

Measuring Crowd Collectiveness

Bolei Zhou, Xiaou Tang, *Fellow, IEEE*, Hepeng Zhang, and Xiaogang Wang, *Member, IEEE*

Abstract—Collective motions of crowds are common in nature and have attracted a great deal of attention in a variety of multidisciplinary fields. Collectiveness, which indicates the degree of individuals acting as a union, is a fundamental and universal measurement for various crowd systems. By quantifying the topological structures of collective manifolds of crowd, this paper proposes a descriptor of collectiveness and its efficient computation for the crowd and its constituent individuals. The *Collective Merging* algorithm is then proposed to detect collective motions from random motions. We validate the effectiveness and robustness of the proposed collectiveness on the system of self-driven particles as well as other real crowd systems such as pedestrian crowds and bacteria colony. We compare the collectiveness descriptor with human perception for collective motion and show their high consistency. As a universal descriptor, the proposed crowd collectiveness can be used to compare different crowd systems. It has a wide range of applications, such as detecting collective motions from crowd clutters, monitoring crowd dynamics, and generating maps of collectiveness for crowded scenes. A new Collective Motion Database, which consists of 413 video clips from 62 crowded scenes, is released to the public.

Index Terms—Crowd behavior analysis, collective motion, video analysis, graph connectivity.

1 INTRODUCTION

One of the most captivating phenomena in nature is the collective motions of crowds. From bacterial colonies and insect swarms to fish shoals, collective motions widely exist in different crowd systems and reflect the ordered macroscopic behaviors of constituent individuals. Many interdisciplinary efforts have been made to explore the underlying principles of this phenomenon. Physicists treat crowds as sets of particles and use equations from fluid mechanics to characterize individual movements and their interactions [1]. Behavioral studies show that complex crowd behaviors may result from repeated simple interactions among its constituent individuals, *i.e.*, individuals locally coordinate their behaviors with their neighbors, and then the crowd is self-organized into collective motions without external control [2], [3]. Meanwhile, animal aggregation is considered as an evolutionary advantage for species survival, since the integrated whole of individuals can generate complex patterns, quickly process information, and engage in collective decision-making [4].

One remarkable observation of collective motions in various crowd systems is that some spatially coherent structures emerge from the movements of different individuals, such as the arch-like geometric structures illustrated in Fig.1. We refer to the spatially coherent structures of collective motions as *Collective Manifold*. One of the important structural properties of collective manifolds is that behavioral consistency remains high among individuals in local neighborhoods, but low among those that are far

apart, even if they are on the same collective manifold. In fact, individuals in crowds only have limited sensing range, then they often base their movements on locally acquired information such as the positions and motions of their neighbors. However, information can propagate to distant areas through local interactions and coordination, which lead to the collective motion of the whole crowd. Some empirical studies have explored the importance of topological relations and information transmission among neighboring individuals in crowd [5], [6]. However, there is a lack of quantitative analysis of the collective manifold and its structural properties corresponding to the overall crowd dynamics.

Collectiveness describes the degree of individuals acting as a union in collective motions. It depends on multiple factors, such as the decision making process of individuals, crowd density, and scene structures. Quantitatively measuring this universal property and comparing it across different crowd systems play an important role to understand the general principles of various crowd behaviors. It is important to many applications, such as monitoring the transition of a crowd system from disordered to ordered states, studying correlation between collectiveness and other crowd properties such as population density, characterizing the dynamic evolution of collective motions, and comparing the collectiveness of different crowd systems. Most existing crowd surveillance technologies [7], [8] cannot compare crowd behaviors across different scenes because they lack universal descriptors with which to characterize the crowd behaviors. Monitoring collectiveness is also useful in crowd management, control of swarming desert locusts [9], prevention of disease spreading [10], and many other fields. However, this important property lacks accurate measurements. Existing works [6], [11] simply measure the average velocity of all the individuals to indicate the collectiveness of the whole crowd, which is neither accurate nor robust. The collectiveness of individuals in crowd is

- B. Zhou is with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, USA.
- X. Tang is with the Department of Information Engineering, the Chinese University of Hong Kong, Hong Kong.
- H. Zhang is with the Physics Department, Shanghai Jiao Tong University, China.
- X. Wang is with the Department of Electronic Engineering, the Chinese University of Hong Kong, Hong Kong.

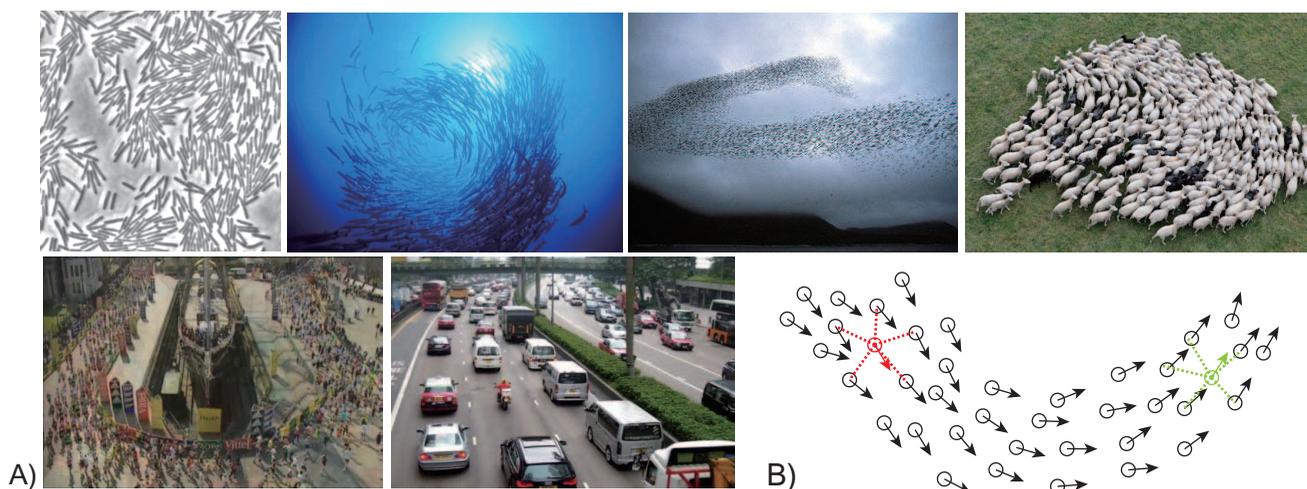


Fig. 1. A) Collective motions of the bacterial colony, fish shoal, bird flock, sheep herd, athletic group, and traffic flow. B) One common spatially coherent structure, *i.e.*, *Collective Manifold*, emerges in these different crowds. Since individuals in a crowd system only coordinate their behaviors with their local neighbors, individuals at a distance may have low velocity correlation even though they are on the same collective manifold. Consider the red and green individuals as examples. By characterizing the structural property of collective manifolds in various crowds, this work aims at proposing a general measurement of *collectiveness* for different crowd systems and its efficient computation. Colored dash links represent neighborhoods.

also ill-defined.

In this paper, by quantifying the structural properties of collective manifolds of crowds, we propose a descriptor of collectiveness for crowds as well as their constituent individuals. Based on collectiveness, we propose an algorithm called *Collective Merging* to detect collective motions from random motions. We validate the effectiveness and robustness of the proposed collectiveness on self-driven particles [11]. It is further compared with human motion perception on a new Collective Video Database with ground-truth. In addition, our experiments of detecting collective motions and measuring crowd collectiveness in videos of pedestrian crowds and bacterial colonies demonstrate the wide applications of the collectiveness descriptor¹.

The novelty and contributions of this work are summarized as follows. 1) Rather than learning/extracting motion patterns from time-series data, we propose a general descriptor of collectiveness to measure crowd dynamics, which is based on a common observation of collective manifolds in collective motions of different crowds. 2) This collectiveness descriptor along with the collective motion detection algorithm can be effectively applied to a variety of data (such as self-driven particle systems, pedestrian crowd videos, natural scene videos, and real bacteria motion data) and a range of novel applications (such as detecting collective motions, monitoring crowd dynamics, and generating collective maps of scenes). 3) By applying this collectiveness descriptor to motion analysis of bacterial colony, we illustrate that our work has potential contributions to the scientific studies of collective motions. 4) The first video database of evaluating collectiveness of various crowd systems is introduced to the computer

vision community. The conference version of this work was published in [12] as an oral presentation. More technical details, theoretic analysis, experimental evaluations and applications are provided in this paper.

2 RELATED WORKS

Crowd behavior analysis has recently become a hot topic in computer vision because of the large demands on crowd surveillance. Many studies [13], [14] show that various crowd systems do share a set of universal properties because some general principles underlie different types of crowd behaviors. Quantitatively measuring such properties and comparing them across different crowd systems is important for understanding the underlying general principles of various crowd behaviors and plays an important role in surveillance applications.

2.1 Scientific Studies on Collective Motions

The remarkable collective motions of organisms have long captivated the attention of scientists from different scientific fields. Understanding the collective behaviors of crowds is a fundamental problem in social science. It has shown that individuals staying in crowd tend to lose their personalities. Instead of behaving independently they tend to follow the behaviors of others and move along the same direction as their neighbors [15], [16]. Some collective behaviors of crowds such as herding have been studied in the field of social psychology [17].

In biology, considerable progress has been made in revealing the principles of collective behaviors of different crowds using empirical or theoretical approaches. A compact review can be found in [18]. As for empirical approaches, behavior data of different crowds such as

1. Data and codes are available at <http://mmlab.ie.cuhk.edu.hk/project/collectiveness/>

bacterial colonies [10], locust swarm [19], fish shoals [20], and bird flocks [6] has been collected and analyzed. People study the mechanism underlying the collective organization of individuals [4], the evolutionary origin of animal aggregation [14] and collective information processing in crowds [2] at both macroscopic and microscopic levels. Many important factors and laws affecting collective motions have also been discovered and analyzed, such as phase transition, criticality of crowd density, and self-organization [21], [22]. Zhang *et al.* [10] studied the correlation between collective motions and bacteria densities. Makris *et al.* [20] conducted quantitative study on the collective temporal and spatial processes formed by vast oceanic shoals.

As for theoretical approaches, differential equations of continuum [23], statistical physics models, such as the maximum entropy model [24], have been proposed to understand the mechanism of collective motions. Besides, based on some local behavioral rules many agent-based models have been proposed for crowd modeling and simulation [3], [11], [25]. Vicsek *et al.* [11] and Chate *et al.* [26] proposed self-driven particle (SDP) models and observed a rapid phase transition from disordered to highly synchronized behaviours in crowd systems as crowd density reaches a critical level. Understanding such a transition is crucial to many applications related to crowd systems, such as management of locust outbreaks. Detailed quantitative analysis of collective motions becomes essential in order to understand when, how and where they happen and how to improve the control of crowd systems. Buhl *et al.* [9] videotaped locust motions, empirically studied the transition phase and validated SDP.

In the field of network science and control theory, the flock of interacting units (human beings, robots, agents, and dynamic systems) are connected into dynamically changing networks. Then, community structures with shared collective behaviors are detected [27]. The community and network evolving dynamics are further analyzed [28]. Meanwhile, flocking and consensus algorithms from control theory have been proposed to control these distributed complex systems in which these units can be synchronized into collective movements despite of the absence of centralized coordination [29].

Given the huge literature on collective behaviors, however, the study on quantitative measurements of collectiveness is very limited. Most existing works [11] simply measure collectiveness as the average velocity of particles in a crowd system, and assume that the magnitude of the average velocity of a disordered crowd system is close to zero. Petitjean *et al.* [30] computed it as the average of $\cos(\phi_i)$ where ϕ_i is the moving direction of each individual i . Such measurements cannot deal with collective motions with manifold structures or a mixture of heterogeneous collective patterns as shown in Figure 1.

2.2 Collective Motions in Computer Vision

In computer vision, a lot of works have been done on learning global motion patterns related to crowd behaviors

[7], [8], [31]–[39], detecting coherent or incoherent motions from crowd clutters [40]–[49], and analyzing interactions among individuals in crowd [25], [50]–[57]. A brief review is given below.

There has been significant amount of work on learning the motion patterns of crowd behaviors. Ali *et al.* [31] and Lin *et al.* [7], [58] modeled crowd flows with Lagrangian coherent structures or Lie algebra based on computed flow fields. Mehran *et al.* [32] proposed a streakline representation for crowd flows. With topic models, Wang *et al.* [33] explored the co-occurrence of moving pixels to learn the motion patterns in crowd. Topic models have been augmented by adding spatio-temporal dependency among motion patterns [34], [35]. Some approaches [8], [36]–[39] learn motion patterns through clustering trajectories or tracklets in crowded scenes. For example, Zhou *et al.* [8] used a mixture of dynamic systems to learn pedestrian dynamics and applied it to crowd simulation. However, none of the above-mentioned approaches measured the collectiveness of crowd behaviors or explored its potential applications.

Meanwhile, detecting coherent and incoherent (abnormal) behaviors in crowd is of great interests in surveillance and crowd management. Rabaud *et al.* [44] and Brostow *et al.* [45] detected independent motions in order to count moving objects. Zhou *et al.* [46] proposed a graph-based method to detect coherent motions from tracklets. Brox *et al.* [47] extended spectral clustering to group long-term dense trajectories for the segmentation of moving objects in videos. These coherent motion detection methods extract and cluster collectively moving targets from randomly moving points. Some approaches have been proposed to model local spatio-temporal variations for abnormality detection with dynamic texture [48], [49], HMM [43], distributions of spatio-temporal oriented energy [42], chaotic invariants [41], and local motion descriptors [40]. These methods are scene-specific and their features or descriptors cannot be used to compare crowd videos captured from different scenes.

Individuals in social groups have the same destination and closer relationship. They are more likely to form collective behaviors. To analyze interactions and social influence among pedestrians, the social force model, first proposed by Helbing *et al.* [25] for crowd simulation, was introduced to the computer vision community recently and was applied to multi-target tracking [56] and abnormality detection [55]. Ge *et al.* [52] proposed a hierarchical clustering method to detect groups and Chang *et al.* [59] proposed a probabilistic strategy to softly assign individuals into groups. Moussaid *et al.* [60] modified the social force model to account for the influence of social groups. Lan *et al.* [50] analyzed individual behaviors considering the context of social groups with hierarchical models. Recently Kratz *et al.* [54] proposed *efficiency* to measure the difference between the actual motion and intended motion of pedestrians in crowd for tracking and abnormality detection.

In computer graphics, simulating collective behaviors of virtual crowds has attracted many attentions due to its wide

applications to digital entertainment, emergency training, and urban planning. A recent survey can be found in [61]. Agent-based models are commonly used to model complex behaviors of crowd [3], [62], [63]. The seminal work of Reynolds [3] demonstrated emergent flocking and other collective behaviors using simple local rules. Flow-based models have been used to simulate crowd flows and dynamics [64]. Some approaches can learn model parameters from the real-world data of crowd movements then simulate the crowd [65]. Recently an information-theoretic metric has been proposed to measure the similarity between real-world crowd movements and crowd simulation results [66]. It is related to our collectiveness measurement. Rather than directly comparing crowd movements as [66], we first compute the collectiveness of a single crowd, then compare the collectiveness of different crowds. It would be interesting to integrate these two measurements in the future to estimate cross-crowd collectiveness.

3 THEORY OF COLLECTIVENESS

3.1 Emergence of Collective Manifolds in Crowds

As shown in Fig.1A, from lower-level bacterial colony to higher-level human crowd, one common observation of collective motions in different crowd systems is that some spatiotemporal coherent structures emerge from the collective movements of constituent individuals in crowds. We define such structures as *collective manifolds* of collective motions. There are two key properties of collective manifolds, which distinguish collective motions from random movements.

- Behavior consistency in neighborhoods: neighboring constituent individuals have consistent behaviors.
- Global consistency among non-neighbors: although constituent individuals at a distance may have different behaviors, they are correlated by behavior similarity through intermediate individuals in neighborhoods on the manifold.

Thus individuals in a crowd system coordinate their behaviors in local neighborhood, but the behavior consistency between individuals at a distance may have some uncertainty, such as the low velocity correlation between the red and green individuals illustrated in Fig.1B.

Crowd collectiveness, described as the degree of individuals acting as a union in collective motion, measures the holistic behavior consistency of collective manifold. Meanwhile, individual collectiveness measures the behavior consistency of each individual with all the other individuals. But due to the behavior uncertainty between individuals at a distance, we could not directly measure the individual collectiveness, which prevents the estimation of crowd collectiveness. To handle this problem, we study behavior consistency along paths on the collective manifolds. Thus, in our framework collectiveness is measured in a bottom-up way: from behavior consistency in neighborhood of individuals to behavior consistency among all pairwise individuals along paths in the crowd, then from individual collectiveness to crowd collectiveness.

3.2 Behavior Consistency in Neighborhood

We first measure the similarity of individual behaviors in neighborhood. When individual j is in the neighborhood of i , i.e., $j \in \mathcal{N}(i)$ at time t , the similarity is defined as

$$w_t(i, j) = \max(C_t(i, j), 0), \quad (1)$$

where $C_t(i, j)$ is the velocity correlation coefficient at time t between i and j , i.e., $C_t(i, j) = \frac{\mathbf{v}_i \cdot \mathbf{v}_j}{\|\mathbf{v}_i\|_2 \|\mathbf{v}_j\|_2}$. \mathcal{N} is defined as K -nearest-neighbor, motivated by existing empirical studies of collective motions, which have shown that animals maintain local interaction among neighbors with a fixed number of neighbors on topological distance, rather than with all neighbors within a fixed spatial distance [5]. Thus, $w_t(i, j) \in [0, 1]$ measures an individual's behavior consistency in its neighborhood.

In fact, K represents the topological range of neighborhood. A large K may result in inaccurate estimation of behavior consistency in neighborhood. Later we will have a discussion on how K affects the estimation of collectiveness.

Because of the behavior uncertainty between non-neighbors, this pairwise similarity estimation would be unreliable if two individuals are not in neighborhood. A better behavior similarity based on the connectivity of collective manifold is proposed below.

3.3 Global Consistency via Paths

Since behavior consistency cannot be directly estimated when two individuals are not in neighborhood, we propose a new pairwise similarity based on an important structure of collective manifold: paths, which represent the connectivity of the network associated with a graph [67]. In crowd systems, paths have important roles in characterizing the behavior consistency among individuals outside neighborhood in crowds.

Let \mathbf{W} be the weighted adjacency matrix of the graph, where an edge $w_t(i, j)$ is the similarity between individual i and j in its neighborhood defined in Eq.1. Let $\gamma_l = \{p_0 \rightarrow p_1 \rightarrow \dots \rightarrow p_l\}$ ($p_0 = i, p_l = j$) denote a path of length l through nodes p_0, p_1, \dots, p_l on \mathbf{W} between individual i and j . Then $\nu_{\gamma_l} = \prod_{k=0}^{l-1} w_t(p_k, p_{k+1})$ is defined as the *path similarity* on a *specific* path γ_l . Fig.2 illustrates five specific paths of four different lengths between two individuals which have no neighborhood relation.

Since there can be more than one path of length l between i and j , let the set \mathcal{P}_l contain all the paths of length l between i and j , then the l -path similarity is defined as

$$\nu_l(i, j) = \sum_{\gamma_l \in \mathcal{P}_l} \nu_{\gamma_l}(i, j). \quad (2)$$

$\nu_l(i, j)$ can be efficiently computed with Theorem 1.

Theorem 1. $\nu_l(i, j)$ is the (i, j) entry of matrix \mathbf{W}^l .

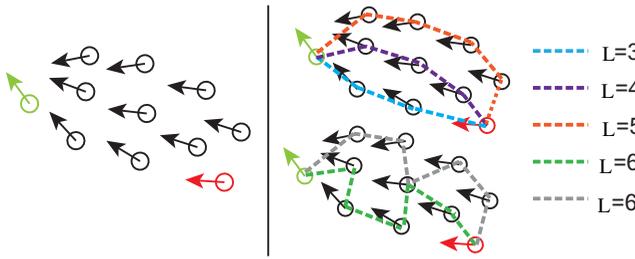


Fig. 2. Paths connecting two individuals. Colored dashed lines shows some representative paths of length $l = 3, 4, 5, 6$ between red and green points. The paths could be longer and more complicated. Only a few of them are shown due to space limit. Red point and green points have no neighborhood relation, but their behavior consistency can be measured via integrating the path similarities between them.

Proof: According to the algebraic graph theory [67],

$$\mathbf{W}^l(i, j) = \sum_{\substack{\gamma_l \in \mathcal{P}_l \\ p_0=i, p_l=j}} \prod_{k=0}^{l-1} w_i(p_k, p_{k+1}) = \sum_{\gamma_l \in \mathcal{P}_l} \nu_{\gamma_l}(i, j) = \nu_l(i, j),$$

where set \mathcal{P}_l contains all the paths with length l between p_0 and p_l . $\mathbf{W}^l(i, j)$ is the sum of the products of the weights over all paths with length l that start at vertex i and finish at vertex j in the weighted graph \mathbf{W} . \square

3.4 Collectiveness at l -path scale

Since l -path similarity $\nu_l(i, j)$ measures the behavior consistency between i and j at l -path scale, we define the individual collectiveness of individual i at l -path scale as

$$\phi_l(i) = \sum_{j \in \mathcal{C}} \nu_l(i, j) = [\mathbf{W}^l \mathbf{e}]_i. \quad (3)$$

Here \mathcal{C} is the set containing all the individuals in the crowd and \mathbf{e} is a vector with all elements as 1, $[\cdot]_i$ denotes i -th element of a vector. Then we define the crowd collectiveness at l -path scale as the mean of individual collectiveness

$$\Phi_l = \frac{1}{|\mathcal{C}|} \mathbf{e}^\top \mathbf{W}^l \mathbf{e}. \quad (4)$$

To further measure the overall collectiveness of individuals and crowd, intuitively we should integrate the individual collectiveness and crowd collectiveness over all the path scales, *i.e.*, $l = \{1, 2, \dots, \infty\}$. However, due to the exponential growth of ϕ_l and Φ_l with l shown in Property 2 and Property 3 of Collectiveness in Section 4, individual collectiveness at different path scales cannot be directly summed. Therefore, we define a generating function with regularization to integrate all path similarities, such that the convergence of summation can be guaranteed.

3.5 Regularizing Collectiveness over All Scales

Generating function regularization is used to assign a meaningful value for the sum of a possibly divergent series

[68]. There are different forms of generating functions. We define the generating function for the l -path similarities as

$$\tau_{i,j} = \sum_{l=1}^{\infty} z^l \nu_l(i, j), \quad (5)$$

where z is a real-valued regularization factor, and z^l can be interpreted as the weight for l -path similarity. $z < 1$ and cancels the effect that ϕ_l and Φ_l exponentially grow with l . $\tau_{i,j}$ can be computed with Theorem 2.

Theorem 2. $\tau_{i,j}$ is the (i, j) entry of matrix \mathbf{Z} , where $\mathbf{Z} = (\mathbf{I} - z\mathbf{W})^{-1} - \mathbf{I}$ and $0 < z < \frac{1}{\rho(\mathbf{W})}$, where ρ is the spectral radius of matrix \mathbf{W} .

Proof: Let the eigendecomposition of the matrix \mathbf{W} be $\mathbf{W} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1}$, and $\lambda_{1 \sim |\mathcal{C}|}$ be the diagonal elements of the diagonal matrix $\mathbf{\Lambda}$. The matrix summation is

$$\begin{aligned} \mathbf{Z} &= z\mathbf{W} + z^2\mathbf{W}^2 + \dots + z^\infty\mathbf{W}^\infty \\ &= z\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1} + z^2\mathbf{Q}\mathbf{\Lambda}^2\mathbf{Q}^{-1} + \dots + z^\infty\mathbf{Q}\mathbf{\Lambda}^\infty\mathbf{Q}^{-1} \\ &= \mathbf{Q}(\mathbf{I} + z\mathbf{\Lambda} + z^2\mathbf{\Lambda}^2 + \dots + z^\infty\mathbf{\Lambda}^\infty)\mathbf{Q}^{-1} - \mathbf{I} \\ &= \mathbf{Q} \begin{bmatrix} \sum_{l=0}^{\infty} (z\lambda_1)^l & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 & \sum_{l=0}^{\infty} (z\lambda_{|\mathcal{C}|})^l \end{bmatrix} \mathbf{Q}^{-1} - \mathbf{I} \\ &= \mathbf{Q} \begin{bmatrix} \frac{1}{1-z\lambda_1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 & \frac{1}{1-z\lambda_{|\mathcal{C}|}} \end{bmatrix} \mathbf{Q}^{-1} - \mathbf{I} \\ &= (\mathbf{I} - z\mathbf{W})^{-1} - \mathbf{I}, \end{aligned}$$

where $z < \frac{1}{\rho(\mathbf{W})}$ guarantees the convergence of series $\sum_{l=0}^{\infty} (z\lambda_i)^l$ as $\frac{1}{1-z\lambda_i}$. \square

Thus, the individual collectiveness from the generating function regularization on all the path similarities can be written as

$$\phi(i) = \sum_{l=1}^{\infty} z^l \phi_l(i) = [\mathbf{Z}\mathbf{e}]_i. \quad (6)$$

Crowd collectiveness of a crowd system \mathcal{C} is defined as the mean of all the individual collectiveness, which can be explicitly written in a closed form as,

$$\Phi = \frac{1}{|\mathcal{C}|} \sum_{i=1}^{|\mathcal{C}|} \phi(i) = \frac{1}{|\mathcal{C}|} \mathbf{e}^\top ((\mathbf{I} - z\mathbf{W})^{-1} - \mathbf{I}) \mathbf{e}. \quad (7)$$

A strong convergence condition of collectiveness will be given in Property 1 in Section 4.

4 PROPERTIES OF THE COLLECTIVENESS

We derive some important properties of collectiveness.

Property 1. (Strong Convergence Condition) \mathbf{Z} converges when $z < \frac{1}{K}$.

Proof: From Lemma 1, we know that $\rho(\mathbf{W}) < \rho(\mathbf{A}) = K$. Thus, when $z < \frac{1}{K} \leq \frac{1}{\rho(\mathbf{W})}$, according to Theorem 2, \mathbf{Z} converges. \square

Lemma 1. $\rho(\mathbf{W}) \leq \rho(\mathbf{A}) = K$, where \mathbf{A} is any $(0,1)$ -adjacency matrix according to K nearest neighbors, \mathbf{W} is the weighted adjacency matrix corresponding to \mathbf{A} , and all the weights $w(i, j) \in [0, +1]$.

Proof: Since adjacency matrix \mathbf{A} is computed from K nearest neighbors, we have $\mathbf{Ae} = K\mathbf{e}$, thus K is an eigenvalue of \mathbf{A} . Let λ_i be any eigenvalue of \mathbf{A} , since $a_{ii} = 0$ according to Gershgorin Circle Theorem we have

$$|\lambda_i| \leq \sum_{j \neq i} |a_{ij}| = K.$$

Now we know $K = \max_i (|\lambda_i|)$, thus $\rho(\mathbf{A}) = K$. Similarly, for any eigenvalue λ_i of \mathbf{W} , $\max_i (|\lambda_i|) \leq \sum_{j \neq i} |w_{ij}| \leq K$. Thus $\rho(\mathbf{W}) \leq \rho(\mathbf{A}) = K$ \square

Note that it is computationally expensive to choose z by comparing it with $\rho(\mathbf{W})$, especially for a large crowd system, since we need to compute the eigenvalues of \mathbf{W} to get $\rho(\mathbf{W})$ with complexity $\mathcal{O}(n^3)$. Because of Property 1, the value of z can be determined without computing $\rho(\mathbf{W})$.

Property 2. (Bounds of Φ_l) $0 \leq \Phi_l \leq K^l$.

Proof: Let \mathbf{A} be the $(0,1)$ K -nearest-neighbor adjacency matrix corresponding to \mathbf{W} . Then we easily have $\mathbf{Ae} = K\mathbf{e}$. We use mathematical induction to prove the statement $\mathbf{e}^\top \mathbf{W}^n \mathbf{e} \leq |\mathcal{C}|K^n$ holds for any n as follows.

(1) When $n = 1$, since $w(i, j) \leq a(i, j) = 1$ if $j \in \mathcal{N}(i)$, we have

$$\mathbf{e}^\top \mathbf{W} \mathbf{e} \leq \sum_{i,j} w(i, j) \leq \sum_{i,j} a(i, j) = \mathbf{e}^\top \mathbf{A} \mathbf{e} = |\mathcal{C}|K.$$

(2) When $n = m$, we assume the statement $\mathbf{e}^\top \mathbf{W}^m \mathbf{e} \leq |\mathcal{C}|K^m$ holds, and let $w_m(i, j)$ denote the (i, j) entry of \mathbf{W}^m , then

$$\begin{aligned} \mathbf{e}^\top \mathbf{W}^{m+1} \mathbf{e} &= \mathbf{e}^\top \mathbf{W}^m \mathbf{W} \mathbf{e} = \sum_{i,j,k} w_m(i, k) w(k, j) \\ &\leq \sum_{i,j,k} w_m(i, k) a(k, j) = \mathbf{e}^\top \mathbf{W}^m \mathbf{A} \mathbf{e} \\ &= K \mathbf{e}^\top \mathbf{W}^m \mathbf{e} = |\mathcal{C}|K^{m+1}. \end{aligned}$$

Then the statement $\mathbf{e}^\top \mathbf{W}^{m+1} \mathbf{e} \leq |\mathcal{C}|K^{m+1}$ holds. Thus $\mathbf{e}^\top \mathbf{W}^n \mathbf{e} \leq |\mathcal{C}|K^n$ holds for any n . Meanwhile, it is obvious to see $\Phi_l \geq 0$.

We proved that $\Phi_l = \frac{1}{|\mathcal{C}|} \mathbf{e}^\top \mathbf{W}^l \mathbf{e} \leq K^l$. The upper bound is reached when $\mathbf{W} = \mathbf{A}$, where \mathbf{A} is $(0,1)$ adjacency matrix according to K -nearest-neighbor. \square

Property 3. (Asymptotic limit of Φ_l) $\lim_{l \rightarrow \infty} \frac{\ln \Phi_l}{l} = \ln \lambda \leq \ln K$, where λ is the largest eigenvalue of \mathbf{W} .

Proof: According to the Perron-Frobenius theorem, we have $\lim_{l \rightarrow \infty} \frac{\mathbf{W}^l}{\lambda^l} = v w^\top$, where v and w are left and right eigenvectors of \mathbf{W} corresponding to λ and are normalized to $w^\top v = 1$. Then we have

$$\begin{aligned} \lim_{l \rightarrow \infty} \frac{\ln \Phi_l}{l} - \ln \lambda &= \lim_{l \rightarrow \infty} \frac{\ln \frac{\Phi_l}{\lambda^l}}{l} = \lim_{l \rightarrow \infty} \frac{\ln \frac{\mathbf{e}^\top \mathbf{W}^l \mathbf{e}}{|\mathcal{C}| \lambda^l}}{l} \\ &= \lim_{l \rightarrow \infty} \frac{\ln \frac{\mathbf{e}^\top v w^\top \mathbf{e}}{|\mathcal{C}|}}{l} = 0. \end{aligned}$$

This quantity is related to the topological entropy of a graph \mathbf{W} [69], where the maximal entropy rate of random walk on the graph is bounded by $\ln \lambda$. It can be interpreted as that as collective motion emerges among individuals of crowd, the links of the whole crowd become homogenous, and the random surfer on the collective manifold reaches the maximal entropy rate. It also shows the exponential growth of Φ_l with l .

Property 4. (Bounds of Φ) $0 \leq \Phi \leq \frac{zK}{1-zK}$, if $z < \frac{1}{K}$.

Proof: From the proof of Property 2 we know that $\mathbf{e}^\top \mathbf{W}^n \mathbf{e} \leq |\mathcal{C}|K^n$ holds for any n . By expanding $\mathbf{e}^\top \mathbf{Z} \mathbf{e}$ we have

$$\begin{aligned} \Phi &= \frac{1}{|\mathcal{C}|} (z \mathbf{e}^\top \mathbf{W} \mathbf{e} + z^2 \mathbf{e}^\top \mathbf{W}^2 \mathbf{e} \dots + z^\infty \mathbf{e}^\top \mathbf{W}^\infty \mathbf{e}) \\ &\leq zK + z^2 K^2 + \dots + z^\infty K^\infty = \frac{zK}{1-zK}. \end{aligned} \quad (8)$$

And it is obvious that $\Phi \geq 0$. \square

The equality stands when $\mathbf{W} = \mathbf{A}$, where \mathbf{A} is the $(0,1)$ adjacency matrix according to K -nearest-neighbor. It indicates that there are perfect velocity correlations among neighbors, i.e., $w_t(i, j) = 1$ if $j \in \mathcal{N}(i)$ for any i , and thus all the constituent individuals in neighborhood move in the same direction.

Note that K determines the topological range of neighborhood. Different K and z result in different $\sup \Phi \equiv \frac{zK}{1-zK}$. We can define the normalized crowd collectiveness $\hat{\Phi} = \frac{\Phi}{\sup \Phi}$. Then The range of $\hat{\Phi}$ is scaled to $[0, 1]$. For simplicity, in most of our experiments we let $K = 20$ and $z = 0.025$, so the range of Φ is $[0, 1]$. Relations among Φ , K and z are discussed in Section 6.3.

Property 5. (Upper bound of entries of \mathbf{Z}) $\tau_{i,j} < \frac{z}{1-zK}$, for every entry (i,j) of \mathbf{Z} .

Proof: Let the eigendecomposition of matrix \mathbf{W} be $\mathbf{W} = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^{-1}$, and $\lambda_{1 \sim |\mathcal{C}|}$ be the diagonal elements of the diagonal matrix $\mathbf{\Lambda}$. From Lemma 1 we know that $K > \max_i (\lambda_i)$. Thus we have

$$\begin{aligned} \mathbf{Z} &= (\mathbf{I} - z\mathbf{W})^{-1} - \mathbf{I} \\ &= \mathbf{Q} \begin{bmatrix} \frac{1}{1-z\lambda_1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 & \frac{1}{1-z\lambda_{|\mathcal{C}|}} \end{bmatrix} \mathbf{Q}^{-1} - \mathbf{I} \\ &= \mathbf{Q} \begin{bmatrix} \frac{z\lambda_1}{1-z\lambda_1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 & \frac{z\lambda_{|\mathcal{C}|}}{1-z\lambda_{|\mathcal{C}|}} \end{bmatrix} \mathbf{Q}^{-1} \\ &< \mathbf{Q} \begin{bmatrix} \frac{z\lambda_1}{1-zK} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 & \frac{z\lambda_{|\mathcal{C}|}}{1-zK} \end{bmatrix} \mathbf{Q}^{-1} \\ &= \frac{z}{1-zK} \mathbf{Q} \begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 & \lambda_{|\mathcal{C}|} \end{bmatrix} \mathbf{Q}^{-1} = \frac{z}{1-zK} \mathbf{W}. \end{aligned} \quad (9)$$

Since the maximum entry on \mathbf{W} is no larger than 1, $\tau_{i,j} < \frac{z}{1-zK}$. \square

Algorithm 1 Collective Merging

INPUT: $\{\mathbf{x}_i, \mathbf{v}_i | i \in \mathcal{C}\}_t$.
 OUTPUT: clusters of individuals $\{\mathcal{F}_1, \dots, \mathcal{F}_N\}$.
 1: Compute \mathbf{W} from K -NN using Eq. 1.
 2: $\mathbf{Z} = (\mathbf{I} - z\mathbf{W})^{-1} - \mathbf{I}$.
 3: Set the entry $\mathbf{Z}(i, j)$ to 1 if $\mathbf{Z}(i, j) \geq \kappa$, otherwise to 0.
 4: Search the connected components of the thresholded \mathbf{Z} as the collective clusters $\{\mathcal{F}_1, \dots, \mathcal{F}_N\}$.

This property will be used in the following algorithm of detecting collective motion patterns from crowd clutters.

Property 6. (Approximate error bound of \mathbf{Z}) $\|\mathbf{Z} - \mathbf{Z}_{1 \sim n}\|_2 \leq \frac{(z\sigma(\mathbf{W}))^{n+1}}{1 - z\sigma(\mathbf{W})}$, where $\mathbf{Z}_{1 \sim n}$ denotes the sum of first n powers of \mathbf{W} and $\sigma(\mathbf{W})$ is the largest singular value of matrix \mathbf{W} .

Proof: Since $\mathbf{Z}_{1 \sim n} \equiv z\mathbf{W} + z^2\mathbf{W}^2 + \dots + z^n\mathbf{W}^n$, we can get $\mathbf{Z}_{1 \sim n} - \mathbf{Z}_{1 \sim n}z\mathbf{W} = z\mathbf{W} - z^{n+1}\mathbf{W}^{n+1}$. Then

$$\mathbf{Z}_{1 \sim n} = (\mathbf{I} - z\mathbf{W})^{-1}(z\mathbf{W} - (z\mathbf{W})^{n+1}). \quad (10)$$

Meanwhile, with $\mathbf{Z} = (\mathbf{I} - z\mathbf{W})^{-1} - \mathbf{I} = (\mathbf{I} - z\mathbf{W})^{-1}z\mathbf{W}$, we get

$$\begin{aligned} \|\mathbf{Z} - \mathbf{Z}_{1 \sim n}\|_2 &= \|(\mathbf{I} - z\mathbf{W})^{-1}(z\mathbf{W})^{n+1}\|_2 \\ &\leq \|(\mathbf{I} - z\mathbf{W})^{-1}\|_2 \|(z\mathbf{W})^{n+1}\|_2 \\ &= \frac{1}{1 - \|z\mathbf{W}\|_2} \|z\mathbf{W}\|_2^{n+1} = \frac{(z\sigma(\mathbf{W}))^{n+1}}{1 - z\sigma(\mathbf{W})}. \end{aligned} \quad (11)$$

Here $\|z\mathbf{W}\|_2 = z\sigma(\mathbf{W})$ and $zK < 1$. □

According to this property, we can approximate \mathbf{Z} by $\mathbf{Z}_{1 \sim n}$ with the error bound $\frac{(z\sigma(\mathbf{W}))^{n+1}}{1 - z\sigma(\mathbf{W})}$.

5 COLLECTIVE MOTION DETECTION

Based on the proposed collectiveness, an algorithm called Collective Merging is proposed to detect collective motions from crowd clutters with random motions (see Algorithm 1). The algorithm in the case of single frames is summarized in Algorithm 1. Given the spatial locations \mathbf{x}_i and velocities \mathbf{v}_i of individuals i at time t , we first compute \mathbf{W} . Then by thresholding the values on \mathbf{Z} , we can easily remove outlier particles with low collectiveness and get the clusters of collective motion patterns as the connected components from thresholded \mathbf{Z} . As for the threshold κ , according to the bound in Property 3 we let $\kappa = \frac{\alpha z}{1 - zK}$ where $0.4 < \alpha < 0.8$. On a Intel Core i5-3210M laptop, this four-lined algorithm in Matlab implementation runs as 10 FPS with $|\mathcal{C}| = 500$. It does not include the time of tracking keypoints. The main computation bottleneck comes from the matrix inversion, which can be solved with current fastest algorithm with $O(n^{2.38})$ time [70]. In the experiment section, we demonstrate its effectiveness on detecting collective motion patterns on various videos.

6 NUMERICAL ANALYSIS ON CROWD SYSTEMS OF SELF-DRIVEN PARTICLES

We take the Self-Driven Particle model (SDP) [11] to evaluate the proposed collectiveness, because SDP has been used extensively for studying collective motion and

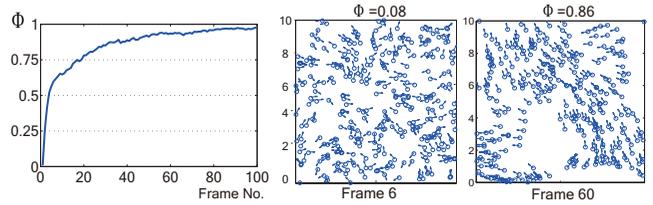


Fig. 3. Emergence of collective motion in SDP. At the beginning, Φ is low since the spatial locations and moving directions of individuals are randomly assigned. The behaviors of individuals gradually turn into collective motion from random movements, and Φ accurately reflects the phase transition of crowd dynamics. Here $K = 20$, $z = 0.025$, and $\eta = 0$, then the upper bound of Φ is 1.

shows high similarity with various crowd systems in nature [9], [10]. Importantly, the groundtruth of collectiveness in SDP is known for evaluation. SDP was firstly proposed to investigate the emergence of collective motion in a system of moving particles. These simple particles are driven with a constant speed, and the directions of their velocities are updated to the average direction of the particles in their neighborhood at each frame. It is shown that the level of random perturbation η on the aligned direction in neighborhood would cause the phase transition of this crowd system from disordered movements into collective motion. The update of velocity direction θ for every individual i in SDP is

$$\theta_i(t+1) = \langle \theta_j(t) \rangle_{j \in \mathcal{N}(i)} + \Delta\theta, \quad (12)$$

where $\langle \theta_j \rangle_{j \in \mathcal{N}(i)}$ denotes the average direction of velocities of particles within the neighborhood of i , $\Delta\theta$ is a random angle chosen with a uniform distribution within the interval $[-\eta\pi, \eta\pi]$. η tunes the noise level of alignment².

6.1 Crowd Collectiveness of SDP

As shown in Fig.3, we compute crowd collectiveness Φ at each time t . Φ monitors the emergence of collective motion over time. At initialization, the spatial locations and velocity directions of all the particles are randomly assigned. The crowd gradually turns into the state of collective motion. The dynamic variation of Φ accurately reflects this phase transition.

As η increases, particles in SDP become disordered. As shown in Fig.4, Φ accurately measures the collectiveness of crowd systems under different levels of random perturbation η . For comparison, Fig.4B plots the average normalized velocity $v = \|\frac{1}{N} \sum_{i=1}^N \frac{\mathbf{v}_i}{\|\mathbf{v}_i\|}\|$, which was commonly used as a measure of collectiveness in existing works [6], [11]. From the large standard deviation of v under multiple simulations with the same η , we see that v is unstable and sensitive to initialization conditions of SDP. On the contrary, Φ shows its robustness for measuring crowd collectiveness.

² In our implementation of SDP, the absolute value of velocity $\|v\|=0.03$, the number of individuals $N = 400$, and interaction radius $r = 1$. Experimental results in [11] have shown that these three parameters only have a marginal effect on the general behaviors of SDP.

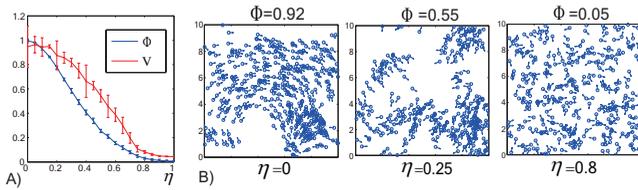


Fig. 4. A) Φ and v with increasing η . The bars indicate the standard deviations of these two measurements. The large deviations of v show that v is unstable and sensitive to initialization of SDP. At each η , simulation repeats for 20 times. B) For a low η , all the individuals are in a global collective motion, and Φ is close to the upper bound. For a relatively larger η , individuals form multiple clusters of collective motions. For a high η , individuals move randomly and Φ is low.

6.2 Collectiveness in Mixed-Crowd Systems

SDP assumes that all the individuals are homogeneous. Studies on complex systems [71] have shown that individuals in most crowd systems in nature are inhomogeneous. To evaluate the robustness of our collectiveness descriptor, we extend SDP to a mixture model by adding outlier particles, which do not have alignment in neighborhood and move randomly all the time. We measure individual collectiveness in this mixed-crowd system. As shown in Fig.5A, individuals are randomly initialized at the start, so the histogram of individual collectiveness has a single mode. When self-driven particles gradually turn into clusters of collective motions, there is a clear separation between two modes in the histogram of individual collectiveness. By removing individuals with collectiveness smaller than 0.5, we can effectively extract collectively moving self-driven particles from outliers as shown in Fig.5B. The threshold is chosen empirically. Notice that although crowd collectiveness is non-negative according to Property 4, the computed individual collectiveness could be negative.

6.3 Collectiveness at l -path Scale

We evaluate the behaviors of collectiveness at l -path scale while SDP are at three different levels of collective motions in Fig.6. We can see that as the SDP gradually turns into collective motion, the regularized $\frac{1}{K^l} \phi_l$ at large l -path scale approaches to 1, which makes $\sum_{l=1}^{\infty} z^l \phi_l$ not converge, and estimated crowd collectiveness Φ unstable when $z = 1/K$. In order to make it converge, we choose $z < 1/K$.

In Fig.7, we plot the regularized crowd collectiveness with SDP at two different levels of collective motions in Fig.7. In Fig.7B when $z = \frac{0.5}{K} = 0.025$, $z^l \Phi_l$ decrease quickly with l . It can be seen that when regularizing collectiveness we consider the crowd collectiveness at higher l -path scale Φ_l has less contributions to the overall regularized Φ . In Fig.7C, we illustrate the asymptotic limit of Φ_l at different l . In Fig.7D, we illustrate the approximate error $\|\mathbf{Z} - \mathbf{Z}_{1 \sim n}\|_2$ with error bound $\frac{(z\sigma(\mathbf{W}))^{n+1}}{1-z\sigma(\mathbf{W})}$ at different n .

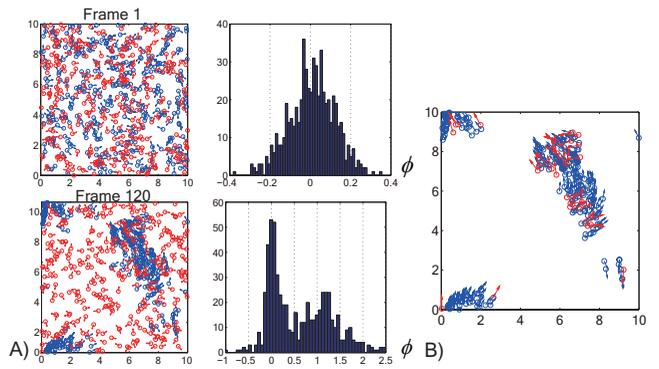


Fig. 5. A) Two frames of the mixture crowd system and their histograms of individual collectiveness. After a while, self-driven particles are organized into clusters of collective motions. The histogram of ϕ_t is clearly separated into two modes. B) By removing particles with individual collectiveness lower than 0.5, we can extract self-driven particles in collective motions. Blue and red points represent self-driven particles and outliers. The number of outliers is equal to that of self-driven particles and $\eta = 0$.

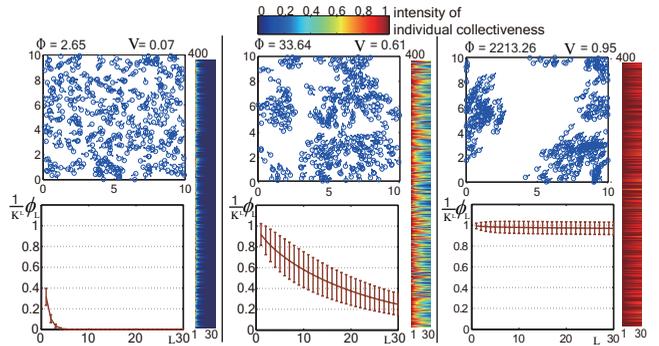


Fig. 6. Regularized individual collectiveness at l -path scale $\frac{1}{K^l} \phi_l$ while SDP are at three different levels of collective motions. In each diagram, the left-hand side shows the average $z^l \phi_l$ with $l = 1 \sim 30$ and the right-hand side shows the visualization of all the values of $z^l \phi_l(i)$ with $l = 1 \sim 30$ and $i = 1 \sim 400$. Since the convergence condition is not satisfied, Φ become unstable when SDP is in a high level of collective motion.

6.4 Convergence Condition of Collectiveness

There are two parameters z and K for computing collectiveness in practical applications: K defines the topological range of neighborhood and z makes the series summation converge. K affects similarity estimation in neighborhood. A large K makes the estimation inaccurate due to the behavior uncertainty between non-neighbors on collective manifold. In Fig.8A, we keep $z = \frac{0.5}{K}$ and let K take increasingly large values, the estimated Φ become unstable. Empirically K could be 5%~10% of $|\mathcal{C}|$. In all our experiments, we fix $K = 20$. Meanwhile, z is constrained by K in Property 1. With different K and z , the upper bound of Φ varies, as shown in Fig.8C. With a larger upper bound, the derivative $\frac{d\Phi}{d\eta}$ is larger and the measurement is

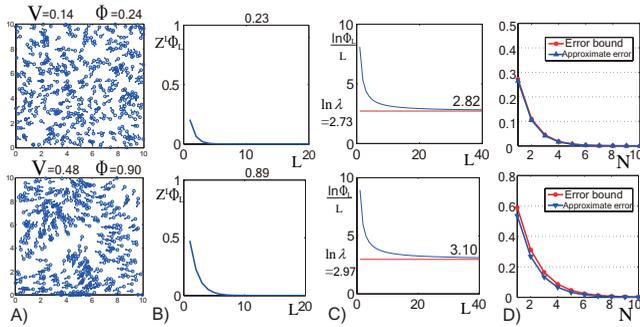


Fig. 7. A) Two frames of SDP under different levels of collective motion. B) Regularized crowd collectiveness at l -path scale $z^l \Phi_l$ with increasing l . The values above are the summation of the crowd collectiveness at top 10 l -path scale. C) The asymptotic limit of Φ_l . λ is the largest eigenvalue of \mathbf{W} computed at current frame of SDP. D) Approximate errors and error bounds of \mathbf{Z} at different n .

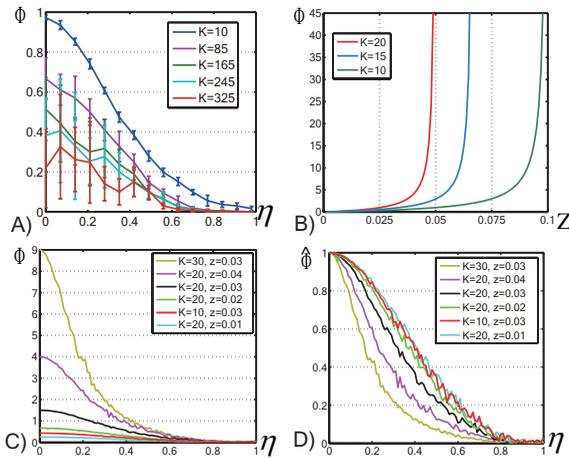


Fig. 8. A) Large K results in the inaccurate estimation of Φ due to the behavior uncertainty between non-neighbors on collective manifold. Given each η , SDP simulation repeats for 20 times. B) Given K fixed, the upper bound of Φ grows quickly when z approaches to $\frac{1}{K}$, which makes Φ unstable. C) Φ with increasing η at different K and z in SDP. D) By dividing $\sup \Phi$, the rescaled $\hat{\Phi}$ have different sensitivity of dynamic order.

more sensitive to the change of crowd motion. Φ can also be re-scaled to $[0, 1]$ by dividing it by the upper bound, the plot of $\hat{\Phi}$ is shown in Fig.8D. Thus by tuning z and K we can control the sensitivity of collectiveness in practical applications. The upper bound of Φ grows quickly when z approaches to $\frac{1}{K}$, which makes the value of Φ unstable, as shown in Fig.8B. The ideal range is $\frac{0.4}{K} < z < \frac{0.8}{K}$.

7 FURTHER EVALUATION AND APPLICATIONS

We evaluate the consistency between our collectiveness and human perception, and apply the proposed descriptor and algorithm on various videos of pedestrian crowds and bacterial colony.

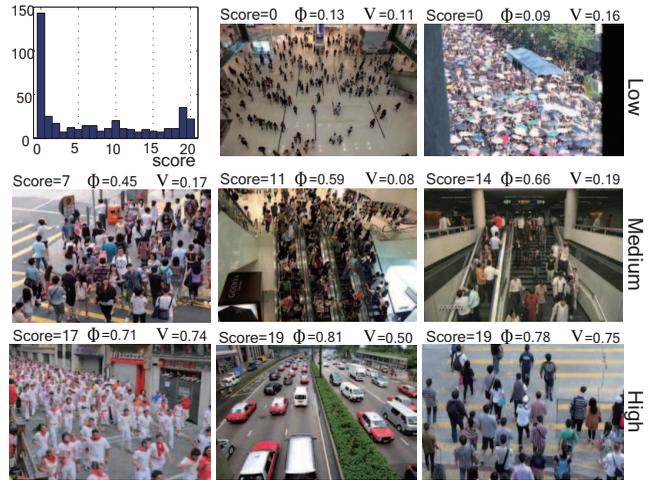


Fig. 9. Histogram of collective scores of all the videos in the Collective Motion Database and some representative video frames, along with their collective scores, Φ , and v . The three rows are from the three collectiveness categories.

7.1 Human Perception for Collective Motion

To quantitatively evaluate the proposed crowd collectiveness, we compare it with human motion perception on a new Collective Motion Database, and then analyze the consistency and correlation with human-labeled ground-truth for collective motions. The Collective Motion Database consists of 413 video clips from 62 crowded scenes. 116 clips are selected from Getty Image [72], 297 clips are collected by us. This database contains different levels of collective motions with 100 frames per clips. Some representative frames are shown in Fig.9. To get the ground-truth, 10 subjects are invited to rate all the videos independently. A subject is asked to rate the level of collective motions in a video from three options: low, medium, and high. Then we propose two criteria to evaluate the consistency between human labeled ground-truth and the proposed collectiveness.

The first is the correlation between the human scores and our collectiveness descriptor. We count the low option as 0, the medium option as 1, and the high option as 2. Since each video is labeled by 10 subjects, we sum up all the scores as the collective score for a video. The range of collective scores is $[0, 20]$. The histogram of collective scores for the whole database is plotted in Fig.9. We compute the crowd collectiveness Φ at each frame using the motion features extracted with a generalized KLT (gKLT) tracker derived from [73], and take the average value of Φ over all the frames as the collectiveness for this video. We compute average v over all the frames using the same motion features as a comparison baseline. Fig.9 shows the collective scores, Φ , and v for some representative videos. Fig.10A scatters the collective scores with Φ and v of all the videos, respectively. There is a high correlation between collective scores and Φ , and the proposed collectiveness is consistent with human perception.

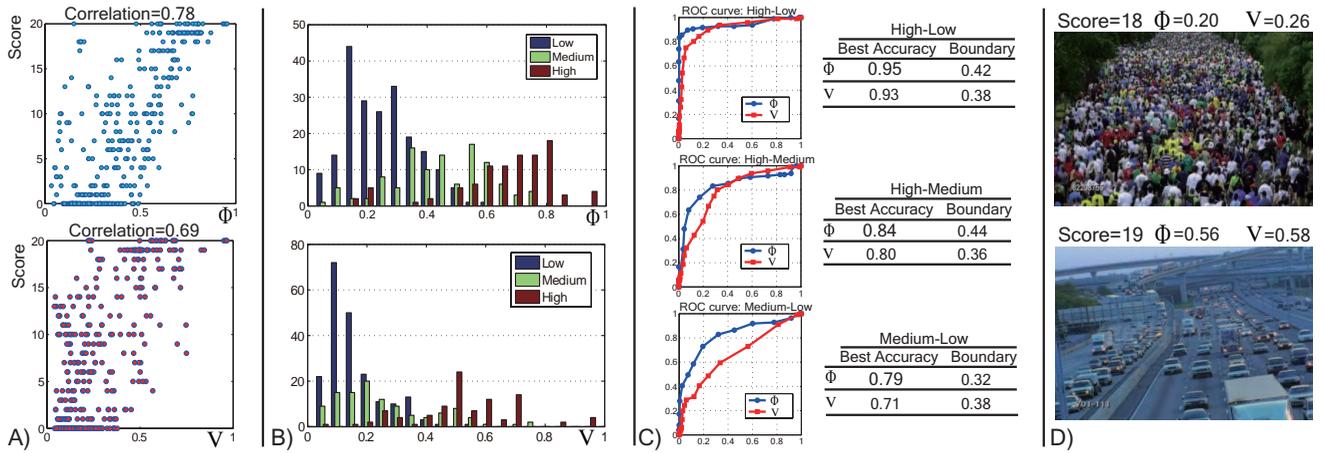


Fig. 10. A) Scatters of collective scores with Φ and v for all the videos. B) Histograms of Φ and v for the three categories of videos. C) ROC curves and best accuracies for high-low, high-medium, and medium-low classification. D) Error examples of collectiveness due to tracking failure or perspective distortion.

The second is the classification accuracy based on the collectiveness descriptor. We divide all the videos into three collectiveness categories by majority voting of subjects' rating, then evaluate how the proposed collectiveness descriptor can classify them. Histograms of Φ and v for the three categories are plotted in Fig.10B. Φ has better discrimination capability than v . Fig.10C plots the ROC curves and the best accuracies which can be achieved with all the possible decision boundaries for binary classification of high and low, high and medium, and medium and low categories based on Φ and v , respectively. Φ can better classify different levels of collective motions than v , especially on the binary classification of high-medium categories and medium-low categories of videos. It indicates our collectiveness descriptor can delicately measure the dynamic state of crowd motions.

Classification failures come from two sources. Since there are overlapping areas between high-medium and medium-low collective motions and it is difficult to quantify human perception of collective motion into three classes, some samples are even difficult for humans to reach consensus and are also difficult to our descriptor. Meanwhile, collectiveness may not be properly computed due to tracking failures, projective distortion, and special scene structures. Two failure examples are shown in Fig.10D. The computed collectiveness in the two videos is low because the KLT tracker does not capture the motions well due to the perspective distortion and the extremely low frame rate, while all 10 subjects give high collective scores because of the regular pedestrian and traffic flows in the scenes.

Since collectiveness in videos is computed based on keypoint tracking on image plane while human perception is based on velocities in the ground plane, here we further evaluate how scene perspective distortion affects the proposed collectiveness. As shown in Fig.11A, we compare the collectiveness of self-driven particles in the ground plane with that after three different perspective projections. When η is small, perspective distortion leads to smaller collectiveness. When η is large, collectiveness

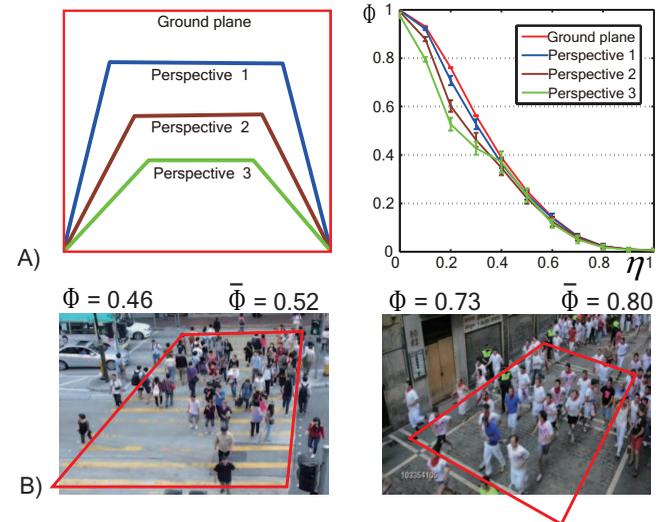


Fig. 11. A) Comparing the collectiveness of SDP with different perspective projections and different η . B) Comparing the collectiveness of real crowd scenes before (ϕ) and after ($\bar{\phi}$) calibration to the ground plane. Red polygons are manually annotated ground planes.

is large and perspective distortion does not make much difference. This is understandable. If two points are very close in space and move in parallel, they still move in similar directions in the image space even with perspective distortion. However, if they are close in space but move in different directions, perspective projection may increase the angles between velocities. In Fig.11B, two real scenes are selected and calibrated to the ground plane. After calibration, the collectiveness slightly increases.

7.2 Collective Motion Detection in Videos

We apply the proposed collectiveness descriptor and collective merging algorithm to analyze collective motions in videos. We first detect collective motions in various videos from the Collective Motion Database. A variety of

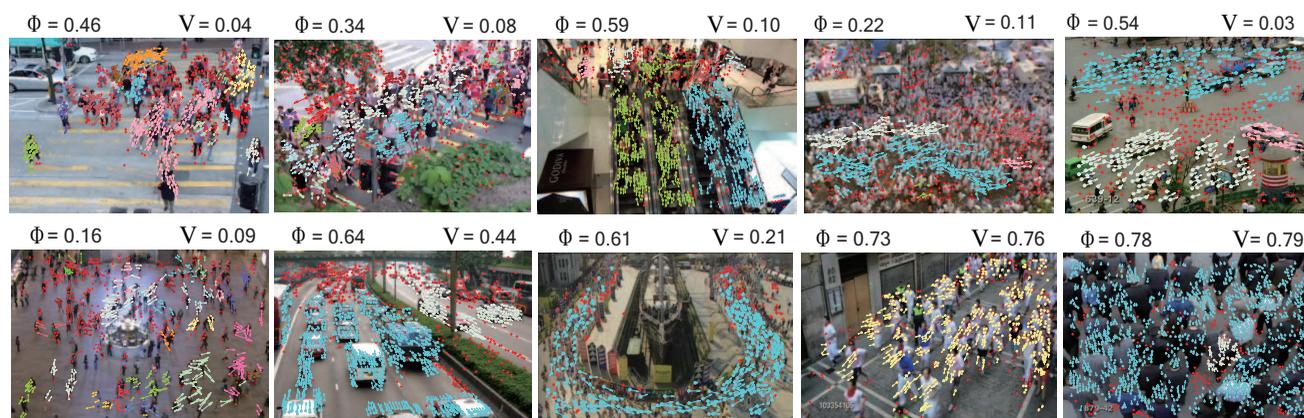


Fig. 12. Detecting collective motions from crowd videos. Keypoints with the same color belong to the same cluster of collective motion. Red crosses are detected outliers.

collective motion patterns are extracted from the crowd movements and their collectiveness is measured.

Collective motion detection in crowd videos is challenging due to the short and fragmented nature of extracted trajectories, as well as the existence of outlier trajectories. Fig.12A shows the detected collective motions by Collective Merging in 10 videos, along with their computed Φ and v . The detected collective motion patterns correspond to a variety of behaviors, such as group walking, lane formation, and different traffic modes, which are of great interest for further video analysis and scene understanding. The estimated crowd collectiveness also varies across scenes and reflects different levels of collective motions in videos. However v cannot accurately reflect the collectiveness of crowd motions in these videos. As the crowds in the videos at first row of Figure 12, v is falsely rather small because the groups of people collectively move in the opposite directions so that the whole average velocity is compromised.

In the algorithm of Collective Merging, α determines the scale of collective motion patterns to be detected. Fig.13 shows the collective motion detection results with threshold $\alpha = 0.5$, $\alpha = 0.6$ and $\alpha = 0.7$ in three scenes. We can see that when the threshold is high, some weak links near the boundary of collective motions are filtered so that the detected collective motion patterns are divided into small fragments. When the threshold is low, the detected collective motion patterns become large and some noise may be included. The setting of α is related to the scale of collective motion patterns to be detected.

7.3 Monitoring Crowd Dynamics in Videos

We use the proposed crowd collectiveness descriptor to monitor the crowd dynamics over time. The real-time computed collectiveness accurately records the evolving states of crowd systems, as shown in Fig.14. In the first scene, the collectiveness changes abruptly when two groups of pedestrians pass with each other. In the second scene where athletics start running, collectiveness reflects the phase transition of the athlete crowd. In the third scene, the collectiveness keeps relatively consistent since the flow of

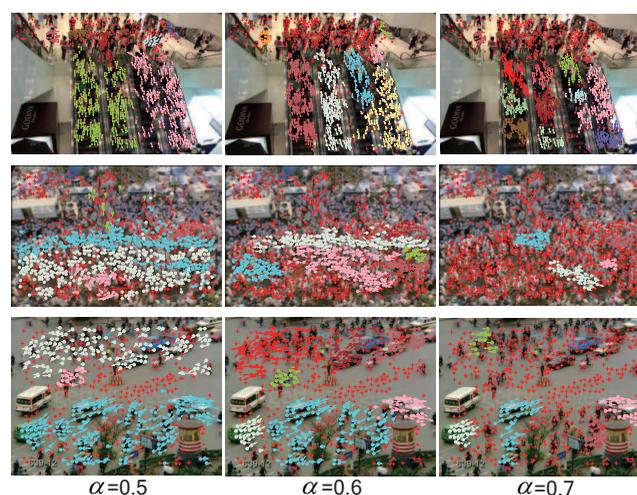


Fig. 13. Collective motion detection results of the Collective Merging algorithm with threshold $\alpha = 0.5$, $\alpha = 0.6$, and $\alpha = 0.7$, respectively.

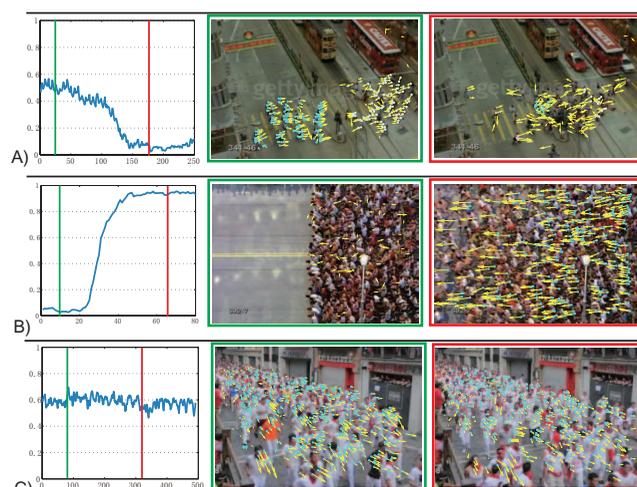


Fig. 14. Monitoring crowd dynamics with collectiveness. Two frames indicate the representative states of the crowd along the time line.



Fig. 15. Collective map of different scenes. High energy areas indicate scene regions with high collectiveness value. These regions have fluent traffic flows.

running people continues as the same. Such events indicate rapid phase transition of a crowd system or some critical point has been reached. They are useful for crowd control and scientific studies.

7.4 Generating Collective Maps of Scenes

We estimate the collective maps of scenes by accumulating the individual collectiveness at each location over time. They reveals valuable information on the interaction between scene structures and crowd behaviors. At each time t we estimate $\phi(i)$ for each keypoints i at spatial location \mathbf{x}_i of one scene. By accumulating the collectiveness values at the spatial locations of keypoints over time, we get the collective map of the scene, which represents the spatial distribution of collectiveness in the scene. It might provide valuable information for crowd management and public facility optimization.

Fig.15 shows the estimated collective maps of six different scenes. Areas with high collectiveness value in these scenes are 1) traffic roads and escalators regularized by scene structures, such as shown in the three collective maps in the first row, and 2) pedestrian lanes formed from self-organization, such as the first collective map in the second row. Empirically we know that regions with high collectiveness values would have fluent traffic thus low probability of accidents. However, in real life crowds of pedestrians are often hard to self-organize into walking lanes with different moving direction. To increase the crowd collectiveness of the scenes and promote fluent flow, physical barriers could be used to divide the crowd into different lanes(such as the trash bin and the barriers between escalators in the last two collective maps in the second row in Fig.15).

7.5 Collective Motions in Bacterial Colony

In this experiment, we use the proposed collectiveness to study collective motions emerging in a bacterial colony. The wild-type *Bacillus subtilis* colony grows on agar substrates, and bacteria inside the colony freely swim on the agar surface. The real motion data of individual bacteria comes from [10]. There are 200 ~ 400 bacteria moving around at every frame.

Fig.16A plot Φ and v with bacteria number over time respectively. Crowd density was proved to be one of the key factors for the formation of collective motion [10],

[11]. A lot of scientific studies are conducted to analyze their correlation. For the same type of bacteria in the same environment, bacteria collectiveness should monotonically increase with density. Fig.16A shows that bacteria density has a much better correlation with Φ than v . In Fig.16B, we scatter Φ and v with the number of bacteria in every frames, respectively. A fractional polynomial is fit to the data of Φ with bacteria numbers. From the polynomial relation we could know the expected collectiveness under some given bacteria number. v does not have some sufficient polynomial relation with bacteria numbers. Fig.16C shows representative frames and the collective motion patterns detected by Collective Merging. Our proposed collectiveness measurement has promising potentials for scientific studies.

8 CONCLUSIONS AND FUTURE WORK

We proposed a collectiveness descriptor for crowd systems as well as their constituent individuals along with the efficient computation. Collective Merging can be used to detect collective motions from randomly moving outliers. We have validated the effectiveness and robustness of the proposed collectiveness on the system of self-driven particles, and shown the high consistency with human perception for collective motion. Further experiments on videos of pedestrian crowds and bacteria colony demonstrate its potential applications in video surveillance and scientific studies.

As a new universal descriptor for various types of crowd systems, the proposed crowd collectiveness should inspire many interesting applications and extensions in the future work. Individuals in a crowd system can move collectively in a single group or in several groups with different collective patterns, even though the system has the same value of Φ . Our single collectiveness measurement can be well extended to a spectrum vector of characterizing collectiveness at different length scales. It is also desirable to enhance the descriptive power of collectiveness by modeling its spatial and temporal variations. The enhanced descriptor can be applied to cross-scene crowd video retrieval, which is difficult previously because universal properties of crowd systems could not be well quantitatively measured. Collectiveness also provides useful information in crowd saliency detection and abnormality detection. This paper is an important starting point in these exciting research directions.

ACKNOWLEDGMENT

This work is partially supported by the General Research Fund sponsored by the Research Grants Council of Hong Kong (Project No. CUHK417110 and CUHK417011) and National Natural Science Foundation of China (Project No.61005057 and 2192019).

REFERENCES

- [1] R. Hughes, "The flow of human crowds," *Annual Review of Fluid Mechanics*, vol. 35, pp. 169–182, 2003.
- [2] M. Moussaid, S. Garnier, G. Theraulaz, and D. Helbing, "Collective information processing and pattern formation in swarms, flocks, and crowds," *Topics in Cognitive Science*, vol. 1, pp. 469–497, 2009.

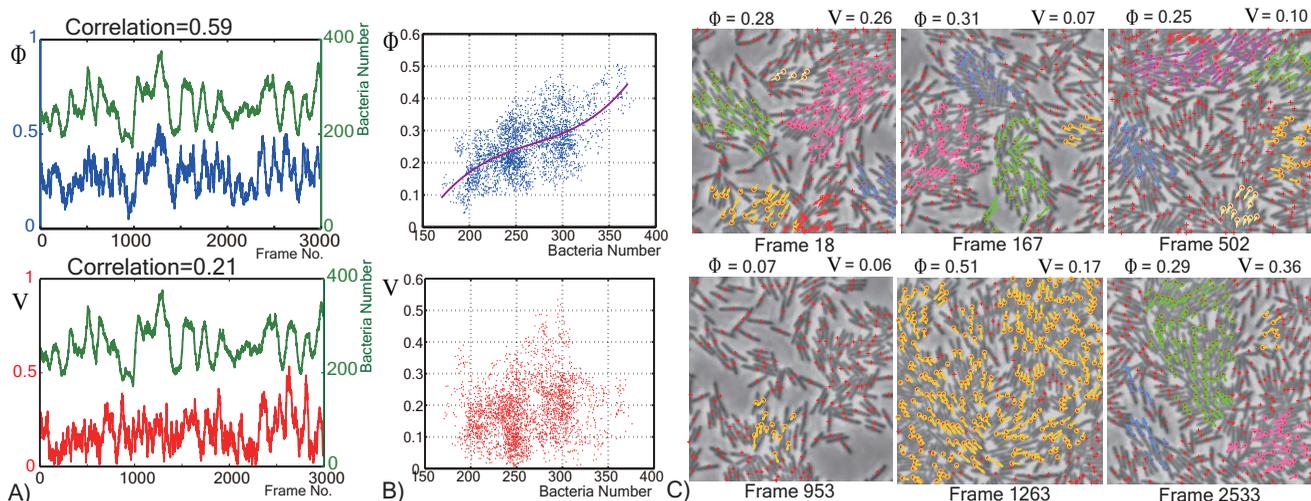


Fig. 16. A) The crowd collectiveness of the bacterial colony and the average normalized velocity change over time. The correlation between Φ and bacteria number ($r = 0.59$) is much higher than the correlation between v and the bacteria number ($r = 0.21$). B) The scatters of Φ and v with bacteria numbers respectively. A fractional polynomial is fit to the data of Φ with bacteria numbers. C) Representative frames of collective motion patterns detected by Collective Merging and their Φ and v . Arrows with different colors indicate different clusters of detected collectively moving bacteria. Red crosses indicate detected randomly moving bacteria.

[3] C. W. Reynolds, "Flocks, herds and schools: A distributed behavioral model," in *ACM SIGGRAPH Computer Graphics*, 1987.

[4] I. Couzin, "Collective cognition in animal groups," *Trends in Cognitive Sciences*, vol. 13, pp. 36–43, 2009.

[5] M. Ballerini, N. Cabibbo, R. Candelier, A. Cavagna, E. Cisbani, I. Giardina, V. Lecomte, A. Orlandi, G. Parisi, A. Procaccini *et al.*, "Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study," *Proceedings of the National Academy of Sciences*, vol. 105, pp. 1232–1237, 2008.

[6] M. Ballerini, N. Cabibbo *et al.*, "Empirical investigation of starling flocks: a benchmark study in collective animal behaviour," *Animal behaviour*, vol. 76, pp. 201–215, 2008.

[7] D. Lin, E. Grimson, and J. Fisher, "Learning visual flows: A lie algebraic approach," in *Proc. CVPR*, 2009.

[8] B. Zhou, X. Wang, and X. Tang, "Understanding collective crowd behaviors: Learning a mixture model of dynamic pedestrian-agents," in *Proc. CVPR*, 2012.

[9] J. Buhl, D. Sumpter, I. Couzin, J. Hale, E. Despland, E. Miller, and S. Simpson, "From disorder to order in marching locusts," *Science*, vol. 312, pp. 1402–1406, 2006.

[10] H. Zhang, A. Ber, E. Florin, and H. Swinney, "Collective motion and density fluctuations in bacterial colonies," *Proceedings of the National Academy of Sciences*, vol. 107, pp. 13 626–13 630, 2010.

[11] 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*, vol. 75, pp. 1226–1229, 1995.

[12] B. Zhou, X. Tang, and X. Wang, "Measuring crowd collectiveness," in *Proc. CVPR*, 2013.

[13] J. Toner, Y. Tu, and S. Ramaswamy, "Hydrodynamics and phases of flocks," *Annals of Physics*, vol. 318, pp. 170–244, 2005.

[14] J. Parrish and L. Edelstein-Keshet, "Complexity, pattern, and evolutionary trade-offs in animal aggregation," *Science*, vol. 284, pp. 99–101, 1999.

[15] D. Forsyth, *Group dynamics*. Wadsworth Pub Co, 2009.

[16] G. Le Bon, "The crowd: A study of the popular mind," *The Macmillan Co. New York*, 1897.

[17] R. M. Raafat, N. Chater, and C. Frith, "Herding in humans," *Trends in Cognitive Sciences*, vol. 13, pp. 420–428, 2009.

[18] Z. A. Vicsek T, "Collective motion," *arXiv:1010.5017*, 2010.

[19] J. Buhl, D. Sumpter, I. Couzin, J. Hale, E. Despland, E. Miller, and S. Simpson, "From disorder to order in marching locusts," *Science*, 2006.

[20] N. Makris, P. Ratilal, S. Jagannathan, Z. Gong, M. Andrews, I. Bertatos, O. Godø, R. Nero, and J. Jech, "Critical population density triggers rapid formation of vast oceanic fish shoals," *Science*, vol. 323, no. 5922, pp. 1734–1737, 2009.

[21] S. Camazine, *Self-organization in biological systems*. Princeton Univ Pr, 2003.

[22] I. Couzin and J. Krause, "Self-organization and collective behavior in vertebrates," *Advances in the Study of Behavior*, vol. 52, pp. 1–75, 2003.

[23] J. Toner and Y. Tu, "Flocks, herds, and schools: A quantitative theory of flocking," *Physical Review E*, vol. 58, pp. 4828–4858, 1998.

[24] W. Bialek, A. Cavagna, I. Giardina, T. Mora, E. Silvestri, M. Viale, and A. Walczak, "Statistical mechanics for natural flocks of birds," *Arxiv preprint arXiv:1107.0604*, 2011.

[25] D. Helbing and P. Molnar, "Social force model for pedestrian dynamics," *Physical review E*, vol. 51, no. 5, pp. 4282–4286, 1995.

[26] H. Chate and F. Ginelli, "Collective motion of self-propelled particles interacting without chesion," *Physics Review*, vol. 77, 1995.

[27] M. Girvan and M. E. Newman, "Community structure in social and biological networks," *Proceedings of the National Academy of Sciences*, vol. 99, pp. 7821–7826, 2002.

[28] G. Palla, A.-L. Barabási, and T. Vicsek, "Quantifying social group evolution," *Nature*, vol. 446, pp. 664–667, 2007.

[29] R. Olfati-Saber, "Flocking for multi-agent dynamic systems: Algorithms and theory," *Automatic Control, IEEE Transactions on*, 2006.

[30] L. Petitjean, M. Reffay, E. Grasland-Mongrain, M. Poujade, B. Ladoux, A. Buguin, and P. Silberzan, "Velocity fields in a collectively migrating epithelium," *Biophysical Journal*, vol. 98, pp. 1790–1800, 2010.

[31] S. Ali and M. Shah, "A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis," in *Proc. CVPR*, 2007.

[32] R. Mehran, B. Moore, and M. Shah, "A streakline representation of flow in crowded scenes," in *Proc. ECCV*, 2010.

[33] X. Wang, X. Ma, and W. Grimson, "Unsupervised activity perception in crowded and complicated scenes using hierarchical bayesian models," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 539–555, 2009.

[34] T. Hospedales, S. Gong, and T. Xiang, "A markov clustering topic model for mining behaviour in video," in *Proc. ICCV*, 2009.

[35] R. Emonet, J. Varadarajan, and J. Odobez, "Extracting and locating temporal motifs in video scenes using a hierarchical non parametric bayesian model," in *Proc. CVPR*, 2011.

[36] X. Wang, K. Ma, G. Ng, and W. Grimson, "Trajectory analysis and semantic region modeling using a nonparametric bayesian model," in *Proc. CVPR*, 2008.

[37] —, "Trajectory analysis and semantic region modeling using non-

parametric hierarchical bayesian models," *Int'l Journal of Computer Vision*, vol. 95, pp. 287–312, 2011.

[38] B. T. Morris and M. M. Trivedi, "Trajectory learning for activity understanding: Unsupervised, multilevel, and long-term adaptive approach," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 33, pp. 2287–2301, 2011.

[39] B. Zhou, X. Wang, and X. Tang, "Random field topic model for semantic region analysis in crowded scenes from tracklets," in *Proc. CVPR*, 2011.

[40] V. Saligrama and Z. Chen, "Video anomaly detection based on local statistical aggregates," in *Proc. CVPR*, 2012.

[41] S. Wu, B. E. Moore, and M. Shah, "Chaotic invariants of lagrangian particle trajectories for anomaly detection in crowded scenes," in *Proc. CVPR*, 2010.

[42] M. Rodriguez, S. Ali, and T. Kanade, "Tracking in unstructured crowded scenes," in *Proc. ICCV*, 2009.

[43] L. Kratz and K. Nishino, "Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models," in *Proc. CVPR*, 2009.

[44] V. Rabaud and S. Belongie, "Counting crowded moving objects," in *Proc. CVPR*, 2006.

[45] G. Brostow and R. Cipolla, "Unsupervised bayesian detection of independent motion in crowds," in *Proc. CVPR*, 2006.

[46] B. Zhou, X. Tang, and X. Wang, "Detecting coherent motions from crowd clutters," in *Proc. ECCV*, 2012.

[47] T. Brox and J. Malik, "Object segmentation by long term analysis of point trajectories," in *Proc. ECCV*, 2010.

[48] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, "Anomaly detection in crowded scenes," in *Proc. CVPR*, 2010.

[49] A. B. Chan and N. Vasconcelos, "Modeling, clustering, and segmenting video with mixtures of dynamic textures," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 30, pp. 909–926, 2008.

[50] T. Lan, Y. Wang, W. Yang, S. N. Robinovitch, and G. Mori, "Discriminative latent models for recognizing contextual group activities," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 34, pp. 1549–1562, 2011.

[51] T. Lan, L. Sigal, and G. Mori, "Social roles in hierarchical models for human activity recognition," in *Proc. CVPR*, 2012.

[52] W. Ge, R. Collins, and R. Ruback, "Vision-based analysis of small groups in pedestrian crowds," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 34, pp. 1003–1016, 2011.

[53] P. Scovanner and M. Tappen, "Learning pedestrian dynamics from the real world," in *Proc. ICCV*, 2009.

[54] L. Kratz and K. Nishino, "Going with the flow: Pedestrian efficiency in crowded scenes," in *Proc. ECCV*, 2012.

[55] R. Mehran, A. Oyama, and M. Shah, "Abnormal crowd behavior detection using social force model," in *Proc. CVPR*, 2009.

[56] P. Scovanner and M. F. Tappen, "Learning pedestrian dynamics from the real world," in *Proc. ICCV*, 2009.

[57] S. Pellegrini, A. Ess, K. Schindler, and L. Van Gool, "You'll never walk alone: Modeling social behavior for multi-target tracking," in *Proc. ICCV*, 2009.

[58] D. Lin, E. Grimson, and J. Fisher, "Modeling and estimating persistent motion with geometric flows," in *Proc. CVPR*, 2010.

[59] M. Chang, N. Krahnstoeber, and W. Ge, "Probabilistic group-level motion analysis and scenario recognition," in *Proc. ICCV*, 2011.

[60] M. Moussaid, N. Perozo, S. Garnier, D. Helbing, and G. Theraulaz, "The walking behaviour of pedestrian social groups and its impact on crowd dynamics," *PLoS One*, vol. 5, p. e10047, 2010.

[61] N. Pelechano, J. M. Allbeck, and N. I. Badler, "Virtual crowds: Methods, simulation, and control," *Synthesis Lectures on Computer Graphics and Animation*, 2008.

[62] D. Helbing, I. Farkas, and T. Vicsek, "Simulating dynamical features of escape panic," *Nature*, vol. 407, pp. 487–490, 2000.

[63] J. Van Den Berg, S. J. Guy, M. Lin, and D. Manocha, "Reciprocal n-body collision avoidance," *International Symposium on Robotics Research*, 2011.

[64] R. Narain, A. Golas, S. Curtis, and M. Lin, "Aggregate dynamics for dense crowd simulation," in *ACM Transactions on Graphics*, 2009.

[65] A. Lerner, Y. Chrysanthou, and D. Lischinski, "Crowds by example," in *Computer Graphics Forum*, 2007.

[66] S. J. Guy, J. van den Berg, W. Liu, R. Lau, M. C. Lin, and D. Manocha, "A statistical similarity measure for aggregate crowd dynamics," *ACM Transactions on Graphics (TOG)*, 2012.

[67] N. Biggs, *Algebraic graph theory*. Cambridge Univ. Pr., 1993.

[68] D. E. Knuth, *The Art of Computer Programming, Volume 1 Fundamental Algorithms (Third Edition)*. Addison-Wesley, 1997.

[69] J.-C. Delvenne and A.-S. Libert, "Centrality measures and thermodynamic formalism for complex networks," *Physical Review E*, vol. 83, p. 046117, 2011.

[70] D. Coppersmith and S. Winograd, "Matrix multiplication via arithmetic progressions," *Journal of symbolic computation*, 1990.

[71] J. Miller and S. Page, *Complex adaptive systems: An introduction to computational models of social life*. Princeton Univ Pr, 2007.

[72] GettyImage, "http://www.gettyimages.com/."

[73] C. Tomasi and T. Kanade, "Detection and Tracking of Point Features," in *Int'l Journal of Computer Vision*, 1991.



Bolei Zhou (S'11) received M.Phil. degree from the Chinese University of Hong Kong in Information Engineering in 2012 and B.Eng. degree from Shanghai Jiao Tong University in Biomedical Engineering in 2010. He is currently a PhD student in the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology. His research interests include computer vision and pattern recognition.



Xiaou Tang (S'93-M'96-SM'02-F'09) received the B.S. degree from the University of Science and Technology of China, Hefei, in 1990, and the M.S. degree from the University of Rochester in 1991. He received the Ph.D. degree from the Massachusetts Institute of Technology in 1996. He is currently a Professor in the Department of Information Engineering and Associate Dean of the Faculty of Engineering of the Chinese University of Hong Kong. He was a program chair of the IEEE International Conference on Computer Vision (ICCV) 2009 and an Associate Editor of IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) and International Journal of Computer Vision (IJCV). He is a Fellow of IEEE.



Hepeng Zhang received the B.S. degree from Fudan University in 2000, and the Ph.D. degree from the City University of New York in 2004. He is a Distinguished Research Fellow in the Institute of Natural Sciences and Department of Physics in Shanghai Jiao Tong University. He is interested in mechanics and statistics of locomotion.



Xiaogang Wang (S'03-M'10) received the B.S. degree from the Special Class for Gifted Young at University of Science and Technology of China in Electrical Engineering and Information Science in 2001, and the M.Phil. degree from Chinese University of Hong Kong in 2004. He received the PhD degree in Computer Science from the Massachusetts Institute of Technology. He is currently an assistant professor in the Department of Electronic Engineering at the Chinese University of Hong Kong. He was the area chair of IEEE International Conference on Computer Vision (ICCV) 2011. He is the associate editor of the Image and Visual Computing Journal. His research interests include computer vision and machine learning.