Density independent hydrodynamics model for crowd coherency detection

Habib Ullah[a], Muhammad Uzair[a], Mohib Ullah[b], Asif Khan[c], Ayaz Ahmad[a], Wilayat Khan[a]

[a]*Department of Electrical Engineering, COMSATS Institute of Information Technology, Wah Cantt, Pakistan*
[b]*Norwegian University of Science and Technology, Gjovik, Norway*
[c]*Ghulam Ishaq Khan Institute of Engineering Sciences and Technology, Topi, Pakistan*

Abstract

We propose density independent hydrodynamics model (DIHM) which is a novel and automatic method for coherency detection in crowded scenes. One of the major advantages of the DIHM is its capability to handle changing density over time. Moreover, the DIHM avoids oversegmentation and thus achieves refined coherency detection. In the proposed DIHM, we first extract a motion flow field from the input video through particle initialization and dense optical flow. The particles of interest are then collected to retain only the most motile and informative particles. To represent each particle, we accumulate the contribution of each particle in a weighted form, based on a kernel function. Next, the smoothed particle hydrodynamics (SPH) is adopted to detect coherent regions. Finally, the detected coherent regions are refined to remove the effects of oversegmentation. We perform extensive experiments on three benchmark datasets and compare the results with 10 state-of-the-art coherency detection methods. Our results show that DIHM achieves superior coherency detection and outperforms the compared methods in both pixel level and coherent region level average particle error rates (PERs), average coherent number error (CNE) and F-score.

Keywords: Coherency detection, Crowed flow analysis, Smoothed particle hydrodynamics.

1. Introduction

Crowd flows represent movements of group of individuals that are pervasive in many real-world environments. Automatic coherency detection in crowded scenes is a challenging computer vision problem [1][2][3] and is useful in effectively decomposing scenes into meaningful parts. These parts can be exploited for automatic anomalous events detection and recognition.

The coherency and density of crowd are correlated. A uniform density over time represents a coherent flow of crowd. The coherency changes with the changing density flowing in multiple directions. Previous works [4][5][6][19][21] assumed that the level of movement remains uniform in a crowded scene. This means that the coherency of both low-density and high-density crowd flows remain consistent over time. However, this assumption may not hold in many real-world scenarios where the density of people is changing over time. For example, high density of pedestrians on a pedestrian pathway can be observed during office hours whereas the same density reduces in later hours.

To address these challenges, we propose a density independent hydrodynamics model (DIHM) to locally model the movement of crowd without distinguishing pedestrians individually. Our proposed DIHM is based on smoothed particle hydrodynamics (SPH) [9], which is extensively used to solve fluid dynamics problems [10][11][12][13]. Our motivation for the SPH directly comes from the observation that the crowd flows in videos resemble fluid flows. In fact, SPH models both compressible and incompressible liquids, which implies that it is independent of changes in the volume. Considering these capabilities of the SPH, DIHM models coherent regions in the crowd and can cope with the density of people that varies over time.

We consider each moving object as part of the crowd and non-moving objects or groups of people as a background. Our proposed DIHM method is depicted in Fig. 1. Firstly, we extract a motion flow field from the video using the Farnebäck optical flow technique [14]. We then incorporate the enthalpy model [15] to remove the static particles that do not contribute to the detection of coherency. Subsequently, the orientation information of the particles is collected where each particle represents the location of a pixel. Secondly, the coherent regions in the scene are detected by employing the SPH model. Finally, to consolidate and to refine the coherency de-
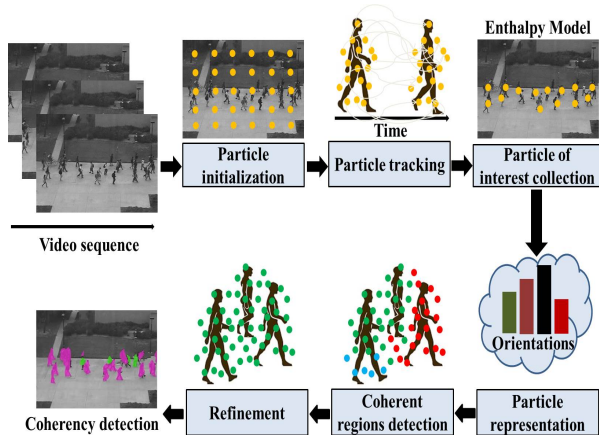
Figure 1: Illustration of the proposed method. Each particle represents the position of a pixel. For the sake of visualization, only a limited set of particles are overlaid on the frame. Particle of interest collection removes static particles associated with non-motion regions. The retained particles are used to detect coherency in the next steps.

tection, an unsupervised, robust and efficient multilayer spectral clustering (MLSC) [16] is exploited to group regions that are coherent both in appearance and motion. Thus, the overall framework renders more consistent coherency detection.

Our main contributions include the development of a novel density independent hydrodynamics model (DIHM) in conjunction with the enthalpy model for improved crowd coherency detection. One of the major attractions of the DIHM is its capability to handle changing density over time. To the best of our knowledge, we are the first to propose these models for coherency detection. Moreover, we extensively evaluate the proposed method on three benchmark datasets and compare our results with 10 state-of-the art methods. Our results show that the proposed method significantly outperforms all 10 state-of-the-art methods both qualitatively and quantitatively. Preliminary results of our proposed work on a few video sequences were presented in [17][18] where we detected only traffic accident [17] and dominant flows [18].

To evaluate the performance of the DHIM, we compare the results with 10 state-of-the-art coherency detection methods including the lagrangian particle dynamics (LPD) [4], the mixtures of dynamic textures (MDT) [5], the motion segmentation in crowds (MSC) [19], the spatio-temporal model (STM) [20], the local trans-

lation domain (LTD) [21], the detection of coherent motion (DCM) [22], the collective motion detection (CMD) [23], the segmentation based on dynamic system (SDS) [24], the trajectory clustering approach (TCA) [25], and the thermal diffusion process (TDP) [6]. Our results show that the DIHM achieves superior coherency detection. Moreover, our proposed DIHM outperforms the compared methods in both pixel level and coherent region level analysis in terms of average particle error rates (PERs), average coherent number error (CNE) and F-score.

The rest of the paper is organized as follows. In Section 2, an overview of related work is provided. The proposed method 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

Since the methods for crowd coherency detection and anomaly detection are related to each other, we divide both of them into three broad categories based on the density of crowd. The methods targeting a maximum of two individuals are categorized as *individual level analysis*. Similarly, the methods targeting 15 and more than 15 individuals are grouped under the terms *low density flow analysis* and *high density flow analysis*, respectively. Table 1 summarizes the methods covered in this section according to their category, features and models used for representing coherency detection and anomaly detection, as well as the datasets on which these methods are evaluated.

In the individual level analysis category, Poling et al. [26] use nonlinear embedding of two-view point correspondences into a 9-dimensional space and identify the different motions by partitioning lower-dimensional subspaces. Narayana et al. [27] use optical flow orientations instead of the complete vectors and exploit the well-known property that under camera translation, optical flow orientations are independent of object depth. They introduce a probabilistic model that automatically estimates the number of observed independent motions and results in a labeling that is consistent with real-world motion in the scene. Shi et al. [28] exploit discrete cosine transform and a two-stage clustering strategy for tracked points to facilitate division of incomplete and corrupted trajectories against severe data missing and noises. Rahmati et al. [29] integrate prior knowledge in the form of weak labeling into motion segmentation. Using the example of Cerebral Palsy detection, motion patterns of infants are segmented into the different body parts by analyzing body movements.

Qin et al. [30] combine the region saliency based on entropy rate superpixel with the affinity propagation clustering algorithm to get seeds in an unsupervised manner, and use random walks method to obtain the coherency results. Zhang et al. [31] use line integral convolution and information entropy to segment the binarization image where the size of the crowd is estimated by least squares fitting to count the number of individuals in a crowd. Wu et al. [32] propose a convex texture image segmentation model by extracting Gabor features and gray level co-occurrence matrix. Then, these features are fused together to effectively construct a discriminative feature space by concatenating with each other. Li et al. [33] deal with challenges in the motion segmentation problem, including perspective effects, missing data, and unknown number of motions. The 3-D motion segmentation is first formulated from two perspective views as a subspace clustering problem. It then combines the point correspondence information across multiple image frames via a collaborative clustering stage.

In the low density flow analysis category, Ma et al. [19] propose spatio-temporal framework for motion segmentation in crowds (MSC). For this purpose, each local spatio-temporal patch is modeled as a dynamic texture. Huerta et al. [34] propose a temporal difference algorithm for objects segmentation. They fuse the knowledge from the color, edge and intensity to solve not only global and local illumination changes but also the camouflage in intensity. Mumtaz et al. [35] propose a motion segmentation approach that consists of a set of location-specific dynamic texture components, for modeling local background motion, and a set of global dynamic texture components, for modeling consistent foreground motion. For this purpose, an EM algorithm is derived and spatial constraints are applied using markov random field. Li et al. [20] use a spatio-temporal model (STM) to perform group motion segmentation. They analyze motion trajectories of multiple objects to extract consistent segments. Chan et al. [5] use the mixtures of dynamic textures (MDT) which is a spatio-temporal generative model for modeling, clustering and segmenting videos. For this purpose, video sequences are represented as observations from a linear dynamical system.

High density flow analysis methods treat the entire scene as a single entity, and are usually capable of obtaining coarser-level information, such as the identification of the main flow, disregarding local and finer information. In high density flow analysis for coherency detection, Ali et al. [4] propose a lagrangian particle dynamics (LPD) approach to detect the coherent flows. For this purpose, they use the Finite Time Lyapunov

Exponent [36] to extract the boundaries between different flow regions in the scene. Zhou et al. [22] propose coherent filtering for the detection of coherent motion (DCM) based on background clutters. Wu et al. [21] propose local-translation domain model (LTD). In this model, they transform the problem of crowd motion segmentation into a problem of scattered motion field segmentation. The optical flow is computed at the salient locations and is subsequently used to derive the evolution of segment boundaries using the Gateaux derivative of an objective function. Wu et al. [23] introduce collective motion detection (CMD) approach to characterize the global consistency for coherency detection. Zhang et al. [24] perform segmentation based on dynamical systems (SDS). Sharma et al. [25] propose a trajectory clustering (TCA) approach for segmenting crowd flow patterns. The purpose of trajectories is to capture the temporal evolution of each region. Wang et al. [6] exploit thermal diffusion process (TDP) to capture motion correlation among particles. Then a two-step clustering technique is used to detect the coherent regions.

Once the coherent regions are detected, it is also possible to detect anomalous events occurring within the flows. Most of the methods for anomaly detection fall under the category of high density flow analysis. Mehran et al. [40] detect abnormal events in terms of escape panics by exploiting the social force model. Wang et al. [43] develop the high-frequency spatio-temporal (HFST) features to detect both global and local abnormal behaviors. The wavelet transform is applied to the plane in the cuboid considering Latent Dirichlet allocation and Multiple Hidden Markov Models. Li et al. [38] and Xu et al. [42] detect anomalies in terms of panic situation and circulation of non-pedestrian entities, by considering global and local spatio-temporal patterns. A sparse reconstruction over the normal bases is proposed by Cong et al. [44] to detect abnormal events. Thida et al. [45] learn the spatial and temporal variations of local motions whereas Kaltsa et al. [37] introduce histograms of oriented swarms combined with histograms of oriented gradients. Li et al. [39] and Mahadevan et al. [7] propose a joint detector of temporal and spatial anomalies based on a video representation that accounts for both appearance and dynamics, using a set of mixture of dynamic textures models. Spatial and temporal anomaly maps are defined at multiple spatial scales that act as potentials of a conditional random field that guarantees global consistency of the anomaly judgments. Wu et al. [41] introduce the concepts of potential destinations and divergent centers to construct the corresponding class-conditional probability density

Table 1: State-of-the-art methods for coherency and anomaly detection. The methods with no descriptions in the Anomaly column are targeting only coherency detection. Methods targeting a maximum two individuals are under the category of individual level analysis, methods targeting a maximum 15 individuals are under low density flow analysis, and methods targeting more than 15 individuals are grouped under high-density flow analysis.

Ref.	Category	Features	Model	Anomaly	Dataset
Narayana et al. [27]	**Individual**	Motion	Probabilistic model	–	Hopkins SegTrack
Poling et al. [26]		Subspace clustering	Global dimension minimization	–	RAS
Qin et al. [30]		Superpixel	Region saliency	–	BSD300 Free 1000
Shi et al. [28]		Trajectories	Discrete cosine transform	–	Hopkins Berkeley
Wu et al. [32]		Gabor features	Convex texture model	–	Synthetic images Squirrel images
Chan et al. [5]	**Low-density**	Mixtures of dynamic textures	Spatio-temporal model	–	Traffic videos UCSD
Ma et al. [19]		Motion	Spatio-temporal	–	i-Lids
Huerta et al. [34]		Color and edge	Temporal difference	–	PETS2001 ATON
Li et al. [20]		Motion	Spatio-temporal model	–	GaTech
Mumtaz et al. [35]		Dynamic textures	Markov random field	–	FBDynScn
Ali et al. [4]	**High-density**	FTLE	Optical flow	Instability	UCF
Zhou et al. [22]		Motion	Spatio-temporal	–	UCF
Kaltsa et al. [37]		Motion and appearance	HOS	Non-pedestrian entities escape panic	UCSD UMN
Wu et al. [23]		Motion	Clustering	–	UCF
Li et al. [38]		Spatio-temporal patterns	Unsupervised statistical	Non-pedestrian entities escape panic	UCSD UMN
Li et al. [39]		Dynamic textures	Joint detector	Non-pedestrian entities escape panic	UCSD UMN
Mahadevan et al. [7]		Dynamic textures	Mixture models	Non-pedestrian entities	UCSD
Mehran et al. [40]		Motion magnitudes	Social force model	Escape panic	UMN
Wu et al. [41]		Motion	Bayesian	Escape panic	UMN PETS2009
Xu et al. [42]		Spatio-temporal patterns	Unsupervised statistical	Non-pedestrian entities	UCSD
Proposed	**Density independent**	Motion and orientation	DIHM	–	PETS2009 UCSD UCD

functions of optical flow. The identified divergent centers indicate possible locations at which the unexpected events occur. Krausz et al. [46] detect motion patterns based on optical flow that characterize crowd behavior in stampedes. Kim and Grauman [47] exploit a mixture of probabilistic PCA models to characterize motion patterns in the local volumes. Furthermore, a global inference by incorporating a Markov Random Field model is applied to detect anomalies locally.

The SPH is used in different applications previously. For example, Gloor et al. [48] consider the SPH for pedestrian simulation. Cohen et al. [49] use the SPH to simulate human swimming. Hieber et al. [50] use remeshed smoothed particle hydrodynamics (rSPH) to simulate human organs for computer aided surgery. In the rSPH, human organs are modeled as linear viscoelastic solids. Raney et al. [51] describe the SPH for the simulation of transportation. Cohen et al. [52] use SPH based simulations to evaluate variants of the swimming stroke technique. In competitive human swimming the dolphin kick stroke i.e. underwater undulatory swimming is used after dives. Unlike these methods, we exploit the SPH model in conjunction with our novel DIHM for crowd coherency detection.

3. Proposed Method

The proposed DIHM (Fig. 1) consists of three main steps: particles of interest collection to remove static particles, coherent region detection using the density independent hydrodynamics model and refining the coherent regions detection. These steps are explained in detail in the following sections.

3.1. Particle of interest collection

We initialize particles over the video frame. The particle initialization represents the localization of pixel positions uniformly spread over the video. We then extract a motion flow field using the Farnebäck optical flow technique [14]. For this purpose, we can use different optical flow techniques. However, the Farnebäck optical flow technique [14] does not assume the spatially homogeneous structure of the scene. Therefore, it is robust to crowd dynamics. It is worth noticing that we are only interested in the motile particles. To remove the static particles associated with non-motion regions, we explore an approach to retain only particles of interest or motile particles.

In our proposed method, we consider crowd as a system where the motion of the crowd scene is modeled in terms of its energy. Therefore, we model the motion

patterns observed in a crowded scene through the enthalpy measure [15]. The attractive features of enthalpy model are two-fold. Firstly, it takes into account the motion information of the surrounding particles for the purpose of retaining only motile particles. Secondly, a compact threshold range between $0 - 1$ is required for removing the static particles instead of considering an arbitrary threshold. To this end, the initialized particles are tracked over time using the optical flow technique [14]. Subsequently, we extract motion information from each frame in terms of velocity magnitudes.

We employ the enthalpy model [15] to retain motile particles that contribute to the coherency detection. In our model, we capture the motion properties of each particle by taking into account its neighborhood. In this way, we treat particles as constituents (sub-populations) of the large crowd, each having its own motion properties. We thus have the possibility to analyze the interactive behavior between sub-populations, in the spatial neighborhood, which have discriminative characteristics represented by the enthalpy model:

$$H = U + pV \qquad (1)$$

where, U is the internal energy, p is the pressure, and V is the volume of the system. The internal energy U of a system is the energy contained within the system. However, we use the kinetic energy ($\frac{1}{2}mv^2$) in term of internal energy, since we are only interested in moving particles. *Pressure* is defined as $p = Force(F)/Area(A)$ and *Force* is $F = mass(m) * acceleration(a)$. For acceleration (a), we compute the average velocity $\langle v \rangle$ in the spatial neighborhood over time. The area A represents the total number of particles in the spatial neighborhood. Mass (m) and volume (V) of each particle may be affiliated with its contribution in the corresponding sub-population, in the spatial neighborhood. However, we set them equal to 1 in our model to maintain consistency. Thus, our enthalpy model is formulated as:

$$H = \frac{1}{2}mv^2 + \left(\frac{\partial \langle v \rangle}{\partial t}\right)\left(\frac{1}{A}\right) \qquad (2)$$

In Eq. (2), H is a threshold to remove static particles. m, v, $\langle v \rangle$, and A are the mass, the temporal velocity, the average velocity, and the area, respectively. We calculate the temporal velocity in the consecutive frames. Fig. 2 highlights the moving particles associated with the motion regions using the enthalpy model.

3.2. Density independent hydrodynamics model (DIHM) for coherency detection

Our proposed DIHM is based on the smoothed particle hydrodynamics (SPH) [9] which is a fluid dynamics
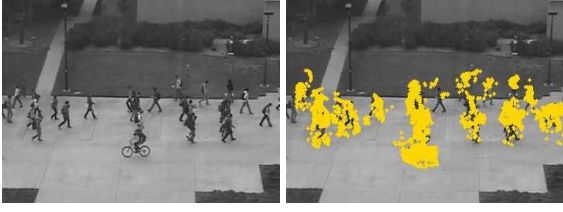
Figure 2: Particle collection via the Enthalpy model. Frame from a video sequence (first column). We can clearly observe how particles that are associated with the motion regions are preserved (second column).

based method. It has been used in many other fields including astrophysics, volcanology, and oceanography. The DIHM exploits a finite set of discrete particles to represent the state of a system. This method deals with fixed particles spatially and can evolve with the crowd dynamics over time. Therefore, the coherency of crowd flow can be naturally modeled using the DIHM. For this purpose, we first collect the orientation information in term of direction. Without loss of generality, we select four different directions quantized with a step size of 90 degrees. We consider particles that are carried with the crowd flow. Then we evaluate their contributions through a weighted average based on smoothed kernel function. Thus it allows encoding the crowd dynamics, adaptively.

In the DIHM, we model the crowd flow into a set of discrete elements in terms of particles. In fact, each particle is represented by its physical properties. In our case, the physical properties of each particle are the motion pattern and density. The former represents the velocity magnitude and the latter represents the number of surrounding particles. Furthermore, these particles have a spatial distance, over which their properties are smoothed by a kernel function. This means that the physical properties of any particle can be modeled by considering the relevant properties of all the particles which lie within the range of the kernel. The contributions of each particle are weighted according to their distance from the particle of interest (central particle), and their density. Mathematically, this is determined by the kernel function $K(r_c - r_i, \lambda)$ as formulated in Eq. (3).

$$K(r_c - r_i, \lambda) = |K_t(r_c - r_i, \lambda) - K_{t-1}(r_c - r_i, \lambda)|$$
$$K_t(r_c - r_i, \lambda) = \frac{315}{64\pi(\lambda)^9}((\lambda)^2 - \|r_c - r_i\|^2)^3 \quad (3)$$

where r_c is the position of the central particle and r_i is

the position of each surrounding particle. t and $t-1$ represent consecutive frames. In Eq. (3), λ is the smoothing length that defines the influence area of the kernel function K. In fact, the smoothing length λ presents a set of neighboring particles associated with a central particle under consideration. The kernel function is like a basis function that is used to represent any quantity by the summation of the quantities at the nearby points multiplied by a weighting function. The kernel function is nonnegative, even, and monotonically decreasing with the increase of distance from the central particle. It determines the influence of particles on one another. Therefore, it provides compact support that is controlled by the influence area λ.

We model the DIHM by considering the Navier-Stokes equation of motion as formulated in Eq. (4).

$$\frac{\partial \mathbf{v}}{\partial t} = -\nabla P + \rho g + \mu \nabla^2 V \quad (4)$$

where ∇, P, ρ, g, μ, and V represent the derivative operator, pressure, density, acceleration, viscosity, and velocity, respectively. Temam et al. [53] designate the three terms on the right hand side of Eq. (4) as the pressure force $F_{press}(-\nabla P)$, the external force $F_{ext}(\rho g)$ and the viscous force $F_{vis}(\mu \nabla^2 V)$, which are formulated as:

$$F_{press} = \sum_{i \in N} m_i K(r_c - r_i, \lambda)$$
$$F_{ext} = \sum_{i \in N} m_i \nabla K(r_c - r_i, \lambda) \quad (5)$$
$$F_{vis} = \sum_{i \in N} m_i \mu v \nabla^2 K(r_c - r_i, \lambda)$$

The pressure force F_{press} upholds the weighted motion pattern of all the particles in the scene. The viscous force F_{vis} considers the contributions of spatial correlation among the particles and the external force F_{ext} ensures that the particles with significant motion magnitude receives more attention. In Eq. (5), m is the mass of a particle, μ is the viscosity, and v is the accumulated velocity of all the particles in the region. We set the mass of each particle equal to 1 for consistency. The viscosity at each time instant is calculated as $\mu = \sum_{i \in N} |r_c - r_i|$ and the accumulated velocity is calculated as $v = \sum_{i \in N} (v_{r_c} + v_{r_i})$. $\nabla K(r_c - r_i, \lambda)$ and $\nabla^2 K(r_c - r_i, \lambda)$ are the gradient and the Laplacian of the kernel function $K(r_c - r_i, \lambda)$ as formulated in Eq. (6).

$$\nabla K(r_c - r_i, \lambda) = \frac{-45}{\pi(\lambda)^6}(\lambda - \|r_c - r_i\|)^2 \frac{r_c - r_i}{\|r_c - r_i\|}$$
$$\nabla^2 K(r_c - r_i, \lambda) = \frac{45}{\pi(\lambda)^6}(\lambda - \|r_c - r_i\|) \quad (6)$$

Considering the three forces, Eq. (4) can be formulated as

$$\frac{\partial \mathbf{v}}{\partial t} = F_{press} + F_{ext} + F_{vis} \quad (7)$$

Eq. (7) can be rearranged as formulated in Eq. (8)

$$\left(\frac{\frac{\partial \mathbf{v}}{\partial t}}{F_{press}} - 1\right)K = \frac{F_{vis}}{F_{press}} + \frac{F_{ext}}{F_{press}}K \quad (8)$$

The term $(\frac{\frac{\partial \mathbf{v}}{\partial t}}{F_{press}} - 1)K$ represents the motion dynamics that models the DIHM magnitude. Thus, to detect a coherent region in the crowd flow, the DIHM magnitude using the three forces is formulated in the Eq. (9)

$$DIHM_{Mag} = \frac{F_{vis}}{F_{press}} + \frac{F_{ext}}{F_{press}}K \quad (9)$$

The DIHM magnitude $DIHM_{Mag}$ in Eq.(9) changes as a function of the pressure, external and viscous forces. Particles satisfying Eq. (9) construct coherent regions $\mathbf{R} = [\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_M]$. However, the orientation information of each particle is used as a prior. It means that surrounding particles are considered in the coherent region, only if their direction of motion correlate with the direction of central particle as depicted in Fig. 3. In fact, we consider the particles one by one starting from the top left corner of a frame. Each particle under consideration at a given time is the central particle. We calculate the DIHM magnitude for each central particle. If the DIHM magnitude for a central particle fall in the range from 0 to 1 that means the central particle satisfies Eq. (9). A central particle satisfying Eq. (9) would construct a coherent region. If the DIHM magnitude does not fall in this range, the central particle either belongs to the background or it could be the neighboring particle of any other central particle. Furthermore, once a set of particles are determined to be the neighboring particles of a central particle, they are not associated with any other central particle. It means that a set of neighboring particles linked with a central particle would not be analyzed again for another central particle.

The ease with which our proposed DIHM can cope with the changing crowd densities, as well as an automatically adaptive nature, are unmatched in other methods [4][5][6][19][21]. Our DIHM is based on the SPH
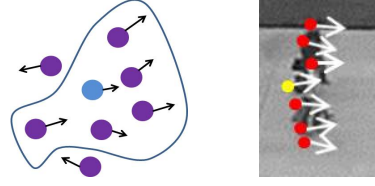

Figure 3: Synthetic example of coherent region (left column), moving particles are grouped together using the proposed DIHM method (right column).

which is a mesh-free method. The mesh-free method is based on interaction of each particle with all its neighbors. Each particle can have variable number of neighbors. Due to the mesh-free nature, SPH can easily deal with complicated physical layout of crowd scenes and different regions of crowd that are completely irregular in term of densities.

To circumvent the unwanted effects of changing density, the key is to consider the correlations among particles to represent coherent regions in crowd scenes. Our DIHM explores the correlation of structures in crowd scenes from particle level to crowd level. This is to mitigate the effects of the changing density for an effective crowd coherency detection. The utilization of our modeling obviates the difficulty to segregate particles in crowd due to the variation of crowd density over time. For this purpose, the evolution of boundaries of different coherent regions is derived from the kernel function $K(r_c - r_i, \lambda)$ of our model. Our DIHM model acquires optimized boundary adherence of coherent crowd regions due to the three forces (F_{press}, F_{ext} and F_{vis}). Since our proposed DIHM is leveraging on the kernel function and the three forces, conformation to varying crowd densities is achieved.

3.3. Refining the detected coherent regions

To refine the coherency detection, different methods can be mulled over. However, we consider the MLSC [16] which is an unsupervised method driven by low computational overheads. Additionally, the method takes into account the coherency in terms of both appearance and motion. For this purpose, the MLSC exploits three types of neighborhood affinities namely intra-frame, inter-layer, and inter-frame. The intra-frame affinity means that regions sharing a common boundary within a frame may be connected. The inter-layer affinity means that regions from the same frame but in adjacent layer may be neighbors as well. Similarly, the inter-frame affinity states that regions at the same layer but in consecutive frames may be connected

by edges representing the temporal relation. In our case, the MLSC is governed by the coherent regions detected by our proposed DIHM. It means that the MLSC performs on the initial coherency information. This initial information is provided to the MLSC in terms of detected coherent regions by the DIHM.

The MLSC [16] method combines both spatial and temporal consistencies that depend on the graph structure and affinities in terms of edge weights. Moreover, it is based on the layer extraction [54][55] that segments regions first, and matches different presegments across the consecutive frames. Each presegment corresponds to a vertex in the graph. The inter-layer affinities put together coarse but stable information, and the affinities between consecutive frames can provide local grouping cues across temporal range. Subsequently, the obtained affinities are used to get the coherency of the crowd flow. Each layer corresponds to the initial segmentation in the multi-layer graph $G^* = (V^*, E^*)$. The nodes $V^* = \cup \{V_t^{(l)}\}_{t=1...T}^{l=1...L}$ are the union of multi-parametric presegment of a video clip, where the node subset $V_t^{(l)}$ corresponds to $Q_t^{(l)}$ image segments of frame t in layer l. The undirected edges in the graph G^* are considered based on the aforementioned three types of neighborhood affinities. According to the intra-frame neighborhood affinity, regions r_i and r_j sharing a common boundary within a frame at the same layer are connected. The similarity between the regions is measured as a Gaussian function of the χ^2 distance between their histogram features as formulated in Eq. (10),

$$w_{ij} = exp(-\tau \cdot D_\chi(g_c(r_i), g_c(r_j))), r_j \in Q^s(r_i)) \quad (10)$$

where $g_c(r_i)$ and $g_c(r_j)$ are the color histograms. The τ is a constant controlling strength of the weight, and $Q^s(r_i)$ is the spatial neighborhood of region r_i. Regions from the same frame in the adjacent layers may be neighbors. Therefore, the affinity between a lower-layer region r_k and its neighboring higher-layer region r_p is formulated in Eq. (11).

$$w_{kp} = \frac{|\tilde{r}_k \cap r_p|}{|r_k|} \quad (11)$$

where $\tilde{r}_k$ is the spatial correspondence region of r_k and $|r_k|$ is the number of pixels in region r_k. The affinity between regions at the same layer but in consecutive frames resort to similarity between their contour shapes [56] since neighboring region boundaries render important information. The similarity between r_j and its temporal neighbor r_q is formulated in Eq. (12).

$$w_{jq} = exp(-\beta \cdot D_\chi(g_s(r_j), g_s(r_q))), r_q \in Q^t(r_j)) \quad (12)$$

To sum up, the edge weights in the multi-layer graph can be formulated in Eq. (13).

$$w_{ij}^\star = \begin{cases} z_1 exp(-\tau \cdot D_\chi(g_c(r_i), g_c(r_j))), & r_j \in Q^s(r_i) \\ z_2 \frac{|\tilde{r}_i \cap r_j|}{|r_i|}, & r_j \in Q^l(r_i) \\ z_3(1 - \frac{1}{2}D_\chi(g_s(r_i), g_s(r_j))), & r_j \in Q^t(r_i) \\ 0, & \text{otherwise} \end{cases} \quad (13)$$

where z_1, z_2, and z_3 are parameters to balance the weights and their sum equals to unity. τ and β vanish away due to the redundancy with z_1 and z_2. The degree matrix D of the affinity matrix $W = [w_{ij}^*]$ is a diagonal matrix with $d = \sum_j W_{ij}$, and the normalized affinity matrix $L = D^{-1}W$. The final coherency is obtained from the clustering result of the lowest subgraph, which is more coherent and keeps necessary information.

4. Experimental results

We evaluate the performance of our proposed DIHM method for coherency detection on three datasets. These include the publicly available benchmark PETS2009 [57], UCSD [7] and our UCD [58] datasets. The PETS2009 dataset consists of low-density to high-density crowd sequences presenting different activities for people tracking, flow analysis, and event recognition. We consider all the views (View1 to View4) from both L1 and L2 subsets from PETS2009 dataset. The UCSD dataset consists of two subsets, ped1 and ped2, corresponding to two different views. Both subsets represent surveillance videos captured by a fixed camera overlooking pedestrian walkways. In Ped1, people are moving towards and away from the camera, with some perspective distortion and ped2 contains videos of people moving parallel to the camera. The resolutions of Ped1 and Ped2 are 158x238 and 240x360, respectively. The normal event appearing in the dataset is sequences of pedestrians on the walkways, with a varying density from sparse to very dense. The non-pedestrian entities include cyclists, skaters, vehicles, people walking on a lawn. The appearance of all non-pedestrian entities occurs naturally, i.e., they were not staged or synthesized for data set collection. The video footage of each scene is divided into clips of 120-200 frames. The UCD dataset contains four low-density video sequences representing flows of students moving outdoors across two buildings.
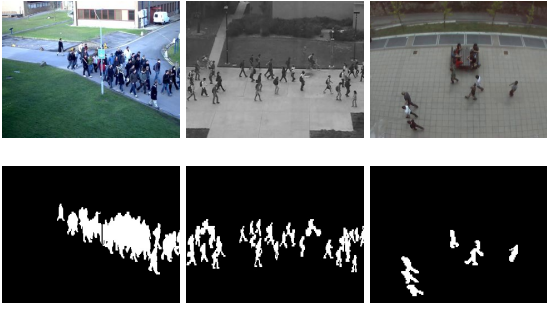
Figure 4: Ground truth (Pixel-based). First row shows input frames from the PETS2009, UCSD and UCD datasets, respectively. The second row shows the corresponding ground truth obtained using RATSNAKE annotation tool [59].
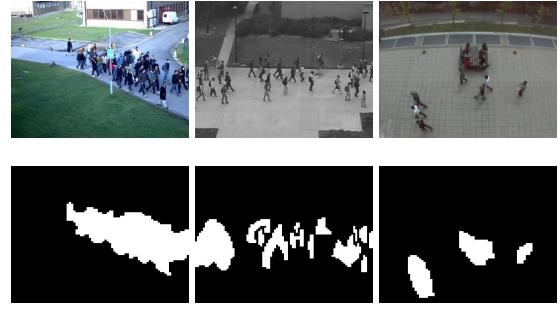

Figure 5: Ground truth (Coherent region-based). First row shows input frames from the PETS2009, UCSD and UCD datasets, respectively. The second row shows the corresponding ground truth obtained using RATSNAKE annotation tool [59].

We compare the performance of our proposed DIHM method with 10 closely related state-of-the-art approaches: the lagrangian particle dynamics (LPD) [4], the mixtures of dynamic textures (MDT) [5], the motion segmentation in crowds (MSC) [19], the spatio-temporal model (STM) [20], the local translation domain (LTD) [21], the detection of coherent motion (DCM) [22], the collective motion detection (CMD) [23], the segmentation based on dynamic system (SDS) [24], the trajectory clustering approach (TCA) [25], and the thermal diffusion process (TDP) [6]. For quantitative evaluation of coherency detection, the average particle error rates (PERs) and the average coherent number error (CNE) are calculated to measure the overall accuracy according to the TDP approach [6]. Additionally, we also calculated the F-score as a performance measure for comprehensive analysis.

The segmentation masks are annotated and compared against the ground-truth mask. We have manually annotated individuals and coherent regions in each video, from all three datasets, using the RATSNAKE annotation tool [59]. It is worth noticing that the annotation of all the three datasets for ground truth generation is a subjective task. Therefore, we requested 20 participants to annotate all the three datasets for both the pixel-level and coherent region-level analysis. For pixel-level, we calculated the ground truth for each video sequence by taking the intersection of all the 20 annotations. It means that we consider a pixel in the ground truth if it is annotated by all the 20 participants. For coherent region-level, we consider the ground truth for each video sequence where majority of the participants came up with the same annotation. It means that a region is

considered coherent if it is annotated by the majority of the participants. For example, the input video frames and the ground truth video frames (annotations of individuals for pixel-level analysis), from the PETS2009, UCSD, and UCD datasets, are shown in the top and bottom rows of Fig.4, respectively. Similarly, the input video frames and the ground truth video frames (annotations of coherent regions for region-level analysis), from the PETS2009, UCSD, and UCD datasets, are shown in the top and bottom rows of Fig.5, respectively. The same annotation tool is used to generate the binary masks for the reference methods and the proposed method. To calculate the PERs, CNE, and F-score, we annotated each tenth frame of each video sequence since the annotation process is very time consuming.

The qualitative performance of our proposed method is presented in Fig. 6. Results for one sequence from each dataset are presented. The first, second, and third rows depict results for the reference methods and our method for the PETS2009, UCSD, and UCD datasets, respectively. For the purpose of visualization, the frames are overlaid by color segmentation results. First column presents the sample frames taken from the original video sequences, while second to last columns illustrate the results obtained using the LPD [4], the MDT [5], the MSC [19], the STM [20], the LTD [21], the DCM [22], the CMD [23], the SDS [24], the TCA [25], the TDP [6], and the proposed approach, respectively. It can be seen that our method detects coherent regions accurately in all rows of the last column. The LPD [4], the MDT [5], the CMD [23], the SDS [24], and the TDP [6] resulted in over-segmentation. The LPD [4] segments the motion also when the boundaries
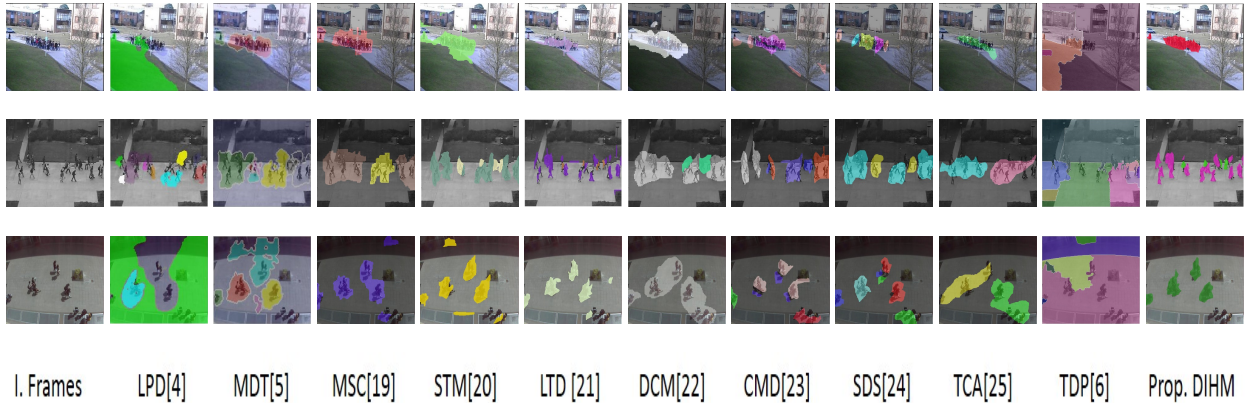
9

Figure 6: Comparison of DIHM with the state-of-the-art coherency detection methods: First column of each row represents input frames (I. Frames) from PETS2009, UCSD and UCD, respectively. The proposed DIHM achieves more refined and accurate segmentation. Moreover, the effects of oversegmentation are also minimized compared to other methods.

in the optical flow field are not consistent. In fact, when the optical flow computation is not accurate due to the lack of coherence in motion, the boundaries may be discontinuous. Furthermore, the merge operation based on Lyapunov divergence is mainly suitable for combining adjacent segments, resulting also in this case in over-segmentation when the density of crowd changes over time. The MDT [5] and the MSC [19] cannot decompose videos into coherent regions due to its inability to take into account the dynamic nature of the underlying motion. In fact, it is important to capture the dimensions along which the video is statistically homogeneous to avoid over-segmentation. The coherent filtering of the DCM [22], the collective density clustering of the CMD [23], and the similarity calculation method of the SDS [24] merge inconsistent regions. Therefore irregular segments in the crowded scenes are constituted. The pixel-wise segmentation process of the TCA [25] cannot cope with the temporal evolution of the coherent regions. Therefore, large segments in the crowded scenes are produced. The TDP [6] uses the diffusion process to consolidate the motion flow field. However, the process also merges the static or non-motion regions, thus losing coherency. The LTD approach [21] and the STM approach [20] cannot handle the coherence and accuracy of the optical flow field in case of varying crowd motion. Furthermore, the LTD [21] and the TDP [6] cannot properly propagate consistent motion information to the neighboring regions using the thermal diffusion process and the local translation domain modeling, respectively.

For quantitative performance, we calculated the average PERs, CNE, and F-scores for the three datasets and reported them in Table 2, Table 3, and Table 4, respectively. We outperform all the reference methods due to strong capability of DIHM to detect coherent regions. Thanks also to the enthalpy model for collecting particle of interest and removing particles associated with static regions. Our method results in smaller PERs and CNE errors for the three datasets as shown in Table 2, and Table 3, respectively. All the reference methods ignore change in the crowd flow that holds only if the flow is consistent over time. As the flow changes from higher density to lower density or vice versa, the performance declines. In fact, such changes result into different disconnected regions, thus depriving of coherency. Moreover, our method shows higher average F-scores comparing against the reference methods for the three datasets as shown in Table 4. The reference methods cannot handle many-fold flows representing different dynamic pieces in a scene. The lowest performance of the LPD [4] is due to the the FTLE field that cannot partition the significant differences in the coherent regions. Also, the reference methods cannot handle the condition when new flow emerges rapidly since accommodating the new change with the available information does not exist.

4.1. Density independence of our proposed DIHM

In real-world scenarios the crowd density changes over time. Therefore, it is significant to cope with the effect of changing density to detect the coherency. The

Table 2: Quantitative analysis. The average particle error rates (PERs) for the reference methods and our proposed method for the three datasets, PETS2009, UCSD, and UCD, are presented.

Datasets	LPD[4]	MDT[5]	MSC[19]	STM[20]	LTD[21]	DCM[22]	CMD[23]	SDS[24]	TCA[25]	TDP[6]	Prop. DIHM
PETS2009	37.27	13.98	22.35	35.77	19.78	15.48	34.11	21.35	22.28	33.10	10.86
UCSD	40.29	18.35	28.45	36.48	22.82	18.02	38.43	24.58	21.19	37.34	13.10
UCD	41.82	17.46	26.07	38.94	23.66	16.87	36.98	22.97	20.48	36.87	11.56
Average	39.79	16.59	25.62	37.06	22.08	16.79	36.51	22.96	21.32	35.77	11.84

Table 3: Quantitative analysis. The average coherent number error (CNE) for the reference methods and our proposed method for the three datasets, PETS2009, UCSD, and UCD, are presented.

Datasets	LPD[4]	MDT[5]	MSC[19]	STM[20]	LTD[21]	DCM[22]	CMD[23]	SDS[24]	TCA[25]	TDP[6]	Prop. DIHM
PETS2009	6.72	3.15	5.05	4.88	3.11	3.18	5.99	4.12	4.76	6.55	2.00
UCSD	7.81	3.74	5.66	5.01	3.89	3.79	6.07	4.64	4.59	6.96	2.63
UCD	7.96	3.63	4.98	5.74	3.56	3.44	5.87	4.96	4.09	7.48	2.44
Average	7.49	3.51	5.23	5.21	3.52	3.47	5.98	4.57	4.48	6.99	2.36

Table 4: Quantitative analysis. F-scores for the reference methods and our proposed method for the three datasets, PETS2009, UCSD, and UCD, are presented.

Datasets	LPD[4]	MDT[5]	MSC[19]	STM[20]	LTD[21]	DCM[22]	CMD[23]	SDS[24]	TCA[25]	TDP[6]	Prop. DIHM
PETS2009	0.27	0.40	0.29	0.28	0.38	0.37	0.26	0.30	0.28	0.32	0.44
UCSD	0.29	0.39	0.28	0.32	0.40	0.39	0.28	0.27	0.28	0.33	0.43
UCD	0.28	0.40	0.29	0.31	0.42	0.39	0.24	0.30	0.31	0.35	0.45
Average	0.28	0.39	0.29	0.30	0.40	0.38	0.26	0.29	0.29	0.33	0.44

state-of-the-art methods ignore this effect since overcoming this effect of changing density is a challenging problem. We propose the DIHM method that absorbs the effect of changing density seamlessly. To illustrate this fact, three frames (L1, View1, frame number 13, 46, and 86) from a video sequence of PETS2009 dataset are depicted in Fig. 7. Comparing to the first column, the density changes in the second column. Similarly, the density changes in the third column comparing to the second column. We also consider three frames from a video frame from both UCSD (ped2, video sequence 1, frame number 10, 71, and 116) and UCD (Video sequence 1, frame number 146, 272, and 342) datasets. We then calculate the PERs, CNE, and F-score for the three frames from each video sequence and reported them in Table 5. The PERs, CNE, and F-score are presented in the columns and the frames are presented in the rows of each cell of the Table 5. It can be seen that our proposed DIHM method presents significant

Figure 7: Density independence of our DIHM. A consistent set of people are shown in the first column. The density changes in the second and third column.

performance comparing against the reference methods when the density is changing in case of second and third frames of each video sequence for all the three datasets.

4.2. Sensitivity Analysis

To demonstrate the robust performance of our approach for the three datasets, we tested our method us-

Table 5: Density independence of our DIHM. The PERs, CNE, and F-scores (shown in columns of each cell of the Table) for the three frames (in rows of each cell of the Table) of each video sequence from the three datasets, PETS2009, UCSD, and UCD, are presented.

Datasets	LPD[4]	MDT[5]	MSC[19]	STM[20]	LTD[21]	DCM[22]	CMD[23]	SDS[24]	TCA[25]	TDP[6]	DIHM
PETS 2009	35-6.0-.29	14-2.5-.42	22-5-.31	34-4-.30	19-2.6-.39	15-2.2-.38	33-5-.28	19-3-.31	20-4-.30	31-5.4-.34	9.5 -2-.45
	38-7.5-.26	16-3.7-.39	23-6-.28	37-6-.25	21-3.7-.36	17-3.3-.36	35-7-.27	22-4-.29	22-5-.28	35-6.9-.30	10 -2-.45
	39-7.0-.25	15-3.9-.40	23-5-.27	38-5-.28	21-3.6-.36	19-3.6-.35	34-6-.26	21-4-.29	23-5-.27	35-7.0-.29	9.8 -2-.46
UCSD	40-7.9-.30	16-3.0-.41	26-5-.30	34-4-.34	20-3.0-.42	17-3.0-.41	36-5-.30	22-4-.29	19-4-.30	35-6-.35	12-2-.44
	42-8.5-.28	19-4.1-.38	27-5-.29	36-5-.30	23-4.2-.40	19-4.3-.38	38-6-.28	25-5-.26	21-4-.29	37-7-.33	13-2-.43
	43-8.9-.28	18-4.0-.38	27-4-.29	37-5-.29	23-4.3-.39	21-4.5-.37	38-6-.28	24-5-.27	22-5-.28	36-7-.32	12-2-.44
UCD	39-7.2-.31	16-3.4-.41	29-6-.29	39-5-.33	22-3.0-.42	16-2.9-.42	35-5-.26	21-4-.33	19-3-.33	33-7-.37	12-2.8-.45
	43-9.1-.27	19-4.1-.39	27-6-.27	42-6-.30	25-4.2-.30	19-3.8-.39	39-6-.23	25-5-.28	21-4-.30	37-8-.33	11-2.3-.46
	41-8.4-.28	17-3.9-.40	27-6-.26	41-6-.31	25-4.5-.30	19-3.7-.38	36-5-.23	24-5-.29	19-3-.30	39-8-.32	11-2.2-.45

Table 6: Configuration set. For sensitivity analysis for our proposed method, 20 different configurations are listed based on enthalpy threshold (H), neighboring area (A), smoothing length (λ), DIHM magnitude (DM), MLSC parameters ($z1$, $z2$, $z3$), and the particle grid (PG). In the neighboring area (A) and particle grid (PG), 3x and 0x represent 3x3 and 0x0, respectively.

Param.	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20
H	.1	.2	.3	.4	.5	.6	.7	.8	.9	1	1	.9	.8	.7	.6	.5	.4	.3	.2	.1
A	3x	5x	7x	9x	11x	13x	15x	17x	19x	21x	3x	5x	7x	9x	11x	13x	15x	17x	19x	21x
λ	2	4	6	8	10	12	14	16	18	20	2	4	6	8	10	12	14	16	18	20
DM	.1	.2	.3	.4	.5	.6	.7	.8	.9	1	1	.9	.8	.7	.6	.5	.4	.3	.2	.1
z1	.1	.2	.3	.4	.8	.7	.6	.5	.4	.2	.2	.4	.5	.6	.7	.8	.4	.3	.2	.1
z2	.2	.3	.4	.5	.1	.2	.3	.3	.3	.4	.4	.3	.3	.3	.2	.1	.5	.4	.3	.2
z3	.7	.5	.3	.1	.1	.1	.1	.2	.3	.4	.4	.3	.2	.1	.1	.1	.1	.3	.5	.7
PG	0x	1x	2x	3x	4x	5x	6x	7x	8x	9x	9x	8x	7x	6x	5x	4x	3x	2x	1x	0x

ing 20 different parameter configuration set as listed in Table 6. These configurations are encoded in the experiments considering different enthalpy thresholds (H), neighboring areas (A), smoothing lengths (λ), DIHM magnitudes (DM), MLSC parameters ($z1$, $z2$, $z3$), and particle grid (PG). All these parameters are independent of each other and their values are chosen randomly. For this purpose, 10 different neighboring areas (A) and smoothing lengths (λ) are taken into account where the enthalpy threshold (H) and the DIHM magnitude can be varied in the range from 0 to 1. Additionally, 10 different MLSC parameter configurations (each sum equals to unity) are listed. In the neighboring area (A) and particle grid (PG), 3x and 0x represent 3x3 and 0x0, respectively. We use this compact representation due to space limitation. The PG 0x means that we are using all pixel positions. Similarly, 1x means one pixel gap between two consecutive particles both vertically and horizontally. To this end, 0x is a dense PG and 9x is a sparse PG.

In Figure 8, gradual improvements in the performances in term of PERs can be noticed from configuration C1 to C4, and configuration C11 to C14 for all the three datasets. However, declines in the performances are significant from configuration C4 to C10 and configuration C14 to C20. In Figure 9 and Figure 10, similar changes in the performances for the same configurations can be noticed for CNE and F-score for all the datasets, respectively. In configuration C1 to C4 and configuration C11 to C14, expanding the neighboring area (A) from 3x3 to 9x9 and the smoothing length (λ) from 2 to 8 improve the performance. Hence, it is worth to con-
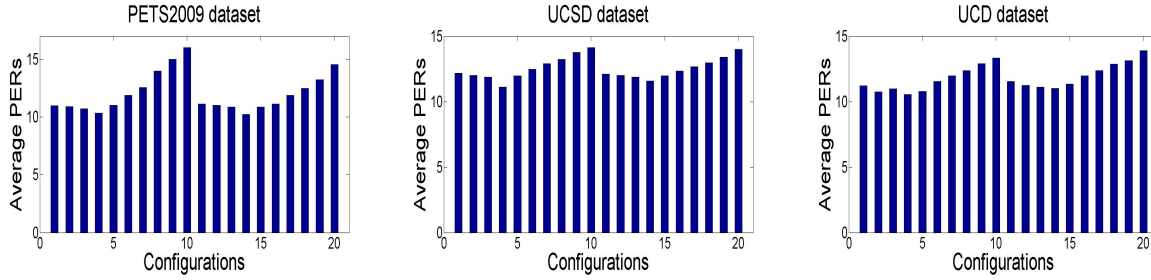
Figure 8: Particle error rates. Average PERs for our proposed method for the three datasets, PETS2009, UCSD, and UCD, are presented. The variations in the results for all the three datasets are not significant except from C4 to C5, C10 to C11, C14 to C15. The declines in the performances from C4 to C10 and C14 to C20 are due to the changes in the neighboring area (A), the smoothing length (λ), and the particle grid (PG).
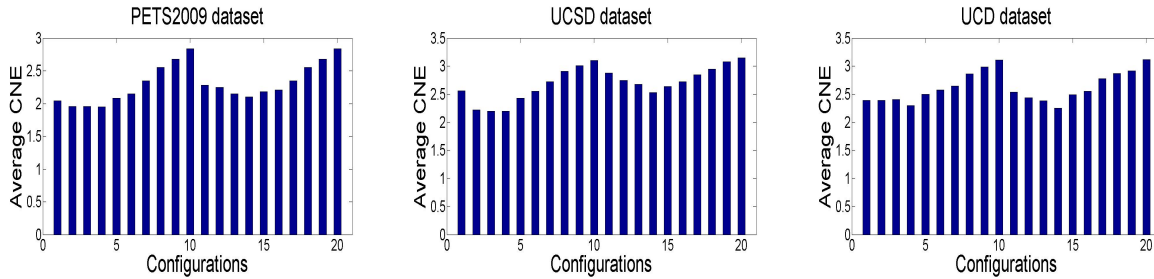

Figure 9: Coherent number error. Average CNE for our proposed method for the three datasets, PETS2009, UCSD, and UCD, are presented.The variations in the results for all the three datasets are not significant except from C4 to C5, C10 to C11, C14 to C15. The declines in the performances from C4 to C10 and C14 to C20 are due to the changes in the neighboring area (A), the smoothing length (λ), and the particle grid (PG).

sider more neighboring particles. However, the performance declines by expanding the neighboring area (A) from 9x9 to 11x11 and smoothing length (λ) from 8 to 10, as shown in the configuration from C4 to C5 and C14 to C15. In fact, the particles located far away do not influence the coherency of the central particle. Similarly, changing the particle grid (PG) from 0x0 to 3x3 does not change the performance significantly. However, changing it from 3x3 to 9x9 declines the performance. Therefore, the performance of our method does not change significantly by changing other parameters except the neighboring area (A), the smoothing length (λ), and the particle grid (PG).

4.3. Computational complexity

The computational complexity of our DIHM method and 10 reference methods are provided in Table 7. The complexity in term of average number of seconds per frame over all the datasets is calculated. It turns out that the TDP [6] and the MDT [5] requiring 28 seconds and

26 seconds per frame are very slow due to high processing requirements. The STM [20], the CMD [23], and our proposed DIHM process each frame in 04 seconds on average. The DCM [22] processes each frame in 03 seconds on average at the cost of coherency detection performance. Thus, our proposed DIHM method presents comparable performance with the DCM [22] in term of computational complexity.

All experiments are conducted on a 16GB RAM computer with a 3.5 GHz CPU, running the algorithms on a C++ platform. It is worth noticing that these algorithms are not optimized. Therefore, the computational complexities of these methods can be reduced with optimized implementation.

4.4. Limitations of our method

Our proposed method performs coarse-level instead of fine-level segmentation. The fine-level segmentation is necessary to segment moving objects subject to their motion in order to distinguish them. However, crowded
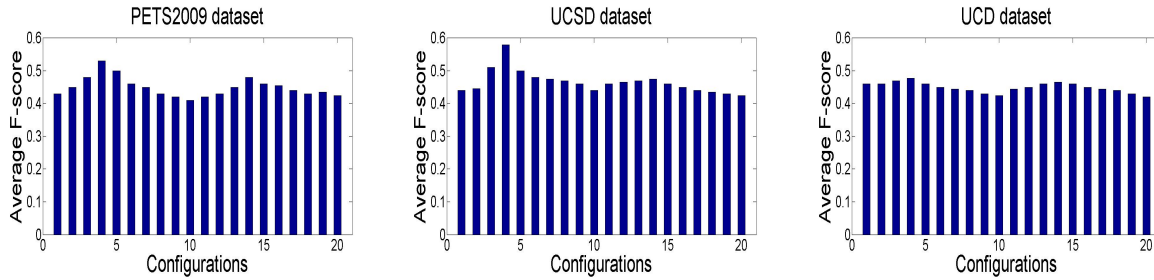
13

Figure 10: F-score. Average F-score for our proposed method for the three datasets, PETS2009, UCSD, and UCD, are presented. The variations in the results for all the three datasets are not significant except from C4 to C5, C10 to C11, C14 to C15. The declines in the performances from C4 to C10 and C14 to C20 are due to the changes in the neighboring area (A), the smoothing length (λ), and the particle grid (PG).

Table 7: Computational overheads. The computational complexity of the reference methods and the proposed DIHM method are listed in term of number of seconds required to process a video frame.

Methods	Sec. per frame	Methods	Sec. per frame
LPD [4]	21	CMD[23]	04
MDT [5]	26	SDS[24]	05
MSC[19]	19	TCA[25]	07
STM [20]	04	TDP [6]	28
LTD [21]	05	Proposed	04
DCM[22]	03	DIHM	

scenes consist of different occluded objects sharing similar movement patterns. In such situations, modeling the movement of each individual is hard as well as worthless if the interest is lying in the movement of crowd as a whole. Therefore, coarse-level segmentation in crowded scenes renders sufficient information for further processing, e.g., detecting sudden change in term of anomaly identification.

5. Conclusion

We presented a novel method for crowd coherency detection. We demonstrated the robustness of our proposed method considering three benchmark datasets namely PETS2009, UCSD, and UCD. The performance of our method has been shown against 10 closely related state-of-the-art methods. Furthermore, we presented the sensitivity of our proposed approach based on 20 different configurations which showed that the performance of our coherency detection method is not affected by

the density changes. As a future work, we would extend our method to detect a set of anomalies on top of the detected coherency information.

References

[1] W. Liu, R. W. Lau, D. Manocha, Robust individual and holistic features for crowd scene classification, Pattern recognition, Elsevier PR 58 (2016) 110–120.

[2] B. Zhou, X. Tang, X. Wang, Learning collective crowd behaviors with dynamic pedestrian-agents, International journal of computer vision, Springer IJCV 111 (1) (2015) 50–68.

[3] J. Shao, K. Kang, C. C. Loy, X. Wang, Deeply learned attributes for crowded scene understanding, in: International conference on computer vision and pattern recognition, IEEE CVPR, 2015, pp. 4657–4666.

[4] S. Ali, M. Shah, A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis, in: International conference on computer vision and pattern recognition, IEEE CVPR, (2007), pp. 1–6.

[5] A. B. Chan, N. Vasconcelos, Modeling, clustering, and segmenting video with mixtures of dynamic textures, Transactions on pattern analysis and machine intelligence, IEEE PAMI 30 (5) (2008) 909–926.

[6] W. Lin, Y. Mi, W. Wang, J. Wu, J. Wang, T. Mei, A diffusion and clustering-based approach for finding coherent motions and understanding crowd scenes, Transactions on Image Processing, IEEE TIP 25 (4) (2016) 1674–1687.

[7] V. Mahadevan, W. Li, V. Bhalodia, N. Vasconcelos, Anomaly detection in crowded scenes, in: IEEE CVPR, 2010, pp. 1975–1981.

[8] R. Mehran, B. E. Moore, M. Shah, A streakline representation of flow in crowded scenes, in: European conference on computer vision, Springer ECCV, 2010, pp. 439–452.

[9] J. Monaghan, Smoothed particle hydrodynamics and its diverse applications, Annual Review of Fluid Mechanics 44 (2012) 323–346.

[10] A. Colagrossi, M. Landrini, Numerical simulation of interfacial flows by smoothed particle hydrodynamics, Journal of computational physics, Elsevier JCP 191 (2) (2003) 448–475.

[11] M. Liu, G. Liu, Smoothed particle hydrodynamics (sph): an overview and recent developments, Archives of computational methods in engineering, Springer ACME 17 (1) (2010) 25–76.

[12] J. P. Morris, Simulating surface tension with smoothed particle hydrodynamics, International journal for numerical methods in fluids, Wiley Online Library IJNMF 33 (3) (2000) 333–353.

[13] J. P. Morris, P. J. Fox, Y. Zhu, Modeling low reynolds number incompressible flows using sph, Journal of computational physics, Elsevier JCP 136 (1) (1997) 214–226.

[14] G. Farnebäck, Two-frame motion estimation based on polynomial expansion, in: Image Analysis, Springer, 2003, pp. 363–370.

[15] J. D. Dunitz, Win some, lose some: enthalpy-entropy compensation in weak intermolecular interactions, Chemistry & biology, Elsevier CB 2 (11) (1995) 709–712.

[16] X. Di, H. Chang, X. Chen, Multi-layer spectral clustering for video segmentation, in: Asian conference on computer vision, Springer ACCV, 2012, pp. 1–12.

[17] H. Ullah, M. Ullah, H. Afridi, N. Conci, F. G. De Natale, Traffic accident detection through a hydrodynamic lens, in: International conference on image processing, IEEE ICIP, 2015.

[18] H. Ullah, M. Ullah, N. Conci, Dominant motion analysis in regular and irregular crowd scenes, in: ECCV workshop on human behavior understanding, Springer ECCVw, 2014, pp. 62–72.

[19] Y. Ma, P. Cisar, A. Kembhavi, Motion segmentation and activity representation in crowds, International journal of imaging systems and technology, Wiley online library IJIST 19 (2) (2009) 80–90.

[20] R. Li, R. Chellappa, Group motion segmentation using a spatiotemporal driving force model, in: International conference on computer vision and pattern recognition, IEEE CVPR, 2010, pp. 2038–2045.

[21] S. Wu, H. San Wong, Crowd motion partitioning in a scattered motion field, Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE SMC 42 (5) (2012) 1443–1454.

[22] B. Zhou, X. Tang, X. Wang, Coherent filtering: detecting coherent motions from crowd clutters, in: European conference on computer vision, Springer ECCV, 2012, pp. 857–871.

[23] Y. Wu, Y. Ye, C. Zhao, Coherent motion detection with collective density clustering, in: International conference on Multimedia, ACM ICM, 2015, pp. 361–370.

[24] D. Zhang, J. Xu, M. Sun, Z. Xiang, High-density crowd behaviors segmentation based on dynamical systems, Journal of multimedia systems, Springer MS (2016) 1–8.

[25] R. Sharma, T. Guha, A trajectory clustering approach to crowd flow segmentation in videos, in: International conference on image processing, IEEE ICIP, 2016, pp. 1200–1204.

[26] B. Poling, G. Lerman, A new approach to two-view motion segmentation using global dimension minimization, International journal of computer vision, Springer IJCV 108 (3) (2014) 165–185.

[27] M. Narayana, A. Hanson, E. Learned-Miller, Coherent motion segmentation in moving camera videos using optical flow orientations, in: International conference on computer vision, IEEE ICCV, 2013, pp. 1577–1584.

[28] F. Shi, Z. Zhou, J. Xiao, W. Wu, Robust trajectory clustering for motion segmentation, in: International conference on computer vision, IEEE ICCV, 2013, pp. 3088–3095.

[29] H. Rahmati, R. Dragon, O. M. Aamo, L. Van Gool, L. Adde, Motion segmentation with weak labeling priors, in: Pattern recognition, Springer PR, 2014, pp. 159–171.

[30] C. Qin, G. Zhang, Y. Zhou, W. Tao, Z. Cao, Integration of the saliency-based seed extraction and random walks for image segmentation, Journal of neurocomputing, Elsevier NC 129 (2014) 378–391.

[31] X. Zhang, H. He, S. Cao, H. Liu, Flow field texture representation-based motion segmentation for crowd counting, Machine Vision and Applications, Springer MVA 26 (7-8) (2015) 871–883.

[32] Q. Wu, Y. Gan, B. Lin, Q. Zhang, H. Chang, An active contour model based on fused texture features for image segmentation, Journal of neurocomputing, Elsevier NC 151 (2015) 1133–1141.

[33] Z. Li, J. Guo, L.-F. Cheong, S. Z. Zhou, Perspective motion segmentation via collaborative clustering, in: International conference on computer vision, IEEE ICCV, 2013, pp. 1369–1376.

[34] I. Huerta, A. Amato, X. Roca, J. Gonzàlez, Exploiting multiple cues in motion segmentation based on background subtraction, Journal of neurocomputing, Elsevier NC 100 (2013) 183–196.

[35] A. Mumtaz, W. Zhang, A. B. Chan, Joint motion segmentation and background estimation in dynamic scenes, in: International conference on computer vision and pattern recognition, IEEE CVPR, 2014, pp. 368–375.

[36] S. Shadden, F. Lekien, J. Marsden, Definition and properties of lagrangian coherent structures from finite-time lyapunov exponents in two-dimensional aperiodic flows, Physica D: Nonlinear Phenomena 212 (3) (2005) 271–304.

[37] V. Kaltsa, A. Briassouli, I. Kompatsiaris, L. J. Hadjileontiadis, M. G. Strintzis, Swarm intelligence for detecting interesting events in crowded environments, Transactions on image processing, IEEE TIP 24 (7) (2015) 2153–2166.

[38] N. Li, X. Wu, D. Xu, H. Guo, W. Feng, Spatio-temporal context analysis within video volumes for anomalous-event detection and localization, Journal of neurocomputing, Elsevier NC 155 (2015) 309–319.

[39] W. Li, V. Mahadevan, N. Vasconcelos, Anomaly detection and localization in crowded scenes, Transactions on pattern analysis and machine intelligence, IEEE PAMI 36 (1) (2014) 18–32.

[40] R. Mehran, A. Oyama, M. Shah, Abnormal crowd behavior detection using social force model, in: IEEE CVPR, 2009, pp. 935–942.

[41] S. Wu, H.-S. Wong, Z. Yu, A bayesian model for crowd escape behavior detection, Transactions on circuits and systems for video technology, IEEE CSVT 24 (1) (2014) 85–98.

[42] D. Xu, R. Song, X. Wu, N. Li, W. Feng, H. Qian, Video anomaly detection based on a hierarchical activity discovery within spatio-temporal contexts, Journal of neurocomputing, Elsevier NC 143 (2014) 144–152.

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

[44] Y. Cong, J. Yuan, J. Liu, Abnormal event detection in crowded scenes using sparse representation, Journal of pattern recognition, Elsevier PR 46 (7) (2013) 1851–1864.

[45] M. Thida, H.-L. Eng, P. Remagnino, Laplacian eigenmap with temporal constraints for local abnormality detection in crowded scenes, Transactions on cybernetics, IEEE C 43 (6) (2013) 2147–2156.

[46] B. Krausz, C. Bauckhage, Loveparade 2010: Automatic video analysis of a crowd disaster, Journal of Computer Vision and Image Understanding, Elsevier CVIU 116 (3) (2012) 307–319.

[47] J. Kim, K. Grauman, Observe locally, infer globally: A space-time mrf for detecting abnormal activities with incremental up-

dates, in: IEEE CVPR, 2009, pp. 2921–2928.

[48] C. Gloor, D. Cavens, E. Lange, K. Nagel, W. Schmid, A pedestrian simulation for very large scale applications, na, 2003.

[49] R. C. Cohen, P. W. Cleary, B. Mason, Simulations of human swimming using smoothed particle hydrodynamics, in: International conference on CFD in the minerals and process industries, 2009.

[50] S. E. Hieber, J. H. Walther, P. Koumoutsakos, Remeshed smoothed particle hydrodynamics simulation of the mechanical behavior of human organs, Technology and Health Care, IOS Press 12 (4) (2004) 305–314.

[51] B. Raney, K. Nagel, Iterative route planning for large-scale modular transportation simulations, Future Generation Computer Systems, Elsevier FGCS 20 (7) (2004) 1101–1118.

[52] R. C. Cohen, P. W. Cleary, B. R. Mason, Simulations of dolphin kick swimming using smoothed particle hydrodynamics, Human Movement Science, Elsevier HMS 31 (3) (2012) 604–619.

[53] R. Temam, Navier-Stokes equations: theory and numerical analysis, Vol. 343, 2001.

[54] Q. Ke, T. Kanade, A subspace approach to layer extraction, in: Computer vision and pattern recognition, IEEE CVPR, Vol. 1, 2001, pp. I–255.

[55] V. Hedau, H. Arora, N. Ahuja, Matching images under unstable segmentations, in: Computer vision and pattern recognition, IEEE CVPR, 2008, pp. 1–8.

[56] G. Hu, Q. Gao, A non-parametric statistics based method for generic curve partition and classification, in: International conference on image processing, IEEE ICIP, 2010, pp. 3041–3044.

[57] PETSdataset, http://www.cvg.rdg.ac.uk/pets2009/a.html, 2009.

[58] H. Ullah, N. Conci, Crowd motion segmentation and anomaly detection via multi-label optimization, in: ICPR workshop on Pattern Recognition and Crowd Analysis, 2012.

[59] D. Iakovidis, T. Goudas, C. Smailis, I. Maglogiannis, Ratsnake: A versatile image annotation tool with application to computer-aided diagnosis, The Scientific World Journal 2014.