Measuring Collectiveness via Refined Topological Similarity

XUELONG LI, Center for Optical Imagery Analysis and Learning (OPTIMAL), State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences
MULIN CHEN and QI WANG, School of Computer Science and Center for Optical Imagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University

Crowd system has motivated a surge of interests in many areas of multimedia, as it contains plenty of information about crowd scenes. In crowd systems, individuals tend to exhibit collective behaviors, and the motion of all those individuals is called collective motion. As a comprehensive descriptor of collective motion, *collectiveness* has been proposed to reflect the degree of individuals moving as an entirety. Nevertheless, existing works mostly have limitations to correctly find the individuals of a crowd system and precisely capture the various relationships between individuals, both of which are essential to measure collectiveness. In this article, we propose a collectiveness-measuring method that is capable of quantifying collectiveness accurately. Our main contributions are threefold: (1) we compute relatively accurate collectiveness by making the tracked feature points represent the individuals more precisely with a point selection strategy; (2) we jointly investigate the spatial-temporal information of individuals and utilize it to characterize the topological relationship between individuals by manifold learning; (3) we propose a *stability* descriptor to deal with the irregular individuals, which influence the calculation of collectiveness. Intensive experiments on the simulated and real world datasets demonstrate that the proposed method is able to compute relatively accurate collectiveness and keep high consistency with human perception.

CCS Concepts: • **Computing methodologies → Computer vision tasks**

Additional Key Words and Phrases: Multimedia, crowd analysis, collectiveness, manifold, feature extraction

ACM Reference Format:
Xuelong Li, Mulin Chen, and Qi Wang. 2016. Measuring collectiveness via refined topological similarity. ACM Trans. Multimedia Comput. Commun. Appl. 12, 2, Article 34 (March 2016), 22 pages.
DOI: http://dx.doi.org/10.1145/2854000

1. INTRODUCTION

In crowd systems, collective motion has been a hot topic that attracts many researchers. It always conveys a large quantity of information about the crowd phenomenon and exists not only in the human community, but also in the natural world as shown in Figure 1. Typically, the information provided by one individual can only reveal some local messages of the scene. But when collective motion is formed, individuals are treated

This work was supported by the National Natural Science Foundation of China under Grants 61105012 and 61379094; by the Natural Science Foundation Research Project of Shaanxi Province under Grant 2015JM6264; by the Fundamental Research Funds for the Central Universities under Grants 3102014JC02020G07 and 3102015BJ(II)JJZ01; and by the Open Research Fund of Key Laboratory of Spectral Imaging Technology, Chinese Academy of Sciences.
Authors' addresses: X. Li, Center for Optical Imagery Analysis and Learning (OPTIMAL), State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, Shaanxi, P. R. China; email: xuelong_li@opt.ac.cn; M. Chen and Q. Wang (corresponding author), School of Computer Science and Center for Optical Imagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an 710072, Shaanxi, P. R. China; emails: chenmulin@mail.nwpu.edu.cn, crabwq@nwpu.edu.cn.

1551-6857/2016/03-ART34 $15.00
DOI: http://dx.doi.org/10.1145/2854000

Fig. 1. Collective motion in fish shoal, bison herd, soldiers, and bacterial colony. URL of the images: https://www.google.com.hk/imghp?gws_rd=cr,ssl.

as part of a union and they share similar properties that are significantly meaningful for studies in many disciplines. For example, behaviorists think simple redundant interactions between individuals lead to various crowd behaviors [Moussaid et al. 2009; Reynolds 1987]. Physicists treat the individuals as particles and characterize their interactions with equations from hydromechanics [Kružkov 1970]. Zoologists consider the collective motion of animals as an evolutionary advantage to go through the difficult living environment [Couzin 2009].

In this work, our focus is mainly on the multimedia aspect. Toward this direction, the study of collective motion is actually with many branches, such as collective behavior analysis, crowd simulation, and crowded scene understanding. Nevertheless, only a few works [Shao et al. 2014a; Zhou et al. 2014] have been done to exploit the general properties of collective motions, which is necessary to characterize different crowd systems. After obtaining the properties of collective motions, we can describe the crowd videos by text, which may inspire some improvements on multimedia applications such as video retrieval and video summary. In order to quantify the properties of collective motions, Zhou et al. [2014] has proposed a fundamental and comprehensive descriptor called *collectiveness*. Collectiveness indicates the degree of individuals performing as a team and reflects how much individuals are united into a uniform group. *Individual-level collectiveness* describes an individuals' relationship with the others, while *entirety-level collectiveness* measures the behavior consistency of all the individuals in a crowd scene. Though collectiveness can properly describe the characteristic of the crowd, the quantitative calculation is not an easy task.

A major difficulty in the measuring of collectiveness is the extraction of individuals. For the investigation of crowded scenes, it is essential to extract the individuals precisely. However, object identification in crowded scenes is still an unresolved issue because occlusion is heavy and the number of individuals is large. Some works detect feature points in crowded scenes and treat them as the individuals of a crowd system, but those feature points cannot always represent individuals correctly, not to mention all the feature points come from the individuals. What is more, occlusion makes it difficult to track the feature points, which limits the dynamic measurement of the crowd system. Since the extraction of individuals is still a problem, it is not easy to quantify the characteristics of crowded scenes.

Another barrier to measure collectiveness comes from the complicated spatial structure of crowd systems. Individuals in a collective motion tend to form a complicated spatial structure, called collective manifold. In collective manifolds, individuals keep high coherence only with their local neighbors, and the behaviors may have big difference between individuals that are far from each other. Even so, all the individuals in a collective manifold may keep a close relationship by the propagation of information along paths. This phenomenon represents a serious impediment in the measuring of individuals' relationship, then hampering the calculation of collectiveness. As Ballerini et al. [2008] pointed out, interactions among individuals depend on their topological structure rather than metric distance, so one way to solve this problem is to exploit the topological relationship of individuals. However, only a few works emphasized the topological structure.

In addition, some ambiguous conceptions about collectiveness hamper its calculation. In crowd scenes, there are always many moving objects besides pedestrians, such as cars. Should these objects also be considered as individuals of crowd scenes? Besides, what is the exact definition of individual-level collectiveness? Zhou et al. [2014] measured the collectiveness of each individual by comparing their behaviors on the current frame, leading to a problem that the unsteady individuals are ignored. Unsteady individuals are those whose behaviors change irregularly, and they seem not to belong to any uniform group. So their collectiveness should be low according to the definition of collectiveness. However, an unsteady individual may keep high behavior consistency with other individuals at some moments, so their collectiveness could be overvalued if they are not detected. To reduce these problems, we redefine those ambiguous conceptions. We treat every moving object (including cars) as an individual in crowd scene, and describe individual-level collectiveness of an individual as its behavior consistency with all the other steady individuals. Moreover, we detect the unsteady individuals and set their collectiveness to 0, then remove them to eliminate their influence on other individuals.

Our goal in this work is to develop a robust method, which is capable of dealing with complex real-world crowd systems, to measure collectiveness more accurately. Similar to the Measuring Crowd Collectiveness (MCC) [Zhou et al. 2014], the proposed method measures collectiveness by investigating the relationship between individuals of a crowd system. *First*, individuals are identified. Feature points on video frames are detected and tracked to represent the initially detected individuals. Then, the most representative points are extracted and selected as the refined individuals. *Secondly*, temporal information that captures the relationship between neighboring frames is combined with spatial information to quantify the similarities between the obtained individuals. *Thirdly*, according to the fact that individuals with unsteady velocities might disturb the measurement of other individuals' collectiveness, we introduce a *stability* descriptor to detect unsteady individuals and discard them. *Finally*, we study the relationship between individuals by manifold learning to compute the collectiveness from individual-level and entirety-level aspects. A flow diagram of the proposed method is shown in Figure 2.

We summarize our contributions as follows:

—A point selection method based on segmentation is proposed to get a more accurate individual representation. Existing works mostly use detect feature points on crowd videos and directly treat them as individuals of the crowded scenes. However, there are inevitably repetitive points on the same individual, which leads to redundancy and inaccuracy. The point selection method is capable of selecting the most representative points and obtaining individuals more precisely.

—Multiclues similarity is exploited and further used to capture the topological relationship between individuals. Spatial-temporal information and manifold learning are jointly incorporated to quantify the relationship between individuals from the view of topology. By exploiting the topological property, our method is more capable of handling the crowd systems with various structures, and calculating collectiveness accurately.

—An individual-level descriptor, stability, is proposed to quantify the degree of an individual keeping steady in a crowd system. Unsteady individuals will affect the calculation of collectiveness in a crowd system. We measure the stability of individuals and restrain them to improve the collectiveness calculation.

This article is organized as follows. Section 2 introduces the related works on the crowd systems. Section 3 describes the point selection method, which aims to accurately

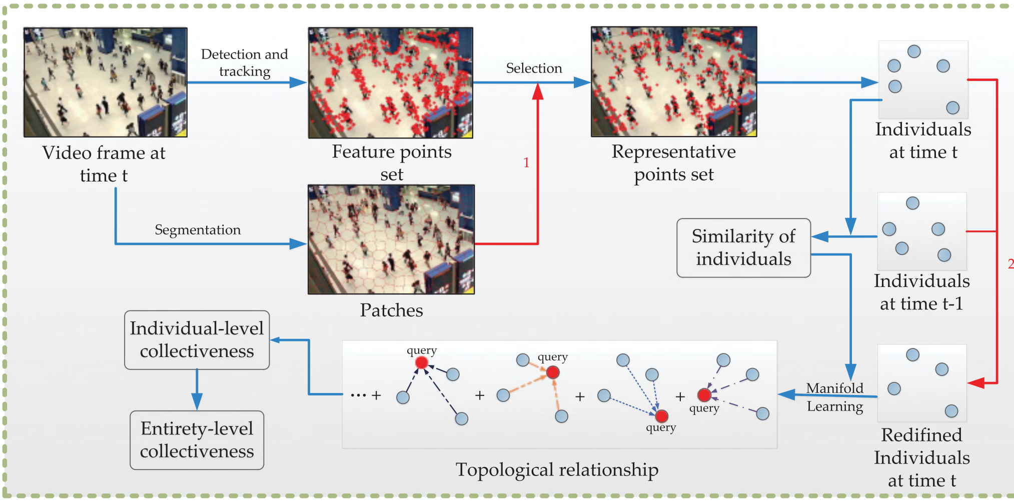


Fig. 2. The flowchart of the proposed method. First, we use feature extraction to generate the original point set (red) and employ segmentation to get a redefined refined point set that can represent individuals more accurately. Second, temporal and spatial information are combined to investigate the relationship between individuals. Third, we measure the stability of every individual and remove the unsteady individuals to better characterize the crowd system. Fourth, topological relationships between individuals are explored, based on which the individual collectiveness and entirety-level collectiveness are calculated. The image is selected from Collective Motion dataset [Zhou et al. 2014].

extract individuals from videos of crowded scenes. Section 4 shows how to measure the collectiveness by exploring the topological relationship of the obtained individuals. We present our extensive experiments in Section 5, and discuss some issues related to the proposed method in Section 6. The conclusion and future work come in Section 7.

2. RELATED WORK

In the area of multimedia, a significant amount of research effort has been made on crowd analysis because crowd systems always convey abundant information and have a wide range of applications. Li et al. [2015] made a brief review on this topic recently.

Among numerous efforts toward this topic, some efforts have been spent on learning models of crowded scene structures and motion patterns [Saleemi et al. 2010; Yang et al. 2009; Jodoin et al. 2013; Hu et al. 2006; Benabbas et al. 2011]. Kratz and Nishino [2012] represented crowd motion with a collection of Hidden Markov Model (HMM) [Kratz and Nishino 2009] trained on local space-time motion patterns. Ali and Shah [2008] employed a scene structure based force model to track individual in a crowded scene. Individuals' movements are converted from the long-term forces into local ones. They also used a Lagrangian particle [Ali and Shah 2007] dynamics to model crowd flows by treating a moving crowd as an aperiodic dynamical system. Wang et al. [2009] investigated motion pattern in crowded scenes with an unsupervised learning framework that is constructed by three hierarchical Bayesian models. Zhou et al. [2011] integrated the Latent Dirichlet Allocation topic model and Markov random fields to learn crowd motion. Another work of theirs [Zhou et al. 2012] studied collective behavior patterns with a mixture model, which models the whole crowd as a mixture of dynamic pedestrian agents. Wang et al. [2008, 2011] learned motion patterns with nonparametric Bayesian models. Zhou et al. [2012] utilized underlying motion priors to detect coherent motion patterns. Ge et al. [2012] discovered collective motions by bottom-up hierarchial clustering with a generalized Hausdorff distance.

In addition, behavior analysis in crowded scenes has also attracted much attention. Huang et al. [2009], Shao et al. [2014b], and Zhao et al. [2015] have introduced the social force model to estimate interaction forces among individuals and detect abnormal behaviors in crowded scenes. Lan et al. [2012b] recognized group activities by jointly capturing the individual actions, the group activity, and interactions among them. Helbing and Molnar [1995] proposed a social force model to measure internal motivations of individuals. Three force terms are used to describe individuals' movement. The first one indicates acceleration along the direction of desired velocity. The second one reflects the phenomenon that pedestrians get used to keeping a distance from others. The third one models attractive effects among individuals. Lan et al. [2012a] described human behavior from low-level actions to high-level events by a hierarchical based max-margin framework and Chang et al. [2011] built a probabilistic grouping strategy to analyze individuals' movements and softly assign them into groups. Zhao et al. [2014] recognized individuals' gestures by structured streaming skeleton. Scovanner and Tappen [2009] learned the parameters of pedestrians' trajectories by variational model learning. Pellegrini et al. [2009] introduced a dynamic social behavior model to track multi-individuals from a vehicle-mounted camera.

Crowd simulation in virtual world has also been a hot research topic recently [Almeida et al. 2013; Golas et al. 2013; Best et al. 2014; Zheng et al. 2014; Zhang et al. 2015; Zawidzki et al. 2014; Zhou et al. 2007; Heck et al. 2007]. Reynolds [1987] proposed a distributed behavioral model to script the path of each individual. This kind of agent-based model can also be found in Helbing et al. [2000] and van den Berg et al. [2009]. Narain et al. [2009] introduced a flow-based model to simulate large, dense crowds on desktop computers. Lerner et al. [2007] introduced an example-based crowd simulation technique, which can copy trajectories taken by individuals from real-world crowd systems. Guy et al. [2012] presented an information-theoretic approach to quantify the ability of a simulator to reproduce the behaviors of crowd systems.

To quantitatively measure the crowd behaviors and compare them across different crowded scenes, some approaches have been proposed to learn collectiveness. Zhou et al. [2014] considered crowd collectiveness as a bottom feature of crowd systems and measured it by a manifold learning method. They use a generalized KLT tracker to find individuals and calculate similarity of individuals on each frame. Collectiveness is then obtained based on the manifold topological relationship of individuals. A group level method to measure collectiveness was proposed by Shao et al. [2014a], which is also based on tracking. These two methods share the same problem that the number of feature points is significantly different from the number of individuals, which leads to an inaccurate collectiveness. There are always more feature points on the big object while fewer on the small one, which makes the number of feature points not capable of reflecting the real number of individuals.

3. POINT SELECTION

The preliminary for crowd analysis is to extract individuals from the input videos. Existing works [Shao et al. 2014a; Zhou et al. 2014; Jodoin et al. 2013; Wang et al. 2013a, 2013b] tackle this problem by object detection and tracking, or regard the specific feature points as individuals directly. Unfortunately, object detection and person identification in crowded scene are still difficult and tracking points do not always represent the individuals accurately. Since collectiveness is analyzed according to the relationship between individuals, an accurate calculation cannot be acquired if we cannot correctly find the individuals in a frame. For this purpose, we propose to choose a set of representative points as individuals instead of directly treating all the feature points as individuals. First, an initial point set is obtained by feature detection and tracking. Then a point selection operation is followed to redefine the points.

3.1. Detection and Tracking of Initial Feature Points

In order to obtain the detailed information of a crowd system, it is essential to detect the individuals. At the same time, it is necessary to track the individuals to get the time-series association. As a consequence, both detection and tracking procedures could significantly influence the investigation of crowd systems. To tackle this problem, we employ a Generalized KLT (gKLT) tracker [Zhou et al. 2014] derived from KLT [Tomasi and Kanade 1991], which jointly combines detection and tracking with efficient computation.

gKLT first detects feature points in the moving foreground, which contains sufficient texture information. Then the feature points are tracked, and their velocities are measured frame by frame according to their displacements. Thus, a set of tracked feature points is obtained.

3.2. Consistency-Based Point Selection

With the preceding processing, we can get a set of tracked feature points. However, these tracking points obtained by KLT cannot represent the individuals very accurately. For the same part of one moving individual, there might be several points within it. Since not all the points' movements are precisely estimated, if we take the whole of them into consideration, it may lead to inaccurate crowd estimation. Even if the motion information is quite right, the redundant representation for the single part is useless. Motivated by this defect, we plan to refine the initially detected and tracked points and search for those stable and representative ones.

For the previously mentioned purpose, we examine the color and spatial location of the points to determine whether they come from the identical individual. To be specific, a segmentation method is employed to fuse these informative clues together. Among various segmentation methods, we utilize the SLIC algorithm [Achanta et al. 2012], which can generate compact and nearly uniform patches. It has good performance at a low computational cost and better to boundaries than other segmentation algorithms.

Generally, points in the same patch always have a high consistency and a better chance to come from the same object. Therefore, with segmented patches, we further check the velocity of the feature points in the same patch. Suppose the velocity correlation between two points at time t is

$$C_t(i, j) = max\left(\frac{\boldsymbol{v}_i \cdot \boldsymbol{v}_j}{|\boldsymbol{v}_i| \cdot |\boldsymbol{v}_j|}, 0\right), \tag{1}$$

where i and j are two tracked feature points, and v_i and v_j are their velocities. The *max* function is used because we do not want the result to be negative. If points i and j are in the same patch and $C_t(i, j) > \eta$ (η is a threshold), they are supposed to belong to an identical object. In this case, we retain the one that keeps higher velocity correlation with all the other points in the patch. By doing this iteratively, some useless tracking points can be abandoned. The remaining ones represent individuals more precisely, as Figure 3 shows. An overview of the point selection procedure is shown in Algorithm 1.

4. MEASURING COLLECTIVENESS

After the point selection procedure, representative feature points are directly considered to be individuals of a crowd system. In this section, the relationship between those individuals is investigated to measure individual-level and entirety-level collectiveness. First, similarities of individuals are calculated by jointly combining spatial and temporal information. Second, a new individual-level descriptor, stability, is proposed to detect unsteady individuals that may affect the measuring of entirety collectiveness. At last, we learn the topological relationship of individuals by utilizing their similarities and the collectiveness of an individual is measured on the basis of their topological

Fig. 3. (a) Generated feature points (yellow). (b) Segmented patches. (c) Refined feature points by our point selection method. We can see that some useless feature points are removed in (c) compared with (a), remaining the points that can represent the individuals more accurately. The image is selected from the collective motion dataset [Zhou et al. 2014].

ALGORITHM 1: Point Selection

Input: Video Frames $I = \{I_t\}_{t=1}^{N}$.
Output: Redefined tracking points set $\Omega = \{\Omega_t\}_{t=1}^{N}$, where Ω_t contains the tracked feature points on frame I_t.
for *each frame* I_t **do**
    Get patch image P_t by segmentation.
end
for *each patch* P_{ti} *in* P_t **do**
    Find the points with high velocity consistency in P_{ti} by Equation (1);
    Select the most representative points.
end

relationships. Based on the collectiveness of all individuals, the entirety-level collectiveness can be calculated.

4.1. Similarity of Individuals

For the purpose of measuring collectiveness of a crowd system, it is necessary to find the relationship between individuals. First of all, spatial similarities of individuals need to be measured. Based on the theory [Ballerini et al. 2008] that individuals tend to interact with a fixed number of its neighbors instead of keeping in touch with all the neighbors within a fixed distance, we select k nearest neighbors of individual i rather than find all its neighbor individuals within a fixed range. The traditional KNN method labels the neighbors as 1 and the nonneighbors as 0, which loses some soft distance information. Our method makes better use of spatial information by recording the distance between neighbors. For every individual i at time t, we find its k nearest neighbors and define a distance matrix D_t as

$$D_t(i, j) = \begin{cases} \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} & \text{if } j \in N(i) \\ 0 & \text{if } j \notin N(i) \end{cases}, \tag{2}$$

where $N(i)$ is a set of K nearest neighbors of individuals i, and (x_i, y_i) is the spatial coordinates of i.

However, although collectiveness is supposed to be a bottom feature that could be measured simply by just one frame in Zhou et al. [2014], we consider collective motion beyond just a collection of individuals, but a dynamic system with some fundamental intrinsic properties. In order to reveal the dynamic characteristic of a crowd system, it is crucial to add temporal information into the similarity computation. Our assumption is that the individuals used to keep their relationship with each other between frames. The correlation in the current frame is highly associated with that in the previous frame. If points i and j are close on spatial location and velocity in the former frame, we suppose they are more likely to have a high degree of similarity in the current frame.

Considering the preceding factors, we incorporate the temporal connection into the similarity definition as

$$S_t(i,j) = \begin{cases} (1-\omega)C_t(i,j)e^{-D_t(i,j)} + \omega S_{t-1}(i,j) & \text{if } D_t(i,j) > 0 \\ 0 & \text{if } D_t(i,j) = 0 \end{cases}, \tag{3}$$

where $S_{t-1}(i, j)$ is the similarity of i and j in the former frame, f is a proportional function, and ω regulates the weight of $S_{t-1}(i, j)$. We initialize all the elements in S_0 as 0. Temporal information is integrated into S_t because similarity in the former frame is used in the calculation of S_t. From this equation, if points i and j are consistent on space and velocity, and have a high degree of similarity in the former frame, their similarity in the current frame will be high too.

As the similarity of individuals jointly combines spatial and temporal information, our method is capable of handling the changeable interaction between individuals.

4.2. Unsteady Individuals Detection

In real-world crowded scenes, there are inevitably some individuals whose velocities change irregularly. Those abnormal individuals behave quite differently from their neighbors. Thus, they could cause a mistaken measurement of other individuals' collectiveness. In order to reduce this problem, we propose the *stability* descriptor to describe the degree of individuals moving steadily in a crowd system.

We first denote the velocity stability of point i at time t as

$$SP_i(t) = \frac{\boldsymbol{v}_i(t) \cdot \boldsymbol{v}_i(t-1)}{|\boldsymbol{v}_i(t)| \cdot |\boldsymbol{v}_i(t-1)|}. \tag{4}$$

It is not reliable to measure the stability just by a point's own velocity, because sometimes all the individuals fiercely change their velocities together. So, it is necessary to record the deviation between the individual and all its neighbors on velocity. The velocity deviation between point i and its neighbors is defined as

$$SN_i(t) = \frac{\boldsymbol{v}_i(t) \cdot \boldsymbol{v}_{iN}(t)}{|\boldsymbol{v}_i(t)| \cdot |\boldsymbol{v}_{iN}(t)|}, \tag{5}$$

where $\boldsymbol{v}_{iN}(t)$ is the mean velocity of the neighbors of i. Thus, the stability of points i at time t is defined as

$$ST_i(t) = SP_i(t) + SN_i(t). \tag{6}$$

Point i is considered to be unsteady if $ST_i(t) < \epsilon$ (ϵ is a threshold). In this case all the elements in the ith row and ith column of S_t are removed, and the individual collectiveness of point i is denoted as 0.

By detecting and discarding all these unsteady individuals, the robustness of the proposed method is improved.

4.3. Collectiveness Calculation by Topological Relationship

According to Ballerini et al. [2008], interaction between individuals of a collective motion does not depend on their metric distance, but on their topological distance. Moreover, manifold structures often emerge in collective motions and this case can be called collective manifold as shown in Figure 4. From this point of view, we explore the topological relationship between individuals on the basis of manifold learning to deal with various collective motions.

Zhou et al. [2014] investigated the topological relationship of two individuals by accumulating their similarity on all paths, so some useless paths are included. And it

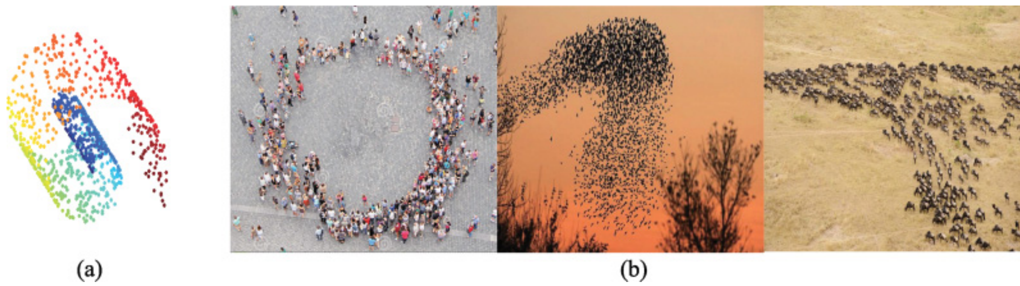


Fig. 4. (a) A schematic diagram of manifold structure. (b) Collective manifolds of humans, horse herd, and bird flock. We can figure out that the movement of individuals in a manifold could be quite different. URL of the images: https://www.google.com.hk/imghp?gws_rd=cr,ssl.

is not so intuitive and reasonable to reveal the intrinsic property of collective manifold, which is formed by the propagation of information through neighbors [Zhou et al. 2003b]. To capture the intrinsic property of collective manifold more intuitively, we exploit the manifold structure of a collective motion by spreading information among individuals. Here we employ the manifold ranking method [Zhou et al. 2003b], which spreads ranking scores through neighbors, to exploit the relationship of individuals in collective manifolds. In Zhou et al. [2003b], the manifold ranking method is designed to classify data with manifold structure. Given a set of data points X, some points are queries and the rest need to be ranked according to their relevance to the queries. Let f denote a ranking function that assigns a ranking score f_i to each data point i, and f can be viewed as a vector $f = [f_1, \ldots, f_n]^T$. $Y = [y_1, \ldots, y_n]^T$ is defined as an indication vector, where $y_i = 1$ if i is a query and $y_i = 0$ otherwise. Given a $n \times n$ symmetrical affinity matrix W, the degree matrix is $D = diag\{\sum_j W_{1j}, \sum_j W_{2j}, \ldots, \sum_j W_{nj}\}$, and the normalized Laplacian matrix is $L = D^{-1/2} W D^{-1/2}$. Initialize f as Y, by iterating

$$f = \alpha L f + (1 - \alpha) Y \tag{7}$$

until convergence, in which α is a parameter in $[0, 1)$, every point spreads its ranking score to its neighbors, and the manifold structure of data can be exploited. Note that, α controls the contribution to the ranking scores from the neighbors. As Yang et al. [2013] pointed out, we can get the final ranking score vector as

$$f^* = (1 - \alpha)(D - \alpha W)^{-1} Y, \tag{8}$$

where a higher score means a closer relevance to the query.

In this work, we view the quantifying of topological relationship as a one-query manifold ranking problem. We set W as $(S_t + S_t^T)/2$ to make it symmetrical, which is necessary for the convergence of the iteration in Equation (7). Let individual j be the only one query; all individuals' ranking score can be computed by

$$f^* = (1 - \alpha)(D - \alpha W)^{-1} e_j, \tag{9}$$

where e_j is a column vector with its j-th element as 1 and the other elements as 0. In Equation (9), the i-th element of f^* indicates the individual i's ranking score, which can be regarded as i's topological relevance to query j, and it is not hard to see that f^* is equivalent to the j-th column of $(D - \alpha W)^{-1}$. Let each individual be the query iteratively; we can get a matrix

$$Z_t = (1 - \alpha)(D - \alpha W)^{-1}, \tag{10}$$

in which $Z_t(i, j)$ indicates the topological relationship between individual i and j in the t-th frame.

Thus, we can define the individual collectiveness of i as the sum of topological relationship between i and all the other individuals, so it can be written as

$$\phi_t(i) = [Z_t e]_i, \tag{11}$$

where e is a column vector with all elements as 1. We denote the set of individuals as C; then, the entirety-level collectiveness of the crowd system at time t is defined as

$$\Phi_t = \frac{1}{|C|} e^T Z_t e, \tag{12}$$

which means the average value of all the individual collectiveness. Note that, unsteady individuals are also included in C, as they are also part of the crowd system, and their individual collectiveness is 0 as mentioned previously. Since S_t contains temporal information, both the individual-level and entirety-level collectiveness are more consistent along time.

By manifold learning, we quantify the complicated topological relationship between individuals, so the proposed method is suitable to deal with the collective manifolds. The whole procedure of measuring collectiveness is shown in Algorithm 2.

ALGORITHM 2: Measuring Collectiveness

Input: Refined points set $\Omega = \{\Omega_t\}_{t=1}^{N}$.
Output: $\phi_t(i)$ and Φ_t.
for *each points set* Ω_t **do**
  Get a distance matrix D_t, which contains spatial information of points by Equation (2);
  Combine D_t with the velocity correlation of points and temporal information between frames and get a similarity matrix S_t by Equation (3).
end
for *each similarity matrix* S_t **do**
  Measure the stability of every point by Equation (6);
  Remove the unsteady points;
  Get the matrix Z_t by Equation (10);
  Calculate the individual-level collectiveness by Equation (11) and the entirety-level collectiveness by Equation (12).
end

5. EXPERIMENT

In this section, we conduct intensive experiments to evaluate the performance of the proposed method. First, the effect of the proposed manifold learning method is evaluated. Then we employ a Self-Driven Particle (SDP) model [Vicsek et al. 1995] to simulate the emergence of collective motion. Through the comparison between the ground truth and the results of collectiveness algorithms, a quantitative and qualitative judgement can be obtained. In the end, we perform experiments on the Collective Motion dataset [Zhou et al. 2014], which contains videos of different real-world crowd systems. In order to evaluate the consistency between our results and human perception, we classify all the videos into three categories according to the calculated collectiveness and compare the results with ground truth. The parameters $\eta, k, \omega, \epsilon, \alpha$ are chosen as $0.9, 30, 0.1, 0.5, 0.99$ empirically throughout all the experiments. Additionally, two state-of-the-art methods MCC [Zhou et al. 2014] and CT [Shao et al. 2014a] for collectiveness measurement are involved in experiments as the baseline for comparison. Note that, Collective Transition (CT) is not taken into account in some experiments because it can neither measure collectiveness on individual level nor measure entirety-level collectiveness of frame by frame.

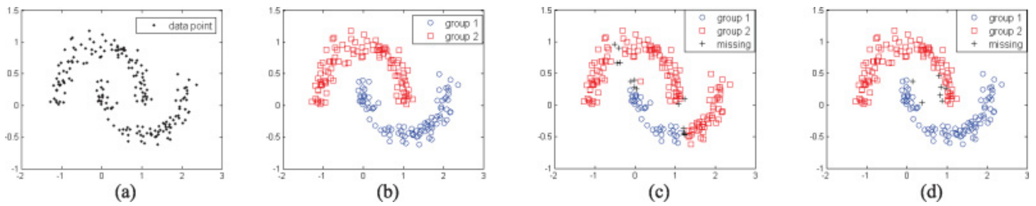


Fig. 5. (a) Data in two moons pattern. (b) Ground-truth clustering result. (c) Clustering result of MCC. (d) Clustering result of the proposed manifold learning method. It can be seen that MCC clusters the data in the below moon into two different groups incorrectly, while our method obtains a good clustering result.

5.1. Comparison of Manifold Learning Methods

Here, the proposed manifold learning method is evaluated to see whether it can capture the topological relationship of individuals. We compare our manifold learning method with MCC [Zhou et al. 2014], which also involves a different type of manifold learning strategy and achieves state-of-the-art performance. To learn the topological relationship between individual i and j, the proposed manifold learning method propagates ranking scores through neighbors, while MCC accumulates similarity of individuals along all paths that connect i and j.

Experiments are conducted on the toy dataset [Zhou et al. 2003a], which is constructed by a set of points in two moons pattern, as shown in Figure 5(a). Points in one moon should keep higher consistency with each other than with points in the other moon. As Zhou et al. [2014] pointed out, a topological relationship can be utilized to cluster individuals. We compare the two methods by their performance on point clustering, and the ground-truth clustering result is shown in Figure 5(b).

Here we set the affinity matrix W as $W_{ij} = e^{-\sqrt{(x_i-x_j)^2+(y_i-y_j)^2}}$, where (x_i, y_i) is the spatial coordinates of i. The proposed method computes the topological relationship matrix as $Z_t = (1-\alpha)(D-\alpha W)^{-1}$, and MCC computes the matrix as $Z_t = (I-zW)^{-1} - I$, where z is set to 0.025 and I is a diagonal matrix with all elements set to 1. Given the topological matrix, points in the toy dataset are clustered into groups using the cluster merging method [Zhou et al. 2014].

Figures 5(c) and 5(d) visualize the results generated by the two methods. MCC incorrectly clusters the data in the bottom moon into two groups, while our method achieves good performance. And we have found that our manifold learning method also performs well when dealing with real-world crowd videos. As MCC accumulates similarity along all paths, some useless paths are included, so it cannot calculate the topological relationship accurately. The proposed manifold learning method propagates information through only neighbors, so it is more suitable to capture the manifold structure formed by the individuals. Consequently, the proposed manifold learning method is more capable of measuring the topological relationship between individuals.

5.2. Experiments on SDP Model

In order to evaluate the performance of our method, we do experiments on the SDP model [Vicsek et al. 1995]. SDP was first used to study the coherent emergency of collective motion in systems of moving particles, and it has been widely employed for investigating collective motions. We employ it because it has high consistency with real-world crowd systems [Zhang et al. 2010; Buhl et al. 2006; Zhou et al. 2014] and its ground truth can be used for evaluation.

In SDP, particles are initialized randomly and driven with a constant speed. The moving direction of individual i at time t is

$$\theta(t) = \langle \theta_{t-1} \rangle_r + \Delta\theta, \tag{13}$$

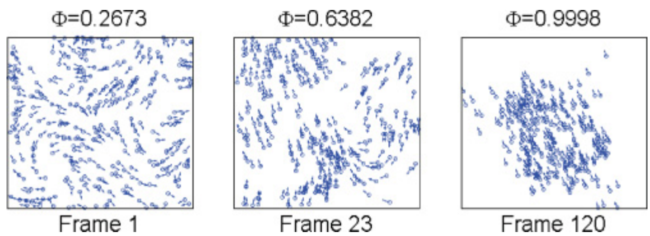


Fig. 6. The simulation results of the SDP model. At the beginning, the particles move randomly and no uniform motion pattern is manifest. Gradually, they move into a collective motion, and the entirety-level collectiveness increases at the same time.

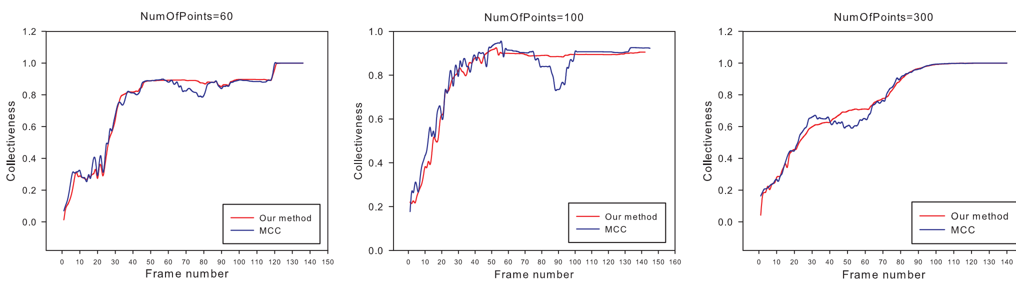


Fig. 7. Entirety-level collectiveness comparison calculated by the proposed method and the MCC method. From left to right, three sets of experiments are conducted with different points number in the SDP simulation procedure. From the figures, we can see clearly that the collectiveness computed by our method changes gradually, while that measured by MCC fluctuates fiercely. So the proposed method can reflect the phase transition of the SDP model more accurately.

where $\theta(t)$ is the moving direction of i, $\langle\theta_{t-1}\rangle_r$ is the average moving direction of the particles surrounding i, and $\Delta\theta$ is a random number chosen from the interval $[-\pi, \pi]$.

5.2.1. Evaluation of Entire-Level Collectiveness Accuracy. In SDP, particles are initialized with random moving directions and spatial locations. According to Equation (13), every particle moves toward the average moving direction of its neighbors. So, all the particles evolve into a collective motion gradually, as shown in Figure 6. Thus, according to the definition of entirety-level collectiveness, its calculated values on SDP should increase with time.

We show the calculation results of entirety-level collectiveness by the proposed method and MCC in Figure 7. MCC measures the collectiveness of these particles according to their spatial location and velocity in the current frame, which is simple and straightforward. However, the particles move toward all directions and their velocities are not stable at the beginning. This makes the collectiveness calculated by MCC fluctuate dramatically. On the contrary, the proposed method utilizes the neighborhood relationships of those particles in the former frame. Since neighboring particles tend to remain in their relationship between frames, our method can capture the individuals' relationships over time. Thus, the proposed method is more capable of reflecting the actual evolvement of a collective motion.

5.2.2. Evaluation of Individual-Level Collectiveness Classification. Crowd systems in nature always consist of various kinds of individuals. Similarly, Zhou et al. [2014] added outlier particles to the SDP model to simulate the mixed-collective motion, as shown in Figure 8(a). In this model, the outlier particles move with constant velocities all the time, while the self-driven particles move toward the average direction of their neighbors. When self-driven particles turn into collective motions, their individual

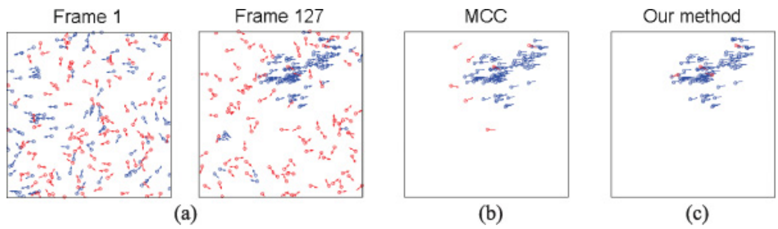


Fig. 8. (a) 300 self-driven particles (blue) with 300 outliers (red). Self-driven particles gradually move into a collective motion while outliers do not. (b) Outliers extracted by MCC. (c) Outliers extracted by our method. It can be seen that our method removes more outliers while extracting the same number of self-driven particles.

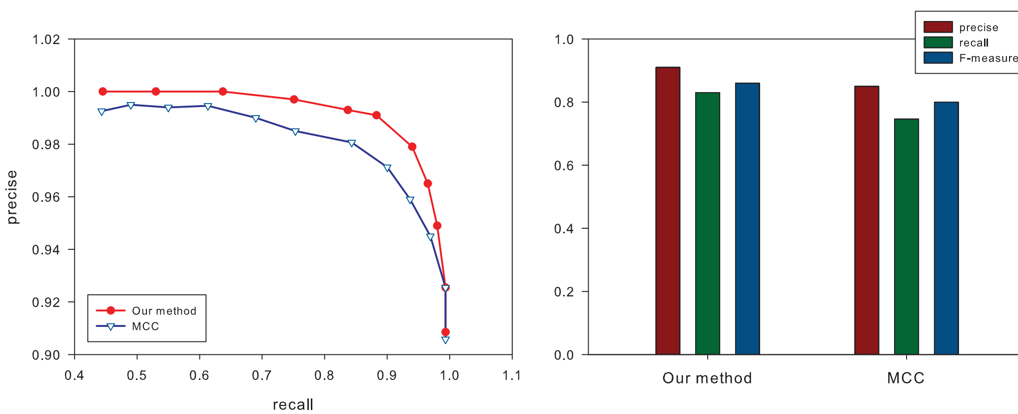


Fig. 9. The precise-recall comparison for extracting self-driven particles from outliers with constant velocities. (a) Precision-recall curves. (b) Average precision, recall, and F-measure bars.

collectiveness increases. But the outlier particles keep low individual collectiveness all the time. Thus, we can extract self-driven particles from outliers by comparing their calculated individual collectiveness, as shown in Figure 8(c).

We then compare the performance of the proposed method with MCC. Figure 9 demonstrates that our method outperforms MCC. Achieving at the same precise value, our method can extract more self-driven particles; with the same recall value, the proposed method is more accurate. This is because MCC measures collectiveness simply by one frame, and it cannot consider the neighbor relationship between individuals over time. Since temporal information is exploited, the proposed method is more capable of detecting outlier particles whose neighbors change frequently. So, our method achieves better performance than MCC on this model.

In real-world crowded scenes, abnormal individuals may change their velocities irregularly. So, it is not enough to assume the outlier particles keep their velocities steady all the time. Rather than just supposing the outlier particles keep their velocities unchanged, we propose a new mixed-SDP model to evaluate the performance. In this developed model, outliers change their moving directions randomly. Then we extract self-driven particles from the outliers. As shown in Figure 10, the proposed method outperforms MCC. Outliers whose velocities change all the time are better detected by our method. This justifies from another aspect that the proposed method outperforms MCC.

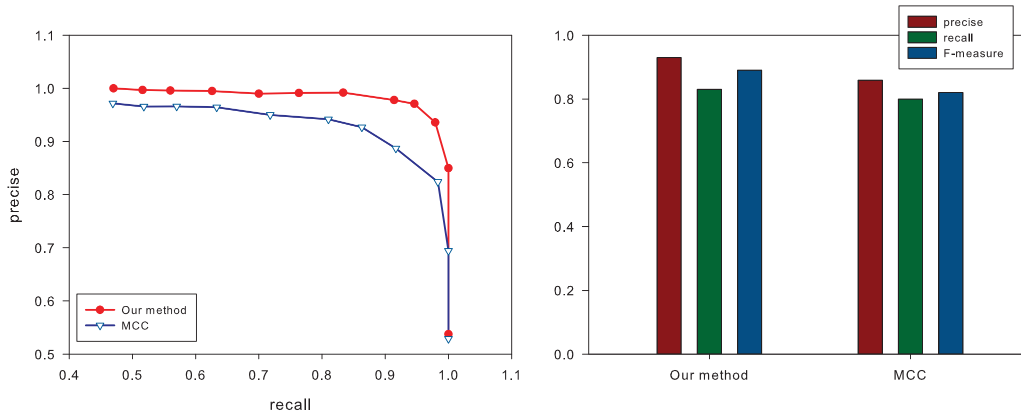


Fig. 10. The precise-recall comparison for extracting self-driven particles from outliers with random velocities. (a) Precision-recall curves. (b) Average precision, recall, and F-measure bars.

5.3. Experiments on Human Labeled Data

To further evaluate the performance of our collectiveness model, we compare its consistency with the labeled human perception on real-world crowd videos.

5.3.1. Date Set. We perform experiments on the Collective Motion Database [Zhou et al. 2014], which contains 413 video clips from 62 different crowded scenes and each video clip consists of 100 frames. The ground truth of this database has been labeled by human rating. To be specific, each video clip was rated by 10 subjects, who scored the video as 0, 1, and 2. A higher score means a higher level of collective motion, and all the videos are sorted into high, medium, and low collectiveness by majority voting.

5.3.2. Performance Evaluation on Manifold Structure. Experiments are conducted on a crowd scene with manifold structure, as shown in Figure 11(a), to evaluate the performance. In Figure 11(a), most of the individuals are moving along the same running track, and they can be highly regarded as the same group. According to the definition of collectiveness, the individuals' collectiveness should be similar and high.

We measure the individuals' collectiveness and show their distribution in Figure 11(e). For comparison, Figure 11(d) shows the result obtained by MCC. Figures 11(b) and 11(c) show the initial tracked feature points (treated as individuals in MCC) and the refined points (treated as individuals in our method), respectively. It can be obviously seen that the distribution obtained by our method is more concentrated than MCC, and most individuals have high collectiveness. That is to say, by using our method, individuals in the running track are more connected and have similar collectiveness. The good result comes mainly from our manifold learning method. So, the proposed method is more capable of handling crowd systems with manifold structures.

5.3.3. Performance Evaluation on Entire-Level Collectiveness. With the previously introduced dataset, we evaluate how the proposed method can classify the videos and compare its results with MCC and CT. Three categories are generated and the classified categories are compared with the human labeled ground truth. The precision-recall curves and the averaged precision, recall and F-measure bars are shown in Figures 12(a)–12(c). We can see that the proposed method has better discriminative capability than MCC and CT. Our method always achieves higher F-measure. MCC and

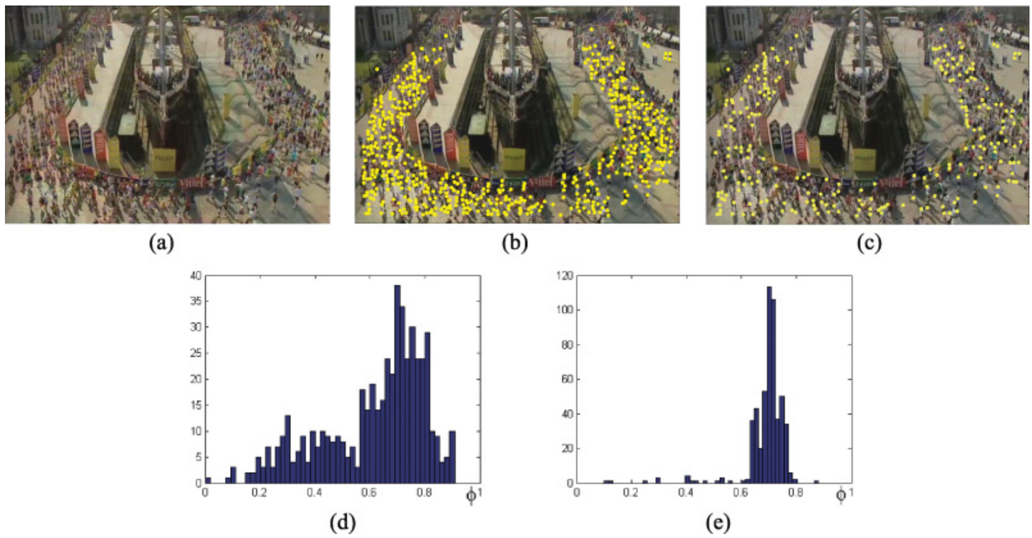


Fig. 11. (a) A marathon running scene. It is intuitive that most individuals belong to the same group. (b) Initial tracked feature points (treated as individuals in MCC). (c) Tracked feature points after point selection (treated as individuals in our method). (d) Histogram of individual collectiveness measured by MCC. (e) Histogram of individual collectiveness measured by the proposed method. It can be obviously seen that in (e) most individuals have similar collectiveness, which is consistent with our motivation. The image is selected from the Collective Motion dataset [Zhou et al. 2014].

CT share the same shortcoming in that they cannot extract individuals precisely. In addition, MCC neglects temporal information, and CT relies on complete trajectories of individuals, which are difficult to obtain. As a result, the proposed method outperforms MCC and CT.

As shown in Figure 12(d), some failure cases of our method still exist. One cause of the classification error is that there are unavoidable overlaps between low-medium and medium-high collective motions. These videos are difficult to classify even for humans, and more accurate ground truth needs to be proposed. Another reason is that inaccurate tracking can affect the measuring of collectiveness significantly, since our method is based on tracking results. This drawback cannot be avoided by most between-frame association based algorithms.

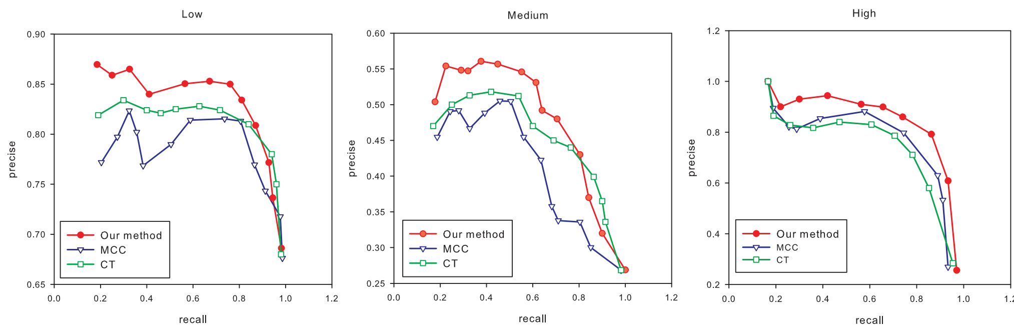
6. DISCUSSION

In this section, some issues about the proposed method are discussed. The first one is about the parameter selection of segmentation algorithm. The second one is whether the point selection method is needed to quantify the collectiveness accurately. In the end, the necessity of using temporal information is analyzed.

For quantitatively discussing these issues, a set of comparative experiments are conducted on the Collective Motion Database, which contains various videos of crowd systems. Under different conditions, collectiveness of the videos is calculated and is classified into three categories. Then the results are compared with the ground-truth labels to see the performance.

6.1. Selection of Segmentation Parameter

The computation of the proposed point selection method is based on the segmentation algorithm. Each frame is divided into patches, and then redundant points are



Low	Our	MCC	CT
precise	0.83	0.75	0.80
recall	0.66	0.61	0.59
F-measure	0.72	0.63	0.67

(a)

High	Our	MCC	CT
precise	0.47	0.38	0.45
recall	0.61	0.56	0.57
F-measure	0.51	0.40	0.49

(b)

High	Our	MCC	CT
precise	0.81	0.73	0.69
recall	0.58	0.51	0.46
F-measure	0.67	0.59	0.56

(c)

(d)

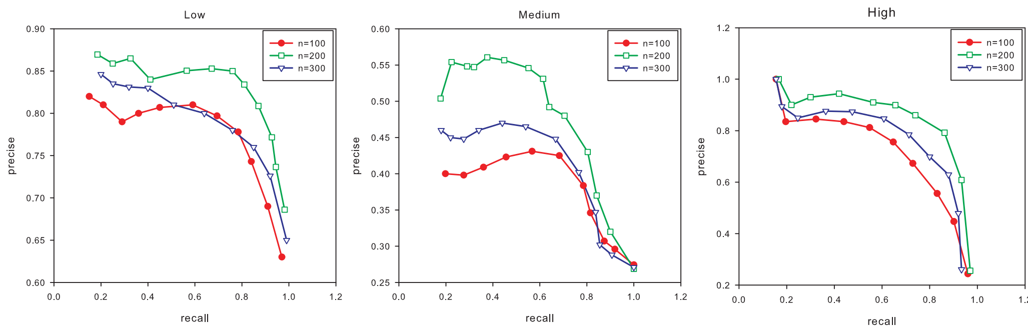
Fig. 12. Comparison of the precision-recall curves and averaged values on the low- (a), medium- (b), and high- (c) level videos. For a clear comparison, the red shows the best result. (d) Typical failures by the proposed method. The images are selected from the Collective Motion Dataset [Zhou et al. 2014].

abandoned. Therefore, the number of patches influences the final results. Too few patches will make some informative points discarded, and too many patches will make some redundant points retained. An appropriate choice of the segmentation parameter is essential to get a good result.

We denote the number of patches as n, and evaluate the performance of our method when n is 100, 200, and 300. Under a certain n, videos are classified into low, medium, and high collectiveness. The precise-recall curves and averaged precision, recall, and F-measure values are shown in Figure 13. It can be seen that the proposed method mostly achieves the best result with $n = 200$. We may infer that 200 is more appropriate. Throughout this article, the segmentation is all conducted under $n = 200$.

6.2. Effects of Point Selection

A Point Selection Method (PSM) is proposed to select the most representative tracking points and extract individuals from frames accurately. In order to evaluate the usefulness of PSM, several experiments are conducted on the Collective Motion Database. Collectiveness of videos is respectively measured with and without PSM by our method. Then we compare the classification results of the three categories with ground truth to evaluate the performance.



Low

n	100	200	300
precise	0.75	0.83	0.79
recall	0.64	0.66	0.65
F-measure	0.58	0.72	0.69

(a)

Medium

n	100	200	300
precise	0.38	0.47	0.44
recall	0.51	0.61	0.55
F-measure	0.45	0.51	0.47

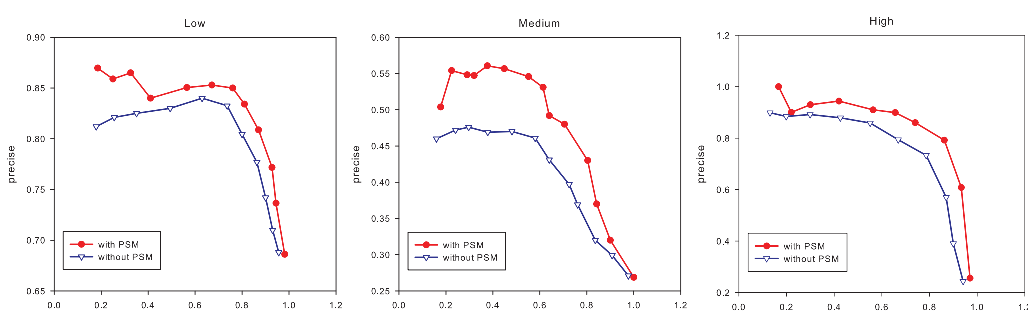
(b)

High

n	100	200	300
precise	0.72	0.81	0.79
recall	0.52	0.58	0.57
F-measure	0.61	0.67	0.63

(c)

Fig. 13. The precision-recall curves and averaged values of classifying low (a), medium (b), and high (c) collectiveness videos by our method when $n = 100$, 200, and 300. For a clear comparison, the red one shows the best result.



Low

	with PSM	without PSM
precise	0.83	0.78
recall	0.66	0.66
F-measure	0.72	0.69

(a)

High

	with PSM	without PSM
precise	0.47	0.40
recall	0.61	0.57
F-measure	0.51	0.45

(b)

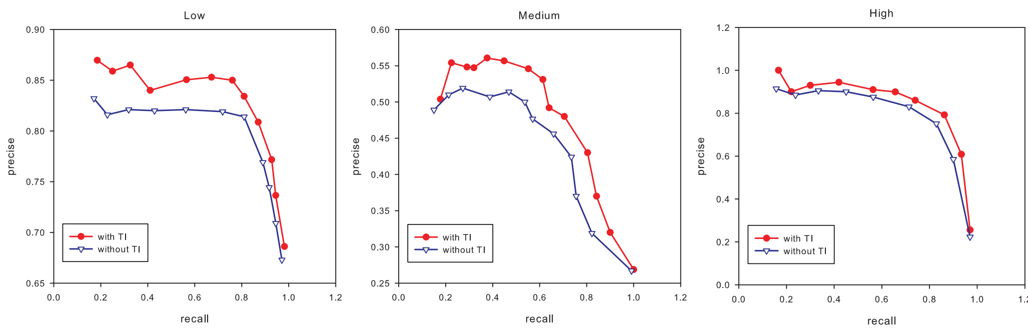
High

	with PSM	without PSM
precise	0.81	0.73
recall	0.58	0.57
F-measure	0.67	0.62

(c)

Fig. 14. The precision-recall curves and averaged values of classifying low (a), medium (b), and high (c) collectiveness videos with and without point selection step. For a clear comparison, the red one shows the best result.

It can be seen in Figure 14 that with PSM, the classification performance is better than that without PSM. With PSM, redundant tracking points are discarded so that individuals can be extracted more precisely from a crowd system. Then the relationship between individuals can be captured more accurately. Since we measure collectiveness by exploiting the relationship between individuals, PSM helps to compute collectiveness more correctly and achieves higher averaged recall, precise and F-measure values in experiments.



Low	with TI	without TI
precise	0.83	0.79
recall	0.66	0.63
F-measure	0.72	0.67

(a)

Medium	with TI	without TI
precise	0.47	0.45
recall	0.61	0.56
F-measure	0.51	0.48

(b)

High	with TI	without TI
precise	0.81	0.79
recall	0.58	0.57
F-measure	0.67	0.64

(c)

Fig. 15. The precision-recall curves and averaged values of classifying low (a), medium (b), and high (c) collectiveness videos by our method with and without temporal information. For a clear comparison, the red one shows the best result.

6.3. Necessity of Temporal Information

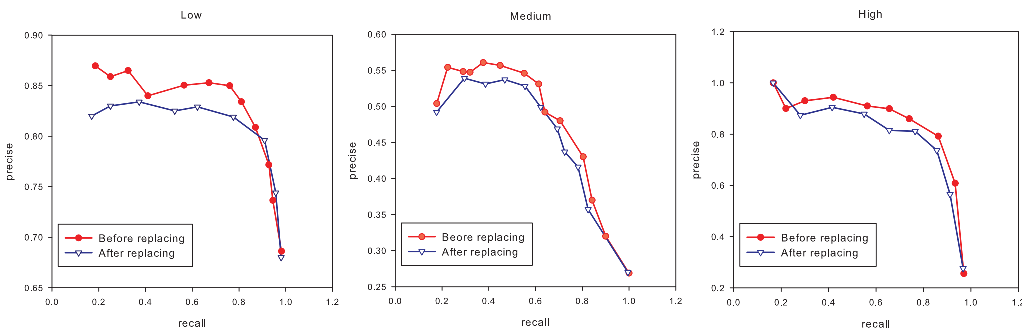
In the calculation of collectiveness, temporal information is utilized. If two individuals are neighbors in the former frame, they tend to keep their neighbor relationship in the current frame. Consequently, their individual-level collectiveness will be more consistent along time. Here we discuss whether the temporal information is necessary for the calculation of collectiveness.

We compute collectiveness of the videos with and without temporal information separately, and use them to classify the videos. Figure 15 plots the precise-recall curves and shows the averaged precise, recall, and F-measure values. It can be seen that with temporal information, the performance is better than that without temporal information. By using temporal information, the relationship between individuals over time can be captured, and the collectiveness can be calculated more accurately and keep high consistency with human perception. Therefore, temporal information is necessary to calculate collectiveness precisely.

6.4. Effects of Manifold Learning Method

Here, we evaluate how the proposed manifold learning method contributes to the performance of the whole algorithm. In the experiments, we replace the proposed manifold learning method with the one used in MCC, and then examine the change of the overall performance.

Figure 16 shows that after replacing the proposed manifold learning method, the performance of our collectiveness-measuring method is not as good as before. This is to say, compared with the manifold learning method of MCC, the proposed manifold learning method demonstrates superiority on the measuring of collectiveness. Therefore, the proposed manifold learning method does improve the overall performance of our collectiveness-measuring method.



Low

	Before	After
precise	0.83	0.76
recall	0.66	0.68
F-measure	0.72	0.70

(a)

Medium

	Before	After
precise	0.47	0.43
recall	0.61	0.59
F-measure	0.51	0.49

(b)

High

	Before	After
precise	0.81	0.75
recall	0.58	0.59
F-measure	0.67	0.64

(c)

Fig. 16. The precision-recall curves and averaged values of classifying low (a), medium (b), and high (c) collectiveness videos by our method before and after replacing the proposed manifold learning method with that of MCC. For a clear comparison, the red one shows the best result.

7. CONCLUSION AND FUTURE WORK

In this article, a new collectiveness-measuring method has been presented. In this procedure, a point selection method is developed and it can select the most useful ones of tracked feature points to represent individuals of a crowd system. A stability descriptor is presented to characterize whether an individual keeps a steady relationship with others. By jointly exploring the spatial and temporal clues of the crowd system, the proposed method can quantitatively calculate a collectiveness measure on the basis of the topological relationship between individuals. To validate the usefulness and effectiveness of our method, we compare its performance on the SDP model and a set of 431 video clips with a state-of-the-art algorithm. Intensive experiments show its robustness and higher consistency with human perception.

In further work, we would like to extend our method for group detection based on the individual collectiveness measurement. Additionally, we are interested in applying our method to some specific applications in the field of multimedia, such as the analysis of crowd behavior. By analyzing the crowd behavior, we can obtain the attributes of a crowd scene and then describe the crowd video by text, which will be a clue for some other applications, such as video retrieval, video summary, and so on. Furthermore, crowd simulation is also one of our interests.

REFERENCES

Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurélien Lucchi, Pascal Fua, and Sabine Süsstrunk. 2012. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34, 11 (2012), 2274–2282.

Saad Ali and Mubarak Shah. 2007. A Lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. In *IEEE Conference on Computer Vision and Pattern Recognition*. 1–6.

Saad Ali and Mubarak Shah. 2008. Floor fields for tracking in high density crowd scenes. In *IEEE Conference on European Conference on Computer Vision*. 1–14.

João Emílio Almeida, Rosaldo J. F. Rossetti, and António Leça Coelho. 2013. Crowd simulation modeling applied to emergency and evacuation simulations using multi-agent systems. *CoRR* abs/1303.4692 (2013).

Michele Ballerini, Nicola Cabibbo, Raphael Candelier, Andrea Cavagna, Evaristo Cisbani, Irene Giardina, Vivien Lecomte, Alberto Orlandi, Giorgio Parisi, Andrea Procaccini, and others. 2008. Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study. *Proceedings of the National Academy of Sciences* 105, 4 (2008), 1232–1237.

Yassine Benabbas, Nacim Ihaddadene, and Chabane Djeraba. 2011. Motion pattern extraction and event detection for automatic visual surveillance. *EURASIP Journal on Image and Video Processing* 2011 (2011), 7.

Andrew Best, Sahil Narang, Sean Curtis, and Dinesh Manocha. 2014. DenseSense: Interactive crowd simulation using density-dependent filters. In *The Eurographics/ACM SIGGRAPH Symposium on Computer Animation (SCA'14)*. 97–102.

Jerome Buhl, David J. Sumpter, Iain D. Couzin, Joseph Hale, Emma Despland, Esther Miller, and Stephen J. Simpson. 2006. From disorder to order in marching locusts. *Science* 312, 5778 (2006), 1402–1406.

Ming-Ching Chang, Nils Krahnstoever, and Weina Ge. 2011. Probabilistic group-level motion analysis and scenario recognition. In *IEEE International Conference on International Conference on Computer Vision*. 747–754.

Iain D. Couzin. 2009. Collective cognition in animal groups. *Trends in Cognitive Sciences* 13, 1 (2009), 36–43.

Weina Ge, Robert T. Collins, and Barry Ruback. 2012. Vision-based analysis of small groups in pedestrian crowds. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34, 5 (2012), 1003–1016.

Abhinav Golas, Rahul Narain, and Ming C. Lin. 2013. Hybrid long-range collision avoidance for crowd simulation. In *Symposium on Interactive 3D Graphics and Games (I3D'13)*. 29–36.

Stephen J. Guy, Jur van den Berg, Wenxi Liu, Rynson Lau, Ming C. Lin, and Dinesh Manocha. 2012. A statistical similarity measure for aggregate crowd dynamics. *ACM Transactions on Graphics* 31, 6 (2012), 190:1–190:11.

Rachel Heck, Michael N. Wallick, and Michael Gleicher. 2007. Virtual videography. *ACM Transactions on Multimedia Computing, Communications and Applications* 3, 1 (2007).

Dirk Helbing, Illés Farkas, and Tamas Vicsek. 2000. Simulating dynamical features of escape panic. *Nature* 407, 6803 (2000), 487–490.

Dirk Helbing and Peter Molnar. 1995. Social force model for pedestrian dynamics. *Physical Review E* 51, 5 (1995), 4282–4286.

Weiming Hu, Xuejuan Xiao, Zhouyu Fu, Dan Xie, Tieniu Tan, and Stephen J. Maybank. 2006. A system for learning statistical motion patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28, 9 (2006), 1450–1464.

Kaiqi Huang, Dacheng Tao, Yuan Yuan, Xuelong Li, and Tieniu Tan. 2009. View-independent behavior analysis. *IEEE Transactions on Systems, Man, and Cybernetics, Part B* 39, 4 (2009), 1028–1035.

Pierre-Marc Jodoin, Yannick Benezeth, and Yi Wang. 2013. Meta-tracking for video scene understanding. In *IEEE Conference on Advanced Video and Signal Based Surveillance*. 1–6.

Louis Kratz and Ko Nishino. 2009. Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models. In *IEEE Conference on Computer Vision and Pattern Recognition*. 1446–1453.

Louis Kratz and Ko Nishino. 2012. Going with the flow: Pedestrian efficiency in crowded scenes. In *Proceedings of the 12th European Conference on Computer Vision (ECCV 2012), Part IV*. 558–572.

Stanislav N. Kružkov. 1970. First order quasilinear equations in several independent variables. *Sbornik: Mathematics* 10, 2 (1970), 217–243.

Tian Lan, Leonid Sigal, and Greg Mori. 2012a. Social roles in hierarchical models for human activity recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*. 1354–1361.

Tian Lan, Yang Wang, Weilong Yang, Stephen N. Robinovitch, and Greg Mori. 2012b. Discriminative latent models for recognizing contextual group activities. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34, 8 (2012), 1549–1562.

Alon Lerner, Yiorgos Chrysanthou, and Dani Lischinski. 2007. Crowds by example. *Computer Graphics Forum* 26, 3 (2007), 655–664.

Teng Li, Huan Chang, Meng Wang, Bingbing Ni, Richang Hong, and Shuicheng Yan. 2015. Crowded scene analysis: A survey. *IEEE Trans. Circuits Syst. Video Techn.* 25, 3 (2015), 367–386.

Mehdi Moussaid, Simon Garnier, Guy Theraulaz, and Dirk Helbing. 2009. Collective information processing and pattern formation in swarms, flocks, and crowds. *Topics in Cognitive Science* 1, 3 (2009), 469–497.

Rahul Narain, Abhinav Golas, Sean Curtis, and Ming C. Lin. 2009. Aggregate dynamics for dense crowd simulation. *ACM Transactions on Graphics* 28, 5 (2009).

Stefano Pellegrini, Andreas Ess, Konrad Schindler, and Luc J. Van Gool. 2009. You'll never walk alone: Modeling social behavior for multi-target tracking. In *IEEE International Conference on International Conference on Computer Vision*. 261–268.

Craig W. Reynolds. 1987. Flocks, herds and schools: A distributed behavioral model. In *Annual Conference on Computer Graphics and Interactive Techniques*. 25–34.

Imran Saleemi, Lance Hartung, and Mubarak Shah. 2010. Scene understanding by statistical modeling of motion patterns. In *IEEE Conference on Computer Vision and Pattern Recognition*. 2069–2076.

Paul Scovanner and Marshall F. Tappen. 2009. Learning pedestrian dynamics from the real world. In *IEEE International Conference on International Conference on Computer Vision*. 381–388.

Jing Shao, Chen Change Loy, and Xiaogang Wang. 2014a. Scene-independent group profiling in crowd. In *IEEE Conference on Computer Vision and Pattern Recognition*. 2227–2234.

Ling Shao, Xiantong Zhen, Dacheng Tao, and Xuelong Li. 2014b. Spatio-temporal Laplacian pyramid coding for action recognition. *IEEE T. Cybernetics* 44, 6 (2014), 817–827.

Carlo Tomasi and Takeo Kanade. 1991. *Detection and Tracking of Point Features*. School of Computer Science, Carnegie Mellon University, Pittsburgh, PA.

Jur van den Berg, Stephen J. Guy, Ming C. Lin, and Dinesh Manocha. 2009. Reciprocal n-body collision avoidance. In *International Symposium on Robotics Research*. 3–19.

Tamás Vicsek, András Czirók, Eshel Ben-Jacob, Inon Cohen, and Ofer Shochet. 1995. Novel type of phase transition in a system of self-driven particles. *Physical Review Letters* 75, 6 (1995), 1226.

Qi Wang, Yuan Yuan, and Pingkun Yan. 2013a. Visual saliency by selective contrast. *IEEE Transactions on Circuits and Systems for Video Technology* 23, 7 (2013), 1150–1155.

Qi Wang, Yuan Yuan, Pingkun Yan, and Xuelong Li. 2013b. Saliency detection by multiple-instance learning. *IEEE Transactions on Cybernetics* 43, 2 (2013), 660–672.

Xiaogang Wang, Keng Teck Ma, Gee Wah Ng, and W. Eric L. Grimson. 2008. Trajectory analysis and semantic region modeling using a nonparametric Bayesian model. In *IEEE Conference on Computer Vision and Pattern Recognition*. 1–8.

Xiaogang Wang, Keng Teck Ma, Gee Wah Ng, and W. Eric L. Grimson. 2011. Trajectory analysis and semantic region modeling using nonparametric hierarchical Bayesian models. *International Journal of Computer Vision* 95, 3 (2011), 287–312.

Xiaogang Wang, Xiaoxu Ma, and W. Eric L. Grimson. 2009. Unsupervised activity perception in crowded and complicated scenes using hierarchical Bayesian models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31, 3 (2009), 539–555.

Chuan Yang, Lihe Zhang, Huchuan Lu, Xiang Ruan, and Ming-Hsuan Yang. 2013. Saliency detection via graph-based manifold ranking. In *2013 IEEE Conference on Computer Vision and Pattern Recognition*. 3166–3173. DOI:http://dx.doi.org/10.1109/CVPR.2013.407

Yang Yang, Jingen Liu, and Mubarak Shah. 2009. Video scene understanding using multi-scale analysis. In *IEEE Conference on Computer Vision*. 1669–1676.

Machi Zawidzki, Mohcine Chraibi, and Katsuhiro Nishinari. 2014. Crowd-Z: The user-friendly framework for crowd simulation on an architectural floor plan. *Pattern Recognition Letters* 44 (2014), 88–97.

H. P. Zhang, A. Ber, E. L. Florin, and H. L. Swinney. 2010. Collective motion and density fluctuations in bacterial colonies. *Proceedings of the National Academy of Science* 107, 31 (2010), 13626–13630.

Peng Zhang, Hong Liu, and Yanhui Ding. 2015. Crowd simulation based on constrained and controlled group formation. *The Visual Computer* 31, 1 (2015), 5–18.

Lin Zhao, Xinbo Gao, Dacheng Tao, and Xuelong Li. 2015. A deep structure for human pose estimation. *Signal Processing* 108 (2015), 36–45.

Xin Zhao, Xue Li, Chaoyi Pang, Quan Z. Sheng, Sen Wang, and Mao Ye. 2014. Structured streaming skeleton—A new feature for online human gesture recognition. *ACM Transactions on Multimedia Computing, Communications and Applications* 11, 1 (2014), 22:1–22:18.

Liping Zheng, Juan Zhang, Chenglong Zhou, and Xiaoping Liu. 2014. Texture diversity generating method for characters in crowd simulation. *Journal of Graphics* 35, 1 (2014), 110–114.

Bolei Zhou, Xiaoou Tang, and Xiaogang Wang. 2012. Coherent filtering: Detecting coherent motions from crowd clutters. In *IEEE Conference on European Conference on Computer Vision*. 857–871.

Bolei Zhou, Xiaoou Tang, Hepeng Zhang, and Xiaogang Wang. 2014. Measuring crowd collectiveness. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36, 8 (2014), 1586–1599.

Bolei Zhou, Xiaogang Wang, and Xiaoou Tang. 2011. Random field topic model for semantic region analysis in crowded scenes from tracklets. In *IEEE Conference on Computer Vision and Pattern Recognition*. 3441–3448.

Bolei Zhou, Xiaogang Wang, and Xiaoou Tang. 2012. Understanding collective crowd behaviors: Learning a mixture model of dynamic pedestrian-agents. In *IEEE Conference on Computer Vision and Pattern Recognition*. 2871–2878.

Dengyong Zhou, Olivier Bousquet, Thomas Navin Lal, Jason Weston, and Bernhard Schölkopf. 2003a. Learning with local and global consistency. In *Advances in Neural Information Processing Systems 16 Neural Information Processing Systems (NIPS 2003)*. 321–328.

Dengyong Zhou, Jason Weston, Arthur Gretton, Olivier Bousquet, and Bernhard Schölkopf. 2003b. Ranking on data manifolds. In *Advances in Neural Information Processing Systems 16 (NIPS 2003)*. 169–176.

Suiping Zhou, Wentong Cai, Stephen John Turner, Bu-Sung Lee, and Junhu Wei. 2007. Critical causal order of events in distributed virtual environments. *ACM Transactions on Multimedia Computing, Communications and Applications* 3, 3 (2007).

Received April 2015; revised September 2015; accepted October 2015