., Wong, H.S., Yu, Z.: A bayesian model for crowd escape behavior detection. Transactions on circuits and systems for video technology, IEEE CSVT **24**(1), 85–98 (2014)
47. Xu, D., Song, R., Wu, X., Li, N., Feng, W., Qian, H.: Video anomaly detection based on a hierarchical activity discovery within spatio-temporal contexts. Journal of neurocomputing, Elsevier NC **143**, 144–152 (2014)
48. Yu, W., Johansson, A.: Modeling crowd turbulence by many-particle simulations. Physical Review E **76**(4), 046,105 (2007)
49. Yves, B.: Pyramidal implementation of the lucas-kanade feature tracker. Microsoft Res. Labs, Tech. Rep (1999)
50. Zhao, T., Nevatia, R.: Bayesian human segmentation in crowded situations. In: International conference on Computer Vision and Pattern Recognition, IEEE CVPR, vol. 2, pp. II–459 (2003)