



Data-Driven and Collision-Free Hybrid Crowd Simulation Model for Real Scenario

Qingrong Cheng, Zhiping Duan, and Xiaodong Gu 

Department of Electronic Engineering, Fudan University,
Shanghai 200433, China
xdgu@fudan.edu.cn

Abstract. In order to take into account evading mechanism and make more realistic simulation results, we propose a data-driven and collision-free hybrid crowd simulation model in this paper. The first part of the model is a data-driven process in which we introduce an algorithm called MS-ISODATA (Main Streams Iterative Self-organizing Data Analysis) to learn motion patterns from real scenarios. The second part introduces an agent-based collision-free mechanism in which a steering approach is improved and this part uses the output from the first part to guide its agents. The hybrid simulation model we propose can reproduce simulated crowds with motion features of real scenarios, and it also enables agents in simulation evade from mutual collisions. The simulation results show that the hybrid crowd simulation model mimics the desired crowd dynamics well. According to a collectiveness measurement, the simulation results and real scenarios are very close. Meanwhile, it reduces the number of virtual crowd collisions and makes the movement of the crowd more effective.

Keywords: Crowd simulation · Data-driven · Main streams · MS-ISODATA
Collision-free · Collectiveness

1 Introduction

Crowd related research has been a significant theme for a long time in the realm of intelligent optimization, computer vision, robotics and computer graphics. People usually come across many circumstance with huge amount of people or other living creatures. It is meaningful to analyze the pattern of crowd motion. Consequently, we can extract a lot of useful information on crowd motion and put it into use. Related work usually involves counting [1], classification [2], tracking [3], abnormality detection [4] and understanding [5]. Pedestrian's movement are affected continuously by other pedestrian and their surrounding environment, so they should change their walking direction and speed frequently. Meanwhile, the state of emergency is different from the normal status. The multi-agent reinforcement learning model of pedestrian emergent behaviors are presented in [6]. Many crowd analysis methods related to crowd motion pattern and crowd simulation using computer vision techniques can be found in a survey [7].

Crowd simulation is an active research field that has drawn strong interest of researchers in industry, academia and government. Reynolds proposed the BOID

model [8] in 1987. He defined three behaviors, i.e. cohesion, alignment and separation for each agent in the group and the virtual group can mimic behaviors of bird group well. Helbing et al. constructed social force model [9] in 1995 and the model is especially helpful in simulating evacuation in public area [10]. In the past decades, many agents based or force based models were put forward and we saw crowd simulation flourished consequently such as [11, 12]. However, although the agent or force based model works well in small group of simulation, it has some drawbacks such as quantities of parameters tuning and complex rules defining, which makes it extremely difficult to fulfil simulation with large numbers of agents.

It is natural to think that we can directly learning some motion patterns from real crowds and simulating virtual crowds directly with these patterns, which is not only free the model from complex parameters and rules, but also enable the simulation results resemble to the real crowds. Accordingly, some approaches based on data-driven method emerged. Lerner presented an example-based crowd simulation technique [13] and examples were created from tracked video segments of real pedestrian crowds. Lee also presented a data-driven method [12] of simulating a crowd of virtual humans that exhibit behaviors imitating real human crowds.

Inspired by these works, we propose a data-driven and agent-based hybrid model for crowd simulation in this paper. The model mainly consists of two parts. The first part is a data-driven process in which motion patterns of specific scenarios are learned from videos of real crowds. The second part is an agent-based collision-free procedure, which is proposed by improving a velocity-based steering approach [14]. In the first part, global features we call Main Streams (MS) of crowds are extracted. While in the second part, the collision-free mechanism which guarantees the microscopic details are physically realistic by avoiding mutual collisions. With both macroscopic and microscopic characteristic we obtain, we can reproduce realistic simulating crowds just as real scenarios.

2 The Hybrid Model

2.1 Overview

The overall framework of the hybrid model is illustrated in Fig. 1.

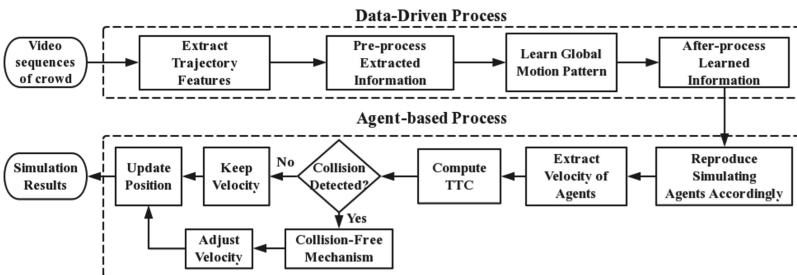


Fig. 1. The overall framework of our hybrid simulation model

2.2 The Data-Driven Process

Extract Trajectory Features. In our model, we would extract these common movement dynamics from observed video sequences and use them as global features of simulation. We know there exist some multi-target tracking algorithms and some of them work well, such as [15, 16]. Most of these algorithms are complex and high computational. In our model, we would extract point moving features from video sequences based on a method [17]. However, these trajectories are often highly fragmented and quite a lot of trajectories are missing because the tracker cannot keep track of all the moving points all the time. So the trajectories cannot be used directly and some pre-processing works are necessary.

Pre-process Extracted Information. The main purpose of the pre-processing procedure is to exclude outlets which result from tracking failures. So we would delete trajectories which are too short and wrongly recorded. As track captured might often shrink, the raw trajectories we get are usually slightly zigzag. So smoothing in a certain extent is also required. In our model, an arithmetic average filtering method is adopted, which is simple but effective.

Learn Global Motion Patterns. In a specific scenario, crowd behaviours actually share some common patterns based on adjacency of departure or destination positions, amplitude of velocities, similarity in motion paths or magnitude of collectiveness [18]. The paths of diverse pedestrians must be different from each other even under the circumstance of sharing the same departure and destination position. However, paths people chose should show some statistic features, which reflect the common preference of crowd flow. For example, in EWAP dataset [19] (Fig. 2), some statistic characteristics can be observed in the video, as is shown in Fig. 2 right. It implies that pedestrians tend to pass through the gallery in the middle.

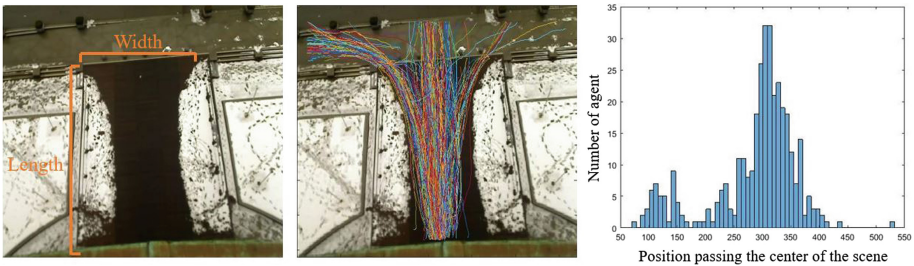


Fig. 2. Left, the background of EWAP Scenario. Middle, trajectories extracted from the video. Right, the histogram demonstrating the statistical feature of agents' position in the middle of its trajectory. (The pedestrian in the Scene of EWAP is low density. The width is about four meters and the length is about eight meters.)

In our work, we learn motion patterns called Main Streams (MS) based on adjacency and similarity of paths and use it as a global attribute for crowd simulation. Main Streams reflect the preference of crowds when they choose their paths. A trajectory of an agent is defined as the realistic path in the crowd. That is

$$\chi^p = \left(x_1^p, x_2^p, \dots, x_{np}^p \right), \quad (1)$$

where χ^p here denotes the l -th path and x_i^p denotes the i -th position recorded of the path. x_1^p and x_{np}^p here denote the departure and destination position of the path.

If there are k categories of patterns among χ , we define the set of paths sharing the same pattern as χ_i where $i = 1, 2, \dots, k$. Then our next concern is to extract each χ_i from the overall set χ . Within χ_i , each path χ_i^p should be similar with each other. So if we regard a path χ_i^p as a sample in sample space χ_i , the distance between samples within sample space should be shorter than distance between samples outside the space. That means if there exists a centre C_i in space χ_i , the distance between any sample of the space and C_i should be the shortest one compared to other centres, which can be abstracted as the following equations. For all the centres defined in χ ,

$$C = \{C_1, C_2, \dots, C_k\}, \quad (2)$$

if the distance between a path χ^p and a centre C_i satisfies

$$D_i = \min\{\|\chi^p - C_i\|, i = 1, 2, \dots, k\}, \quad (3)$$

then we define $\chi^p \in \chi_i$. As a matter of fact, it is actually a process of clustering.

For a specific scene from real world, it is difficult to judge how many categories trajectories should be clustered. So some popular clustering method such as K-means and N-cut are not ideal choices. It is essential that the model is capable of automatically judging how many categories the set should be clustered. ISODATA [20] is an iterative self-organization data analysis or unsupervised clustering algorithm without any prior knowledge. We introduce ISODATA in clustering trajectories' set and extracting patterns Main Streams (MS-ISODATA). It is explained as follows.

Step 1, sample trajectories. Before clustering, we must acquire samples of the same length. Since the length of trajectories in the set χ is varied from each other, sampling trajectories to form sequences of the same length is necessary. Firstly, find out the shortest trajectory and denote its length as Ns . Secondly, linearly divide each path χ^p into Ns parts. It equals to constructing an arithmetic progression of Ns elements from 1 to np .

$$q^p = \left\{ 1, \frac{np}{Ns}, \frac{2np}{Ns}, \dots, \frac{(Ns-1)np}{Ns}, np \right\}. \quad (4)$$

Thirdly, take one element from each parts of χ^p and form a sample of χ^p . It equals to form an index

$$index^l = floor(q^p), \quad (5)$$

by taking the nearest round number of each element of q^p in minus direction and extract point from χ^p according to this index. Then a sample of χ^p is acquired by

$$S^p = \chi^p(\text{index}^p). \quad (6)$$

Step 2, reshape each sample into one-dimension Eigenvector of the trajectory it represents.

Step 3, clustering. We apply ISODATA to cluster the trajectories set χ into categories.

Step 4, extract centre C_i of each clustered trajectories χ_i . These centres are the bases of our global motion patterns MS. They are reshaped into two-dimension position coordinate series.

$$C_i = (C_i^1, C_i^2, \dots, C_i^{N_s}). \quad (7)$$

Step 5, obtain patterns of MS. The above centers cannot be used as MS directly or reflect complete global features, because that trajectories recorded might be highly fragmented. These ISODATA clustered centers are modified and the modified results are denoted as Main Streams of crowds. The modified process takes both trends of these original centers and positions of real departure or destination into account. At $C_i^{N_s}$, we denote direction vector to destination position as \vec{v}_{des} and we compute

$$C_i^{N_s+1} = \frac{1}{\alpha} \vec{v}_{des} + \frac{1}{1-\alpha} (C_i^{N_s} - C_i^{N_s-1}) \quad (8)$$

iteratively until the modified center trajectory getting the edge of the scene. $\alpha \in (0, 1)$ here and we use 0.5 in experiments following. For the departure direction the same process is repeated. If we denote this growing process as function $f(\cdot)$ then we get

In Fig. 3, key processes of extracting Main Streams from crowd scenario, EWAP dataset are demonstrated.

$$MS_i = f(C_i). \quad (9)$$

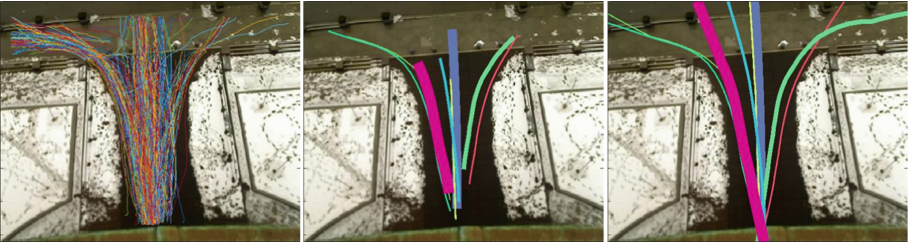


Fig. 3. Left, trajectories extracted from EWAP. Middle, clustering centers of paths. Right, MS extracted from EWAP.

2.3 Agent-Based Collision-Free Model

Paths Preference. In our hybrid model, Main Streams extracted from the real world are used as strong constraints for simulating agents when they choose appropriate paths. As illustrated in Fig. 2, agents prefer to pass through the scene in the middle and observation supports that agents' distribution among space is nearly a Gaussian distribution. Then we construct the model by defining the spatial distribution probability density function of agents as the following equation,

$$f_i(x, MS_i, \delta) = \frac{1}{\delta\sqrt{2\pi}} \left(-\frac{(x - MS_i)^2}{2\delta^2} \right). \quad (10)$$

Where δ is determined by the standard deviation in clustering. If no collision risks are detected around, agents would like to choose their paths according to Main Streams.

Collision-Free Steering Mechanism. Humans control their speed and direction based on their vision to avoid static and moving obstacles. Data-driven approaches cannot guarantee that agents are free of collision. It is necessary to add collision-free steering mechanism in crowd simulation. Collision avoidance research has drawn much attention in agent-based crowd simulation. Several approaches have been proposed to tackle the interaction and collision between agents. Helbing et al. [9] proposed the social forces model to avoid collision by repulse each other. However, Ondřej's vision-based approach (VISION) [21] performs more exactly and efficiently than RVO model and Helbing's model. Therefore, we introduce a collision-free mechanism by improving the VISION approach in our model. In this approach, velocities and directions of simulating agents can be controlled according to a threshold function

$$\tau_1(tti) = \left\{ \begin{array}{l} \tau_1(tti) = a - b \cdot tti^{-c}, \text{ if } \dot{\alpha} < 0, \\ \tau_1(tti) = a + b \cdot tti^{-c}, \text{ otherwise.} \end{array} \right\} \quad (11)$$

where τ is the threshold value controlling the angle magnitude of agents when making turns, tti is the time to interaction between two different agents, a , b and c are constant parameters, and $\dot{\alpha}$ is the relative angular acceleration between two different agents. In the original VISION, the constant parameter $a = 0$, $b = 0.6$ and $c = 1.5$. However, the author of VISION didn't mention the physical meanings of parameter a , b and c . We find that the better performance can be got during changing the value of a , b or c . In our model, we propose two different improvements with respect to the threshold function, as shown in Fig. 4.

Firstly, we reinforce the steering habit of agent by setting a suitable none-zero offset to parameter a . In experiments, we find that setting parameter a to 0 is not effective enough to fulfill completely collision free mechanism as the steering habit is not strong enough. By giving a suitable offset to a , the agent makes turns with larger magnitude.

Secondly, we restrict the range of reaction by adding cut-off effect to the threshold function. In the original approach of VISION, agents may behave unnecessary steering behaviors. Even agents without mutual collision risk would evade from each other

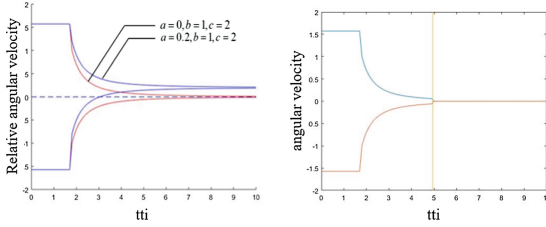


Fig. 4. Left, Reinforce the steering habit by adding an offset to the threshold function. Right, Bring cut-off effect into threshold function with cut-off threshold.

when they get closer if the reaction range is not restricted. The cut-off effect is brought in by refining (11) as follows:

$$\tau_1(tti) = \left\{ \begin{array}{l} \left\{ \begin{array}{l} \tau_1(tti) = a - b \cdot tti^{-c}, \text{ if } \dot{\alpha} < 0, \\ \tau_1(tti) = a + b \cdot tti^{-c}, \text{ otherwise.} \end{array} \right\}, \text{ if } tti \leq \tau_0, \\ 0, \text{ otherwise.} \end{array} \right\} \quad (12)$$

In (12), a threshold value of tti is brought in, which is denoted as τ_0 . If the time of interaction tti between two agents is larger than the cut-off threshold τ_0 , the steering behavior will not be activated. With the cut-off effect, unnecessary evading behaviors are avoided.

2.4 Centralized Post-process

Velocity Optimization. According to mechanism in 2.2 and 2.3, we can make sure that simulating agents move in the right ways and get rid of mutual collision. However, we still cannot guarantee simulation results to resemble the real scenario as we just define the global motion features according to real crowd for the simulation but the local velocity definition is still missing. It is necessary to optimize velocity choices of simulating agents after global paths and collision avoidance meet the requirements. Let’s assume that in example crowd video of the real scenario there are n agents and their velocities are denoted as $V_e = (v_1, v_2, \dots, v_n)$. Their velocity are transformed to angle coordinate system respectively as $v_i = |v_i|\theta_i$. Then we get a vector of $\Theta_e = (\theta_1, \theta_2, \dots, \theta_n)$. Accordingly, the vector of agents’ direction in simulation can be computed as well, which is denoted as Θ_s . We compute the cross-correlation of agents’ directions in each frame as

$$\text{cov}(\Theta_e, \Theta_s) = \sum_{i=1}^n (\Theta_e(i) - E(\Theta_e)) \cdot (\Theta_s(i) - E(\Theta_s))/n. \quad (13)$$

For each simulation epoch or frame, we make velocities magnitude of agents invariable in the basis of 2.2 and 2.3, and change directions of simulating agents within the range of $\theta_i = \pm\pi/4$ in the direction of gradient descent, using (13) as cost function. With this method of optimization, the distribution of simulating agents’ velocities

approximates the distribution of agents in real scenario, which makes simulation resemble real crowd locally in feature of motion.

3 Experimental Results

3.1 Hybrid Simulation Model Results

In the simulation, agents we reproduced succeed in passing through the scenario with the same motion pattern learned from real crowds in dataset Grand Central Hall (Grandhall) [22], in which 45,976 frames and 3,712 recorded agents are extracted. Using our model, we can automatically understand motion features of crowds in a specific real scenario and these features are successively used in crowd simulation. In the experimental results, we can see that our model successfully simulates a crowd of agents moving with global features like the real scenario, as demonstrated in Fig. 5. The simulation experiment comparison of three different approaches is demonstrated in Table 1. Approaches without collision-free mechanism, with VISION and with mechanism we propose are compared.

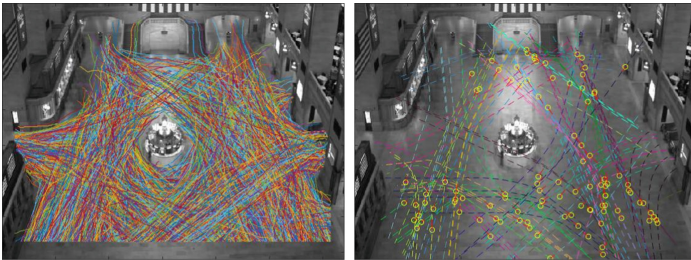


Fig. 5. Left, trajectories extracted from Grandhall. Right, one simulating frame of MS-based model.

Table 1. Comparison of experimental results based on different collision-free mechanism

Scenario	Evasion mechanism	Frames simulating	Number of agents	Times of collision	Agents reach goals
Grandhall	–	300	157	7,126	157
Grandhall	MS-VISION	300	157	0	59
Grandhall	MS-Improved VISION	300	157	0	121

The simulation results of the hybrid model can effectively mimic the behavior of the crowds on the global level, and the behavior of the simulating crowds is very similar to the real crowds' behavior. If the collision-free mechanism is not adopted, the simulation results based on the MS simulation method are developed smoothly along

the MS direction. In the simulation, however, the spatial position is sometimes overlapped by different agents. The simulation results are not real enough in details, as shown in Fig. 6B.

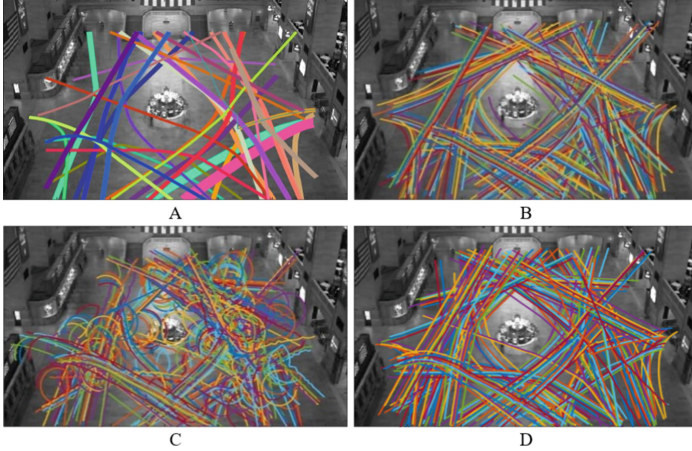


Fig. 6. A, MS extracted from grandhall dataset. B, simulation trajectories with MS. C, simulation trajectories with MS compounding VISION. D, simulation trajectories with hybrid model we proposed.

With the VISION collision-free mechanism, some agents are trapped in local oscillation due to the interference of neighboring agents. These individuals continue to make circles in some local areas and repeat searching for ideal paths. They failed to complete the simulation and reach their destinations, as shown in Fig. 6C. With our improved collision free mechanism, we can eliminate most local oscillation caused by mutual interference between individuals, so that individuals can finally reach their destinations while avoiding mutual collisions, as shown in Fig. 6D.

In this simulation experiment, we simulated 300 frames and the simulation contains 157 agents. We record times of collision occurred in the three approaches and agents finally reaching their goals within 300 simulation epochs. Fig. 6 and Table 1 show that the proposed hybrid simulation model can mimic the motion trajectories feature of real scenario and guide the agents reach their goals without collisions. In further experiments, we introduce collectiveness to measure the similarity between the simulation crowd and the real data in details.

3.2 Collectiveness Evaluation

The simulated crowd model needs a descriptor to evaluate whether its performances are similar to the real scenarios. In the past few years, some theoretic methods have been proposed to measure the similarity between the real-world data and the simulated results, such as [23, 24]. However, both of them are complex and heavy computing

burden. In order to reduce the complexity of similarity measurement, we introduce a simple method to describe the similarity of real scenarios data to crowd simulators. Collectiveness is a common characteristic used to measure the behavior of a crowd motion from the individuals' behavior. Zhou et al. proposed a collectiveness descriptor based on path similarity and velocity correlation in 2013 [18].

Figures 7 and 8 show that the collectiveness of simulating crowds created by our simulation model is very close to the collectiveness of the real crowd scenes. The experiment shows that the maximum collectiveness error is about 0.3 for the first MS, and the error is below 0.1 for another. From the results of the collectiveness comparison, our simulation results are similar to real scenarios.

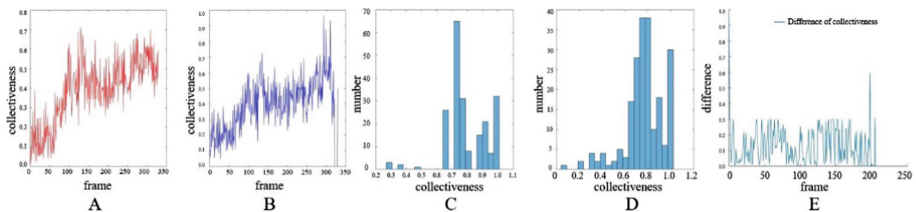


Fig. 7. A, collectiveness of one streams of agents in 330 frames of real scenario. B, collectiveness of one streams of agents in 330 frames of simulation. C, histogram of simulation result's collectiveness. D, histogram of real scenario's collectiveness. E, the difference of collectiveness between simulation and real data.

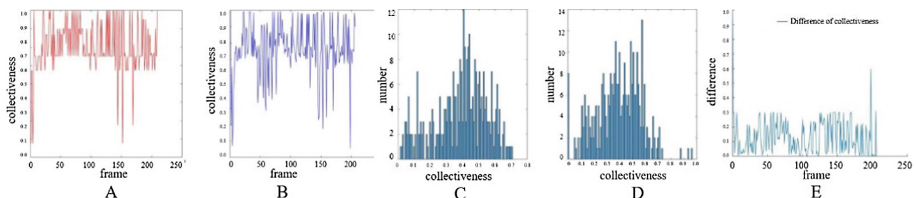


Fig. 8. A, collectiveness of one streams of agents in 215 frames of real scenario. B, collectiveness of one streams of agents in 215 frames of simulation. C, histogram of simulation result's collectiveness. D, histogram of real scenario's collectiveness. E, the difference of collectiveness between simulation and real data.

4 Conclusion

In this paper, we have presented a data-driven and collision-free hybrid simulation model to achieve crowd simulation. As demonstrated above, our model could reproduce the crowd scenarios with real motion features by using Main Streams we introduce. Meanwhile, it combines improved collision-free mechanism, which makes the simulating agents evade from local mutual collisions. In the future work, we will explore more approaches (Deep Learning etc.) to extract global features of real crowds and develop advanced steering mechanisms in crowd simulation. Meanwhile, real-time crowd simulation or crowd analysis of large number of agents will be an important part of our next research.

Acknowledgments. This work is supported by national natural science foundation of China under grants 61771145 and 61371148.

References

1. Chan, A.B., Liang, Z.S.J., Vasconcelos, N.: Privacy preserving crowd monitoring: counting people without people models or tracking. In: 26th IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–7. IEEE Press, Anchorage (2008)
2. Huang, X., Wang, W., Shen, G., Feng, X., Kong, X.: Crowd activity classification using category constrained correlated topic model. *KSII Trans. Internet Inf. Syst.* **10**, 5530–5546 (2016)
3. Bera, A., Manocha, D.: Realtime multilevel crowd tracking using reciprocal velocity obstacles. In: 22nd IEEE International Conference on Pattern Recognition, pp. 4164–4169. IEEE Press, Stockholm (2014)
4. Cui, J., Liu, W., Xing, W.: Crowd behaviors analysis and abnormal detection based on surveillance data. *J. Vis. Lang. Comput.* **25**(6), 628–636 (2014)
5. Shao, J., Loy, C.C., Wang, X.: Learning scene-independent group descriptors for crowd understanding. *IEEE Trans. Circ. Syst. Video Technol.* **27**(6), 1290–1303 (2017)
6. Martínez-Gil, F., Lozano, M., Fernández, F.: Emergent behaviors and scalability for multi-agent reinforcement learning-based pedestrian models. *Simu. Model. Pract. Theory* **74**, 117–133 (2017)
7. Junior, J.C.S.J., Musse, S.R., Jung, C.R.: Crowd analysis using computer vision techniques. *IEEE Sig. Process. Mag.* **27**(5), 66–77 (2010)
8. Reynolds, C.W.: Flocks, herds and schools: a distributed behavioral model. *ACM SIGGRAPH Comput. Graph.* **21**(4), 25–34 (1987)
9. Helbing, D., Molnar, P.: Social force model for pedestrian dynamics. *Phys. Rev. E* **51**(5), 4282 (1995)
10. Helbing, D., Farkas, I., Vicsek, T.: Simulating dynamical features of escape panic. *Nature* **407**(6803), 487–490 (2000)
11. Duan, Z., Gu, X.: Animal group behavioral model with evasion mechanism. In: IEEE International Joint Conference on Neural Networks, pp. 1167–1172. IEEE Press, Beijing (2014)
12. Lerner, A., Chrysanthou, Y., Lischinski, D.: Crowds by example. *Comput. Graph. Forum* **26**(3), 655–664 (2007)
13. Lee, K.H., Choi, M.G., Hong, Q., Lee, J.: Group behavior from video: a data-driven approach to crowd simulation. In: Proceedings of the 2007 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, pp. 109–118. ACM Press, San Diego (2007)
14. Karamouzas, I., Skinner, B., Guy, S.J.: Universal power law governing pedestrian interactions. *Phys. Rev. Lett.* **113**(23), 238701 (2014)
15. Benfold, B., Reid, I.: Stable multi-target tracking in real-time surveillance video. In: 29th IEEE Conference on Computer Vision and Pattern Recognition, pp. 3457–3464. IEEE Press, Colorado (2011)
16. Berclaz, J., Fleuret, F., Fua, P.: Multiple object tracking using flow linear programming. In: 12th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS-Winter), pp. 1–8. IEEE Press, Snowbird (2009)
17. Tomasi, C.: Detection and tracking of point features. Technical report 91(21), 9795–9802 (1991)

18. Zhou, B., Tang, X., Wang, X.: Measuring crowd collectiveness. In: 31st IEEE Conference on Computer Vision and Pattern Recognition, pp. 3049–3056. IEEE Press, Portland (2013)
19. Pellegrini, S., Ess, A., Schindler, K., Van Gool, L.: You'll never walk alone: Modeling social behavior for multi-target tracking. In: 12th IEEE International Conference on Computer Vision, pp. 261–268. Kyoto (2009)
20. Ball, G.H., Hall, D.J.: ISODATA, a novel method of data analysis and pattern classification. Stanford Research Institute, AD-699616 (1965)
21. Ondřej, J., Pettré, J., Olivier, A.H., Donikian, S.: A synthetic-vision based steering approach for crowd simulation. *ACM Trans. Graph. (TOG)* **29**(4), 123 (2010)
22. Zhou, B., Wang, X., Tang, X.: Understanding collective crowd behaviors: learning a mixture model of dynamic pedestrian-agents. In: 30th IEEE Conference on Computer Vision and Pattern Recognition, pp. 2871–2878. IEEE Press, Providence (2012)
23. Guy, S., Van Den Berg, J., Liu, W., Lau, R., Lin, M.C.: A statistical similarity measure for aggregate crowd dynamics. *ACM Trans. Graph. (TOG)* **31**(6), 190 (2012)
24. Wong, S.M., Yao, Y.Y.: A statistical similarity measure. In: Proceedings of the 10th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 3–12. ACM Press, New Orleans (1987)