Measuring Crowd Collectiveness via Global Motion Correlation

Ling Mei[1] , Jian-Huang Lai[2,3,4*] , Ze-Yu Chen[2] , and Xiao-Hua Xie[2,3,4]
[1]School of Electronics and Information Technology, Sun Yat-sen University, China
[2]School of Data and Computer Science, Sun Yat-sen University, China
[3]Guangdong Key Laboratory of Information Security Technology, China
[4]Key Laboratory of Machine Intelligence and Advanced Computing, Ministry of Education, China
{meil3, chenzy5}@mail2.sysu.edu.cn, {stsljh, xiexiaoh6}@mail.sysu.edu.cn

Abstract

Crowd collectiveness refers to the behavior consistency of crowd scenes, which reflects the degree of collective movements among massive individuals in crowd systems. The existing methods focus on measuring the discrepancy of motion direction among the individuals. However, few studies consider the magnitude discrepancy of velocity in a crowd and the collectiveness among different crowds, which can also affect the overall crowd collectiveness. In this paper, we propose a novel descriptor which combines intra-crowd collectiveness with inter-crowd collectiveness to solve the problem. For intra-crowd collectiveness, we introduce the energy spread process to identify the impacting factors of collectiveness, then measure the collectiveness of individuals within a crowd cluster by computing their similarities of magnitude and direction from the optical flow. For inter-crowd collectiveness, we assess the motion consistency among various crowd clusters generated from collective merging. Experimental results demonstrate that how the new collectiveness descriptor improves performance on three different crowd datasets, thus validating the superiority of the proposed descriptor.

1. Introduction

The motion collectiveness of crowds is one of the most common phenomena in nature, which broadly appears in different crowd systems such as human crowd, bacterial colony, and traffic vehicles. Motion collectiveness represents the collective movements of massive individual particles of crowds in both nature and social scenarios. Motion collectiveness is closely related to many low-level and high-level computer vision problems, such as solving problems of depicting collective human behavior [45], motion pattern understanding [1, 48], motion segmentation [16, 30], object tracking [2], and microorganism motion analysis [40]. Meanwhile, it is also applied to the high-level semantic analysis in crowded scenes such as modeling scene structures [11, 34, 47] and recognizing crowd behaviors [48].

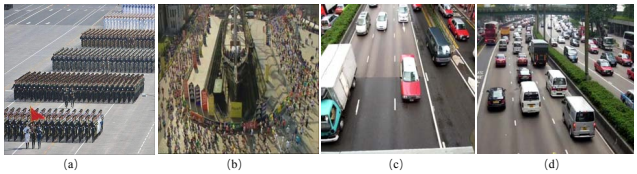

Figure 1: Examples depict the different collectiveness with consistent motion direction. (a) Military formations. (b) Marathon parades. (c) One-way traffic flow. (d) Two-way traffic flow.

Measuring crowd collectiveness indicates the degree of individuals acting as a union in collective motion [45], it is difficult to measure the complex crowd behavior because the integrated individuals in crowds can generate diverse motion patterns, so the problem is essentially whether there is an objective way to identify the key factors for explicitly measuring the underlying principles of the phenomenon? In general, the behavior of crowd primarily affects the motion consistency of crowd. Especially, the direction and velocity of the crowd motion should be related to the crowd collectiveness according to human's visual intuition. For example, as shown in Fig. 1(a) and (b), military formations in a parade and marathon parades have the same consistency in terms of motion direction, but the military formations have a more uniform velocity than the marathon parades, our intuition is that the former should be more consistent. Thus, we need to find a rational way to reveal the relation between intuition and motion collectiveness.

Motivated by the energy diffusion proposed in [19], it transforms input motion field into a coherent crowd motion field named as thermal energy field. Energy diffusion treats the crowd individuals as particles in sets and uses the optical flow based energy mechanics to capture individual movements and their interactions. However, [19] ignores the initial optical flow of particles, which can't get a comprehensive energy field to find all impacting factors of crowd col-

*Corresponding author

lectiveness. Therefore, we propose an energy spread process to solve the problem and try to capture all useful impacting factors of crowd collectiveness.

In addition, traditional approaches typically focus on the collectiveness of intra-crowd in terms of crowd density estimation [14, 18, 43]. However, a crowd may include some clusters that involve different motion patterns. As shown in situations of Fig. 1 (c) and (d), although the motion directions of the vehicle crowd in every road lane is consistent, the collectiveness of Fig. 1 (c) is more consistent in terms of intuition, and the intra-crowd collectiveness can't measure the situation well. Hence, we propose the inter-crowd collectiveness to measure the motion collectiveness among these clusters.

In this study, we propose a fusional descriptor to estimate crowd collectiveness. The novelty of our approach lies in three aspects. First of all, we propose a novel strategy to establish the intra-crowd collectiveness by measuring the discrepancies of both the velocity magnitude and motion direction. Second, an energy spread process is introduced to validate the factors that impact intra-crowd collectiveness. The energy spread process utilizes the theory of zero-input response and zero-state response in terms of optical flow to deduce the diffusion process. Third, we introduce the cluster merging module to compute the inter-crowd collectiveness, which indicates the collectiveness among different crowd clusters. The overall crowd collectiveness is thus composed of the intra-crowd and inter-crowd collectiveness.

The remainder of this paper is organized as follows. Section 2 reviews the related work. Section 3 presents the theoretical framework of the proposed crowd collectiveness descriptor and the energy spread process. The experimental results and corresponding conclusions are given in Sections 4 and 5, respectively.

2. Related Works

In this section, we review relevant studies of crowd motion consistency in the four aspects of the energy spread process, optical flow, merging cluster motion, and the collective motion property.

2.1. Energy Spread Process

The energy spread process based methods [6, 19, 32, 35, 37] are previously used for finding coherent motions. However, our method differs from them. At first it uses the energy spread process to find the key factors that influence motion consistency, while the other methods use it for image segmentation or understanding crowd scenes. Second, our work not only considers the additional effects on the particle from external force but also the initial energy state, while other works only consider the former. It is natural for us to regard the energy change of a particle as a system,

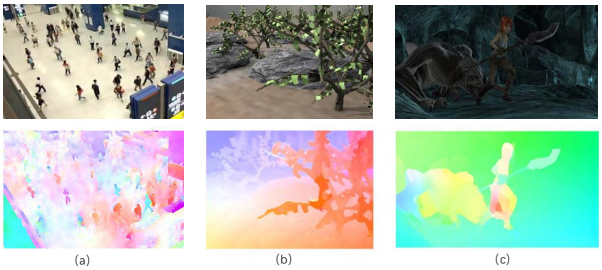

Figure 2: Visualization of optical flow in different scenes. (a) Crowd in market. (b) Grove scene on the Middlebury dataset [4]. (c) Moving scene in animation on the MPI Sintel dataset [9]. The first row is the original image, the second row is the visual results of optical flow.

and the effects of an external force as a signal. If there is no external signal source in the system and the initial conditions have non-zero value, then the solution with these initial conditions is known as the zero-input response [3], which is similar to the initial energy state. If all of the initial conditions have zero values, then the system is said to be in the zero-state, and the solution to non-zero inputs for the system is known as the zero-state response, which is like the effects of external force. The complete response is the sum of the zero-input response and the zero-state response, which can reflect the real motion state of a particle. Thus we can use the theory of zero-input response and zero-state response from the field of signals and systems [26].

Compared with the complete response, the zero-state response is incomplete and limited in its ability to reflect the energy spread process. To tackle this issue, we introduce the complete response to explain the energy spread process and take the zero-input response into consideration.

2.2. Motion Field of Optical Flow

Obtaining an accurate motion field is essential for measuring crowd collectiveness. The motion field can be built by using the optical flow estimation, which is typically solved by an energy minimization framework of the data term and the smoothness term [12,23]. The data term mainly reduces the motion blur and preserves the image plots such as motion boundaries, while the smoothness term can reduce outliers by denoising [24]. As the visualization results shown in Fig. 2, optical flow can precisely extract the motion boundaries both in the cases of small-displacement (Middlebury Dataset [4], see Fig. 2(b)) and large-displacement (MPI Sintel Dataset [9], see Fig. 2(c)). The energy optimization of the optical flow preserves the motion boundary and segments the regions of coherent motion clusters. Energy diffusion by optical flow can be regarded as an external force added onto the individual particle in a cluster to affect its diffusion behavior [19], and the

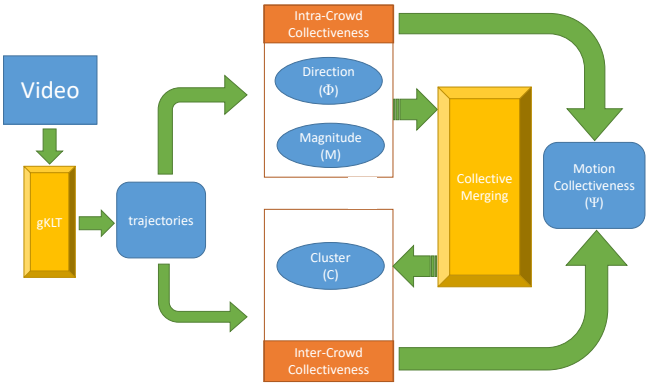

Figure 3: Framework of the proposed method.

information of the optical flow preserves the original motion patterns.

2.3. Merging Cluster Motion

Merging cluster motion is a classical topic with wide applications including crowd motion analysis, and some approaches [11, 25, 33, 47, 48] learn motion patterns by clustering trajectories in crowd scenes. Sharma *et al*. [27] proposed a clustering-based trajectory approach for segmenting flow patterns in high-density crowd videos, and Wu *et al*. [36] proposed the Curl and Divergence of motion Trajectories (CDT) descriptors to describe collective motion patterns. Detecting whether the behavior in crowds is coherent is of great interest in crowd surveillance management. Zhou *et al*. [44] proposed a graph-based method to detect coherent motion from tracklets. Lin *et al*. [19] discovered recurrent activities in crowd scenes by clustering and merging the extracted coherent motion data. Wang *et al*. [31] proposed multi-view clustering method by incorporating their motion and context similarities to detect coherent groups in crowd scenes.

Another class of methods is based on the different similarity measurement of clusters. Recently, Keuper *et al*. [16] proposed correlation co-clustering method by normalizing the motion difference between different trajectories based on defining graph model, which can be thought of measuring the discrepancy of velocity of clusters. On the other hand, Wu *et al*. [38] introduced Collective Density Clustering (CDC) method which using the cosine similarity to measure the motion discrepancy of clusters.

2.4. Collective Motion Property

Scientists in various research fields have long been interested in the collective motion of organisms. Individuals in crowds tend to lose their personalities and follow the behavior of others. In biology, the principles of collective behavior have been revealed through empirical or theoretical methods. In empirical methods, crowd behavior such as bacterial colonies [40] have been extensively collected and analyzed. It is found that factors like crowd density [28] and crowd counting [39] make difference to the crowd collectiveness. Zhang *et al*. [40] used bacteria densities to study collective motion, and Makris *et al*. [22] studied the temporal and spatial collectiveness of vast oceanic shoals. As for theoretical methods, various models and theorems about the properties of collectiveness have been proposed, such as self-driven particle (SDP) models [10] and collective merging [7, 21, 38, 45].The SDP models can establish collectiveness as they are very similar to complex crowd systems in reality. Collective merging detects collective motion from randomly moving particles, and can remove outlier particles with low collectiveness then obtain the connected clusters by a threshold [46]. However, the collective merging is only used in the postprocess or for visualization, and its potential effects to construct the descriptor have not been explored.

Algorithm 1: Measuring Crowd Collectiveness via Global Motion Correlation

Input: Crowd video frame I.
Output: Crowd collectiveness degree Ψ.

Compute the optical flow field of I to get the magnitude and direction of the particles' velocity;
Filter out the outlier particles by the threshold h;
Thresholding matrix $\mathbf{Z}$ to get M clusters$\{c_m\}_{m=1}^{M}$;
for *each cluster* c_m **do**
  for *each particle* $\mathbf{P}$ *in a cluster* c_m **do**
    Compute the similarity $s_t(\mathbf{P}, \mathbf{Q})$ between $\mathbf{P}$ and the neighboring particle $\mathbf{Q}$ using Eq. 5;
    Construct the path-based graph [5] by weighted adjacency matrix $\mathbf{S}$;
    Compute the $l-$path similarity $s_l(\mathbf{P}, \mathbf{Q})$ using Eq. 6 and 7;
    Using Eq. 8, compute the sum $\Psi_l(\mathbf{P})$ of all $l-$path similarity;
    Using Eq. 9, compute the collectiveness of $\mathbf{P}$ on all path similarities;
  Using Eq. 10 to compute the intra-crowd collectiveness Ψ_{intra};
Measure the average normalized velocity by Eq. 11;
Compute the inter-crowd collectiveness Ψ_{inter} using Eq. 12;
Return the overall crowd collectiveness Ψ.

3. Framework of the Proposed Method

In this section, we introduce the proposed descriptor of crowd collectiveness, which merges global motion correlation via intra-crowd collectiveness and inter-crowd collectiveness. As shown in Fig. 3, our framework general-

ly involves three steps: First, find the trajectories of individuals using tracking algorithms such as the generalized KLT (gKLT) tracker derived from [13]. Second, measure the intra-crowd collectiveness which consists of direction module (Φ) and magnitude module ($\mathbf{M}$) according to the corresponding trajectories. Third, use collective merging to measure the inter-crowd collectiveness of cluster module ($\mathbf{C}$), then merge the intra-crowd and inter-crowd collectiveness to generate the global crowd motion collectiveness. As gKLT is the preliminary step which can be found in [45], we start to describe the proposed framework from the second step.

3.1. Proposed Energy Spread Process

To find the useful factor of intra-crowd collectiveness rationally, it is important to construct a collective motion field of correlated particles which can describe the coherent crowd motions accurately. In the micro crowd motion field, correlated motion patterns of these particles usually collide and change trajectory. They diffuse thermal energy to other neighboring particles by colliding, which can change the motion status of the other particles. Therefore, the motion energy of the particles is updated. As [19] mentions, after T seconds, the diffused energy from $\mathbf{Q}$ to $\mathbf{P}$ is

$$e^{\gamma}_{\mathbf{P},T} = u^{\gamma}_{\mathbf{Q}} \times e^{-k_p\|\mathbf{P}-\mathbf{Q}\|^2} \times e^{k_f|\mathbf{F_Q}\cdot(\mathbf{P}-\mathbf{Q})|} \tag{1}$$

where $\gamma \in (x, y)$ denotes the axis, k_p and k_f are the propagation coefficients, $\mathbf{F_Q}$ is the optical flow at location $\mathbf{Q}$, $\mathbf{U_Q} = (u^x_{\mathbf{Q}}, u^y_{\mathbf{Q}})$ is the current motion pattern of $\mathbf{Q}$, which is initialized by $\mathbf{U_Q} = \mathbf{F_Q}$, and the corresponding thermal energy is $|\mathbf{U_Q}|$. However, [19] ignores the initial input motion $\mathbf{F_P}$ in energy spread process, the energy diffusion can't change the motion state of the particle since the initial motion always assumes to zero, so it can't deduce the underlying principle of intra-crowd collectiveness. Hence, our motivation is finding a global energy diffusion way to reveal the principle, which considers the initial input motion.

Motivated by the complete response which considers both the initial and current energy state, the beginning of the energy spread process is supposed to be zero-moment. Let $E_{\mathbf{P}}$ be the total motion energy at location $\mathbf{P}$, which can be split into two parts: the diffused energy from the heat source that not concerned with the initial motion is the zero-state response $E_{\mathbf{P},zs}$; and the initial motion energy before the diffusion process is the zero-input response $E_{\mathbf{P},zi}$. Based on Eq. 1 [19], the final individual thermal energy diffused from the heat source $\mathbf{Q}$ can be expressed as the zero-state response $E_{\mathbf{P},zs}$:

$$\begin{aligned} E_{\mathbf{P},zs} &= \sqrt{(e^x_{\mathbf{P},T})^2 + (e^y_{\mathbf{P},T})^2} \\ &= |\mathbf{U_Q}| \times e^{-k_p\|\mathbf{P}-\mathbf{Q}\|^2} \times e^{k_f|\mathbf{F_Q}\cdot(\mathbf{P}-\mathbf{Q})|} \\ &= |\mathbf{U_Q}| \times e^{-k_p} \times e^{k_f'|\mathbf{F_Q}|} \end{aligned} \tag{2}$$

Here we assume the length of $|\mathbf{P}-\mathbf{Q}|$ is 1 for computational convenience as they are neighboring, k_f' is a coefficient. In this way, optical flow is the current motion pattern of a particle, the energy of a particle's optical flow updates the correlation with neighboring particles, and the new correlation updates its optical flow in turn. Therefore the energy of particle's original optical flow is also diffused by its neighboring particles before the zero-moment according to the energy spread process. Eq. 1 shows that particle $\mathbf{P}$ with an original motion state like the zero-input response, which considers the initial optical flow without energy diffusion from other particles after the zero-moment, but it can be diffused by the neighboring particle's motion before the zero-moment. Combined with Eq. 1, optical flow $\mathbf{F_Q}$ should be projected to the direction of optical flow $\mathbf{F_P}$ to diffuse the energy, and $\mathbf{F_Q}$ is invariant in the zero moment. As Eq. 2 indicates $E_{\mathbf{P},zs} \propto |\mathbf{U_Q}| \times e^{k_f'|\mathbf{F_Q}|}$, the energy of the zero-input response can be denoted as

$$E_{\mathbf{P},zi} = |\mathbf{U_Q}| \times e^{k_i|\mathbf{F_P}|} \times \cos(\mathbf{F_P}, \mathbf{F_Q}) \tag{3}$$

where k_i is a propagation factor in the exponential term to approximate the impact of $\mathbf{F_P}$ in the zero-input response. Thus the total motion energy of $\mathbf{P}$ is the complete response, which combines the zero-state and zero-input responses. The energy of a particle reflects the real motion state in a crowd, and the crowd motion performs more consistent if the energy distribution of particles is uniform, so it is important to compare the energy of different particles. Hereby, the energy ratio r of particle $\mathbf{P}$ and its neighboring $\mathbf{Q}$ is computed as

$$\begin{aligned} r = \frac{E_{\mathbf{P}}}{E_{\mathbf{Q}}} &= \frac{|\mathbf{U_Q}|\,(e^{-k_p+k_f'|\mathbf{F_Q}|} + e^{k_i|\mathbf{F_P}|} \times \cos(\mathbf{F_P}, \mathbf{F_Q}))}{|\mathbf{U_Q}|} \\ &= e^{-k_p+k_f'\lambda|\mathbf{F_P}|} + e^{k_i|\mathbf{F_P}|} \times \cos(\mathbf{F_P}, \mathbf{F_Q}) \\ &= e^{k_i|\mathbf{F_P}|}(e^{-k_p+(k_f'\lambda-k_i)|\mathbf{F_P}|} + \cos(\mathbf{F_P}, \mathbf{F_Q})) \end{aligned} \tag{4}$$

where $\lambda = \frac{|\mathbf{F_Q}|}{|\mathbf{F_P}|}$ is the magnitude ratio of the velocity, the direction similarity can be denoted by the cosine similarity. As k_p, k_f', k_i are constant and $|\mathbf{F_P}|$ in Eq. 4 has been given, the potential discriminative factors remain the velocity magnitude ratio and the motion direction similarity of neighboring particles. Thus, we denote $r_t(\mathbf{P}, \mathbf{Q})$ as the motion correlation at time t between $\mathbf{P}$ and $\mathbf{Q}$.

3.2. Intra-Crowd Collectiveness

The intra-crowd collectiveness can be determined by the collectiveness of its constituent individuals, and can be computed in three steps. First, we compute the motion consistency in neighborhoods of individuals; then, we compute the pairwise collectiveness of the individuals within particular paths; finally the crowd collectiveness is obtained by the mean of all individual collectiveness (see Algorithm 1).

We propose a comprehensive similarity metric to measure the intra-crowd motion consistency between each particle and its neighborhood. The comprehensive motion similarity of intra-crowd is defined as

$$s_t(\mathbf{P}, \mathbf{Q}) = \max(r_t(\mathbf{P}, \mathbf{Q}), 0) \tag{5}$$

where individual $\mathbf{Q}$ is in the neighborhood of $\mathbf{P}$ at time t. r_t consists of the magnitude similarity λ and direction similarity $\cos(\mathbf{F_P}, \mathbf{F_Q})$ of the particle's velocity. Then we normalize $s_t(\mathbf{P}, \mathbf{Q}) \in [0, 1]$ to measure the motion consistency between the adjacent individuals.

After that, we need to estimate the motion similarity between non-adjacent individuals. The path-based graph proposed in [5] can solve this problem. Let $\mathbf{S}$ be the weighted adjacency matrix in the graph that represents the individual distribution of crowd set D, and $s_t(\mathbf{P}, \mathbf{Q})$ be the edge in the graph. Considering all paths $\mathcal{P}_l$ between individuals $\mathbf{P}$ and $\mathbf{Q}$ whose lengths are l , similarity on these paths can be collected as the l-path similarity

$$s_l(\mathbf{P}, \mathbf{Q}) = \sum_{\{\mathcal{P}_l\}} s_{\mathcal{P}_l}(\mathbf{P}, \mathbf{Q}) \tag{6}$$

$$s_{\mathcal{P}_l} = \prod_{i=0}^{l} s_t(q_i, q_{i+1}) \tag{7}$$

where $\{q_i\}$ are the nodes in path $\mathcal{P}_l$. In addition, as there can be more paths $\mathcal{P}_l$ with length l between $\mathbf{P}$ and $\mathbf{Q}$, Eq. 7 denotes each path similarity. As the l-path similarity gives the motion consistency between $\mathbf{P}$ and $\mathbf{Q}$ at l-path scale, we can define the collectiveness of all individuals with l-path to individual $\mathbf{P}$ as

$$\Psi_l(\mathbf{P}) = \sum_{\mathbf{Q} \in D} s_l(\mathbf{P}, \mathbf{Q}) \tag{8}$$

Then we define the individual collectiveness on all path similarities in crowd D as

$$\Psi_D(\mathbf{P}) = \sum\nolimits_{l=1}^{+\infty} z^l \Psi_l(\mathbf{P}) = [\mathbf{Zi}]_k \tag{9}$$

where $[\cdot]_k$ is the k-th element of a vector. Thus the intra-crowd collectiveness of D is defined as the mean of all individual motion collectiveness, which can be written as

$$\Psi_{intra} = \frac{1}{|D|} \sum_{\mathbf{P} \in D} \Psi_D(\mathbf{P}) = \frac{1}{|D|} \mathbf{i}^{\mathbf{T}} \mathbf{Z} \mathbf{i} \tag{10}$$

where $|D|$ is the cardinality of set D, $\mathbf{i}$ and $\mathbf{I}$ are the unit vector and unit matrix, respectively, and the matrix $\mathbf{Z} = (\mathbf{I} - z\mathbf{S})^{-1} - \mathbf{I}$ is a mutual similarity matrix of all particles which converges when $z < \frac{1}{K}$. Here, z is a real-valued regularization factor of the weighted adjacency matrix $\mathbf{S}$, and K is the number of adjacent particles. Thus, the intra-crowd collectiveness has upper and lower bounds which denoted as $0 \leq \Psi_{intra} \leq \frac{zK}{1-zK}$ [41, 45]. The new intra-crowd crowd collectiveness descriptor measures both magnitude and direction discrepancy of the optical flow, which can be used to measure the intra-crowd motion consistency as a whole.

Figure 4: Examples depict the visualization of the merging clusters. Scatters with different colors indicate different detected merging clusters, arrows represent motion orientations, cross points indicate the outliers. (a) Human crowd. (b) Traffic crowd. (c) Bacterial colony.

3.3. Measuring Inter-Crowd Collectiveness

Since the crowd may be constituted of more than one clusters of individuals gathering together. The intra-crowd collectiveness can't measure the motion consistency among different clusters, while the inter-crowd collectiveness can make up the problem. We propose an cluster merging module to detect the inter-crowd collectiveness using a collective merging process, which can measure the collectiveness among different clusters if more than one crowd cluster appears simultaneously in the video. First, we need to remove the outlier particles with low intra-crowd collectiveness and derive the clusters of crowds from the connected components via thresholding matrix $\mathbf{Z}$. According to Property 3 in [45], we let $h = \frac{\alpha K}{1-zK}$ be the threshold to remove outlier particles where $\alpha \in (0.4, 0.8)$. The connected components in the image are then integrated into piles of particle clusters. Thus, the outliers with low collectiveness can be removed, while groups of neighboring particles with high motion consistency can be extracted and separated into M different clusters $\{c_m\}_{m=1}^{M}$. Fig. 4 shows the visualization of the cluster merging.

Second, for a cluster c_m with N particles, we measure the motion similarity between it and other cluster by two strategies. One is the average normalized velocity which denoted as

$$\mathbf{v}_m = \frac{1}{N} \sum\nolimits_{n=1}^{N} \frac{\mathbf{v}_{mn}}{\|\mathbf{v}_{mn}\|} \tag{11}$$

where $\mathbf{v}_{mn}$ denotes the velocity of the nth particle in cluster c_m. Eq. 11 suppresses the impact of the velocity magnitude of particles by normalization, and focuses on the overall motion direction of a cluster. Then the collectiveness Ψ_{inter} of all clusters can be merged as

$$\Psi_{inter} = \left\| \frac{\sum_{m=1}^{M} a_m \mathbf{v}_m}{\sum_{m=1}^{M} a_m \|\mathbf{v}_m\|} \right\| \tag{12}$$

Method	MAE ↓	Correlation (%) ↑
Φ	21.67	78.34
M	62.18	-27.26
C(Velocity)	30.87	50.85
C(Cosine)	42.48	57.74
Φ+M	17.55	79.04
Φ+C(Velocity)	16.28	81.68
Φ+C(Cosine)	16.65	81.31
M+C(Velocity)	26.71	54.39
M+C(Cosine)	25.18	59.63
Φ+M+C(Velocity,Ours)	**15.64**	**82.85**
Φ+M+C(Cosine,Ours)	**15.92**	**82.51**

Table 1: Results of different modules and combined modules on the Collective Motion Dataset. The proposed collectiveness "Ψ" is the same as the combined modules of "Φ+M+C", "Velocity" and "Cosine" denote using average normalized velocity and cosine similarity as the cluster merging measurement, respectively.

(a) GT = 0.9, Ψ = 0.97 Φ = 0.7 (b) GT = 0.8, Ψ = 0.88 Φ = 0.

(c) GT = 0.9, Ψ = 0.71 Φ = 0. (d) GT = 0.9, Ψ = 0.88 Φ = 0.49

Figure 5: Some comparative collectiveness results on Collective Motion Dataset. GT, Ψ, Φ are the ground truth, our collectiveness descriptor and method in [45], respectively. The four scenes are: (a) One-way traffic flow. (b) Marathon parades. (c) Same motion direction in two-lane highway. (d) One pedestrian walks against the crowd.

where a_m is the number of particles in c_m. In Eq. 12, the normalized velocity of all clusters are weighted according to the corresponding size of each cluster, which emphasizes the clusters with larger amount of particles. Eq. 12 finally normalizes the weighting result to obtain the velocity magnitude Ψ_{inter} of merging clusters, which is the collectiveness of the overall clusters.

The other strategy is the cosine similarity, we denote the average cosine similarity between each $\mathbf{v_m}$ and the average velocity $\mathbf{v}$ of all particles as

$$\Psi_{inter} = \frac{\sum_{m=1}^{M} a_m \cos(\mathbf{v}_m, \mathbf{v})}{\sum_{m=1}^{M} a_m} \tag{13}$$

In this way, we can measure the motion direction consistency among multiple clusters.

The value of Ψ_{inter} represents the inter-crowd motion consistency. The overall crowd collectiveness Ψ consists of Ψ_{intra} and Ψ_{inter}.

3.4. Difference from Other Research

Among prior research, [45] is the most relevant work to ours, in which collectiveness is utilized to measure crowd motion consistency. However, the core ideas in the two works are different. In [45], only the discrepancy of motion direction is considered when assessing the similarity of individuals, and not consider whether the velocity of the two crowds is consistent or not. In addition, we validate the necessity of velocity magnitude with the energy spread process, which [45] does not. Our work also focuses on using optical flow to explain the factors affecting motion consistency. Another study [19] also uses the energy spread process, but it focuses on discovering recurrent activities in crowd scenes, while our work use the process to find the factors that affect crowd motion collectiveness. In addition, [19] only considers the zero-state response in the energy spread process without the initial optical flow term, while our approach jointly leverages the initial optical flow term as the zero-input response of the process.

4. Experiments

4.1. Datasets and Settings

Datasets. We perform experiments on three crowd datasets, *i.e.*, the Collective Motion Dataset [45], the Self-Driven Particle (SDP) Dataset [29], and the Bacterial Colony Dataset [40]. The Collective Motion Dataset consists of 413 video clips from 62 different crowds scenes, which includes parades, pedestrian crowds, and vehicle crowds. The SDP Dataset introduces a novel type of dynamics model to investigate the emergence of self-ordered motion in crowds of particles with biologically motivated interaction. The Bacterial Colony Dataset comes from the wild-type *Bacillus subtilis* [15] bacteria colony which exhibits the motion patterns of microorganism crowd. There are two videos in the Bacterial Colony Dataset, one video records the motion of bacteria colony with sparse density named "SM01" and the other records dense bacteria colony named "SM02", their average numbers of bacteria in each frame are 343 and 718 [40].

Settings. To make fair comparisons, we evaluate using the Mean Absolute Error (MAE) and the correlation

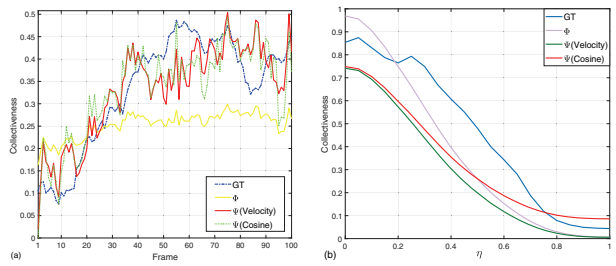

Figure 6: Evolution of SDP collectiveness with time and different noise level. (a) Collectiveness evolution under different frames and $\eta = 0.3$. (b) Average collectiveness within 100 frames under increasing η.

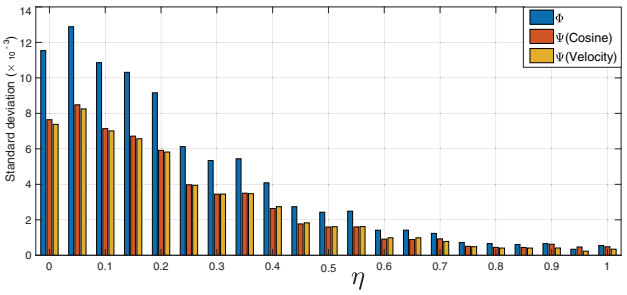

Figure 7: The stability comparison of our methods and Φ with increasing η. The smaller deviation means the method is stable and less sensitive to the initialization of SDP. We repeats the simulation at each η for 20 times as [45].

degree between the ground truth and different collectiveness descriptors. Crowd collectiveness may be related to the crowd counting as mentioned in Section 2.4, so we follow the crowd counting [17, 20, 42] to use MAE as the metric, which complements the evaluation of correlation metric [45]. MAE is computed as

$$MAE = 100 \times \frac{1}{L} \sum_{i=1}^{T} (X_i - Y_i) \tag{14}$$

where X_i and Y_i are the estimated collectiveness and ground truth of the ith frame in a video, respectively. L is the number of frames in the video. MAE reflects the degree of absolute error. The indicator of the correlation metric is a quantitative metric w.r.t. the degree of linear correlation between variables X and Y:

$$Correlation = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} \times 100\% \tag{15}$$

where $Cov(X,Y)$ and $Var(X,Y)$ are the covariance and variance between variables X and Y, respectively. According to our metrics, lower MAE and higher correlation indicates a better collectiveness descriptor.

The ground truth of the Collective Motion Dataset is independently graded by 10 volunteers according to [45]. The collective motion degree of these videos are rated as either low, medium or high, and are scored as 0, 1, and 2, respectively. The range of the total scores is [0, 20], and for the sake of convenience we normalize the scores to [0, 1] as the ground truth (GT). The test conditions of the parameters are set to $(K, z, \alpha, h) = (20, 0.025, 0.6, 0.03)$ in the proposed framework.

4.2. Evaluations of the Collective Motion Dataset

The proposed collective framework involves three modules: direction collectiveness ($\mathbf{\Phi}$), velocity magnitude collectiveness ($\mathbf{M}$), and cluster merging collectiveness ($\mathbf{C}$). As mentioned above, the evaluation can be performed in each module independently or in combined patterns, we perform ablation experiments to evaluate the effect of each module.

The results of MAE and the correlation are shown in Table 1. The performance of our descriptor is found to be better than other modules, and outperforms the descriptor of the direction collectiveness, which improves MAE by 6.03 and raises the correlation by 4.51%. The results show that the correlation of the single velocity magnitude module is negative, which indicates it is not sufficient to only consider the discrepancy of velocity magnitude and it must be considered together with the direction module. For the cluster module, the Velocity strategy in Eq. 12 is better than the Cosine strategy in Eq. 13 since the former handles velocity as vector, which considers the motion consistency of both magnitude and direction among clusters. However, the latter only considers the direction. According to our deduction in energy spread process, the magnitude is also helpful to the collectiveness. As a result, the direction module combined with either the velocity magnitude module or the cluster merging module may also lead to improvements, which indicates the effectiveness of the velocity magnitude collectiveness and the cluster merging collectiveness.

In addition, we provide examples in Fig. 5 to illustrate the performance in difficult situation, particularly for the case in Fig. 5(c), which appears to be a two-way road but is in fact a same direction of traffic moving in both lanes. Intra-crowd descriptor Φ handles the situation as a whole and ignores the inter-crowd collectiveness, while the proposed Ψ overcomes the defect and achieves much better results than Φ in kinds of complex situation in Fig. 5. On the other hand, since our method is based on trajectory information, we compare with recent state-of-the-art methods which based on tracking, we evaluate the collectiveness results by achieving their algorithms [16, 38] or released code from GitHub [45], and our method outperforms them as shown in Table 2. All above experiments validate the superiority of our method on Collective Motion Dataset.

Method	MAE ↓	Correlation (%) ↑
Keuper et al. [16]	30.87	50.85
Wu et al. [38]	42.48	57.74
Zhou et al. [45]	21.67	78.34
Ours	**15.64**	**82.85**

Table 2: Comparison with state-of-the-art methods on the Collective Motion Dataset.

4.3. Evaluations of the Self-Driven Particle Dataset

To demonstrate the effectiveness on the SDP Dataset, we compare our approach with other state-to-the-art descriptors by different sub-modules. SDP has been applied widely in the study of collective motion and resulted in high level of similarity with different real-world crowd systems [8]. The ground truth of the SDP Dataset is the instantaneous average normalized velocity of all N particles

$$v = \left\| \frac{1}{N} \sum\nolimits_{n=1}^{N} \frac{\mathbf{v}_n}{\|\mathbf{v}_n\|} \right\| \tag{16}$$

The particles of SDP model are driven with a constant speed, and their motion directions are updated according to the average directions of their neighboring particles. The velocity magnitude in the SDP model is invariant, which leads to no discrepancy about the velocity magnitude in this dataset. Thus, we test our module without the magnitude module.

We conduct experiments with complex SDP models. In this case, the SDP model is added different levels of random perturbed noise η on the aligned orientation in neighborhood, which would cause the phase transition of crowd system to change from disordered movements into consistent motion. At initialization, the SDP particles are randomly assigned with spatial locations and velocity directions. The moving direction of particle $\mathbf{P}$ is updated as

$$\theta_{\mathbf{P}}(t) = \langle \theta_{\mathbf{Q}}(t-1) \rangle_{\mathbf{Q} \in \mathcal{N}(\mathbf{P})} + \Delta\theta \tag{17}$$

where $\theta_{\mathbf{P}}(t)$ denotes the moving direction of $\mathbf{P}$ at time t, $\theta_{\mathbf{Q}}(t-1)$ is the average direction of velocities within the neighboring particles of $\mathbf{P}$. $\Delta\theta$ is a random angle chosen within the interval $[-\pi, \pi]$. GT is the average normalized velocity of all particles. As Fig. 6(a) shows, the collectiveness is low since the particles are randomly assigned at the initialization, the behaviors of the SDP system turns into collective motion gradually, the two kinds of our methods are more accurate to the GT than Φ. In addition, as shown in Fig. 6(b), the collectiveness of SDP becomes less as the perturbation increases, all curves reflect the trend, so we need to make further comparison to their stability under different η. For further comparison, we plot the histograms of their standard deviation in Fig. 7. From the large standard deviation of Φ under the same η compared with both of our methods, it is shown that our methods are more stable and less sensitive to the initial conditions of SDP model, which validates the robustness of our methods in terms of handling the perturbation.

Method / Video	Φ	Φ+M	Φ+C	Ψ**(Ours)**
SM01	12.44	15.31	16.62	**18.44**
SM02	35.38	35.42	55.32	**55.64**

Table 3: The numerical results of the correlation (%) between the measuring collectiveness and the crowd density.

4.4. Evaluations on Bacterial Colony Dataset

In this experiment, we use the bacterial motion videos to evaluate our method. The crowd density has been proved to be one key factor for the collective bacteria motion [29, 40]. Therefore we conduct our experiment to analyze the correlation between the collectiveness descriptor and the number of the bacteria crowd.

Table 3 gives the numerical results of the correlation, we use the similarity of average normalized velocity as the cluster merging measurement in this experiment. The results in "SM02" is better than "SM01" because its density is higher, which leads it to be more probably measured the trend of bacteria motion. Moreover, Ψ is better than other methods, and adding the magnitude module (**M**) or cluster merging module (**C**) can improve the direction module (Φ). The results indicate the measurement of our crowd collectiveness descriptor has outstanding potentials to apply in the scientific research.

5. Conclusion

In this study, we propose a comprehensive collectiveness descriptor to measure the crowd collectiveness by exploiting the energy spread process. It considers both the discrepancy of the magnitude and the direction in crowd motion to get the intra-crowd collectiveness. Moreover, we measure the motion similarity among different crowd clusters to construct inter-crowd collectiveness. We have validated and demonstrated its robust performance in terms of depicting crowd collective behavior.

Acknowledgements

This work was supported by National Key Research and Development Program of China (2016YFB1001003), the NSFC (61573387, 61876104).

References

[1] S. Ali. Measuring flow complexity in videos. In *ICCV*, pages 1097–1104, 2013.

[2] S. Ali and M. Shah. Floor fields for tracking in high density crowd scenes. In *ECCV*, pages 1–14. Springer, 2008.

[3] T. Bäckström. Windowing and the zero input response. In *Speech Coding*, pages 77–90. Springer, 2017.

[4] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. J. Black, and R. Szeliski. A database and evaluation methodology for optical flow. *IJCV*, 92(1):1–31, 2011.

[5] N. Biggs, N. L. Biggs, and B. Norman. *Algebraic graph theory*, volume 67. Cambridge university press, 1993.

[6] M. J. Black, G. Sapiro, D. H. Marimont, and D. Heeger. Robust anisotropic diffusion. *TIP*, 7(3):421–432, 1998.

[7] T. Brox and J. Malik. Object segmentation by long term analysis of point trajectories. In *ECCV*, pages 282–295. Springer, 2010.

[8] J. Buhl, D. J. Sumpter, I. D. Couzin, J. J. Hale, E. Despland, E. R. Miller, and S. J. Simpson. From disorder to order in marching locusts. *Science*, 312(5778):1402–1406, 2006.

[9] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black. A naturalistic open source movie for optical flow evaluation. In *ECCV*, pages 611–625. Springer, 2012.

[10] H. Chaté, F. Ginelli, G. Grégoire, and F. Raynaud. Collective motion of self-propelled particles interacting without cohesion. *Physical Review E*, 77(4):046113, 2008.

[11] E. Grimson, X. Wang, G.-W. Ng, and K. T. Ma. Trajectory analysis and semantic region modeling using a nonparametric bayesian model. 2008.

[12] T.-W. Hui, X. Tang, and C. Change Loy. Liteflownet: A lightweight convolutional neural network for optical flow estimation. In *CVPR*, pages 8981–8989, 2018.

[13] T. Kanade and C. Tomasi. Detection and tracking of point features. *IJCV*, 1991.

[14] P. Karpagavalli and A. Ramprasad. Estimating the density of the people and counting the number of people in a crowd environment for human safety. In *2013 International Conference on Communication and Signal Processing*, pages 663–667. IEEE, 2013.

[15] D. B. Kearns and R. Losick. Swarming motility in undomesticated bacillus subtilis. *Molecular microbiology*, 49(3):581–590, 2003.

[16] M. Keuper, S. Tang, B. Andres, T. Brox, and B. Schiele. Motion segmentation & multiple object tracking by correlation co-clustering. *TPAMI*, 2018.

[17] Z. Li, T. Dekel, F. Cole, R. Tucker, N. Snavely, C. Liu, and W. T. Freeman. Learning the depths of moving people by watching frozen people. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4521–4530, 2019.

[18] C. Lijun and H. Kaiqi. Video-based crowd density estimation and prediction system for wide-area surveillance. *China Communications*, 10(5):79–88, 2013.

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

[20] Y. Liu, M. Shi, Q. Zhao, and X. Wang. Point in, box out: Beyond counting persons in crowds. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6469–6478, 2019.

[21] C. Lu, J. Feng, Z. Lin, T. Mei, and S. Yan. Subspace clustering by block diagonal representation. *TPAMI*, 41(2):487–501, 2019.

[22] N. C. Makris, P. Ratilal, S. Jagannathan, Z. Gong, M. Andrews, I. Bertsatos, O. R. Godø, R. W. Nero, and J. M. Jech. Critical population density triggers rapid formation of vast oceanic fish shoals. *Science*, 323(5922):1734–1737, 2009.

[23] L. Mei, Z. Chen, and J. Lai. Geodesic-based probability propagation for efficient optical flow. *Electronics Letters*, 54(12):758–760, 2018.

[24] L. Mei, J. Lai, X. Xie, J. Zhu, and J. Chen. Illumination-invariance optical flow estimation using weighted regularization transform. *TCSVT*, 2019.

[25] B. T. Morris and M. M. Trivedi. Trajectory learning for activity understanding: Unsupervised, multilevel, and long-term adaptive approach. *TPAMI*, 33(11):2287–2301, 2011.

[26] A. V. Oppenheim, A. S. Willsky, and S. Nawab. Signals and systems (prentice-hall signal processing series). 1996.

[27] R. Sharma and T. Guha. A trajectory clustering approach to crowd flow segmentation in videos. In *ICIP*, pages 1200–1204. IEEE, 2016.

[28] V. A. Sindagi and V. M. Patel. Generating high-quality crowd density maps using contextual pyramid cnns. In *ICCV*, pages 1861–1870, 2017.

[29] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet. Novel type of phase transition in a system of self-driven particles. *Physical review letters*, 75(6):1226, 1995.

[30] L. Wang, Z. Ding, and Y. Fu. Low-rank transfer human motion segmentation. *TIP*, 28(2):1023–1034, 2019.

[31] Q. Wang, M. Chen, F. Nie, and X. Li. Detecting coherent groups in crowd scenes by multiview clustering. *TPAMI*, 2018.

[32] W. Wang, W. Lin, Y. Chen, J. Wu, J. Wang, and B. Sheng. Finding coherent motions and semantic regions in crowd scenes: A diffusion and clustering approach. In *ECCV*, pages 756–771. Springer, 2014.

[33] X. Wang, K. T. Ma, G.-W. Ng, and W. E. L. Grimson. Trajectory analysis and semantic region modeling using nonparametric hierarchical bayesian models. *IJCV*, 95(3):287–312, 2011.

[34] X. Wang, X. Ma, and W. E. L. Grimson. Unsupervised activity perception in crowded and complicated scenes using hierarchical bayesian models. *IEEE Transactions on pattern analysis and machine intelligence*, 31(3):539–555, 2008.

[35] J. Weickert. *Anisotropic diffusion in image processing*, volume 1. Teubner Stuttgart, 1998.

[36] S. Wu, H. Yang, S. Zheng, H. Su, Y. Fan, and M.-H. Yang. Crowd behavior analysis via curl and divergence of motion trajectories. *International Journal of Computer Vision*, 123(3):499–519, 2017.

[37] Y. Wu, Y. Wang, and Y. Jia. Adaptive diffusion flow active contours for image segmentation. *Computer Vision and Image Understanding*, 117(10):1421–1435, 2013.

[38] Y. Wu, Y. Ye, C. Zhao, and Z. Shi. Collective density clustering for coherent motion detection. *IEEE Trans. on Multimedia*, 20(6):1418–1431, 2018.

[39] C. Zhang, H. Li, X. Wang, and X. Yang. Cross-scene crowd counting via deep convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 833–841, 2015.

[40] H.-P. Zhang, A. Beer, E.-L. Florin, and H. L. Swinney. Collective motion and density fluctuations in bacterial colonies. *Proceedings of the National Academy of Sciences*, 107(31):13626–13630, 2010.

[41] W. Zhang, D. Zhao, and X. Wang. Agglomerative clustering via maximum incremental path integral. *Pattern Recognition*, 46(11):3056–3065, 2013.

[42] Y. Zhang, D. Zhou, S. Chen, S. Gao, and Y. Ma. Single-image crowd counting via multi-column convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 589–597, 2016.

[43] Z. Zhang and M. Li. Crowd density estimation based on statistical analysis of local intra-crowd motions for public area surveillance. *Optical Engineering*, 51(4):047204, 2012.

[44] B. Zhou, X. Tang, and X. Wang. Coherent filtering: Detecting coherent motions from crowd clutters. In *ECCV*, pages 857–871. Springer, 2012.

[45] B. Zhou, X. Tang, and X. Wang. Measuring crowd collectiveness. In *CVPR*, pages 3049–3056, 2013.

[46] B. Zhou, X. Tang, H. Zhang, and X. Wang. Measuring crowd collectiveness. *TPAMI*, 36(8):1586, 2014.

[47] B. Zhou, X. Wang, and X. Tang. Random field topic model for semantic region analysis in crowded scenes from tracklets. In *CVPR 2011*, pages 3441–3448. IEEE, 2011.

[48] B. Zhou, X. Wang, and X. Tang. Understanding collective crowd behaviors: Learning a mixture model of dynamic pedestrian-agents. In *CVPR*, pages 2871–2878, 2012.