

基于熵度量的人群模型评估方法^①



刘 杨, 王 雷, 盛 捷

(中国科学技术大学 信息科学技术学院, 合肥 230027)

通讯作者: 王 雷, E-mail: wangl@ustc.edu.cn

摘 要: 人群模型评估是虚拟人群仿真研究的关键问题, 现有的研究多通过个体仿真轨迹与真实轨迹之间的误差来评估人群模型. 然而人群行为本质上是复杂的随机系统, 简单的轨迹对比并不能有效反映模型能力. 本文应用熵度量的模型评估方法, 通过估计真实人群状态与仿真人群状态的误差分布实现了精确的人群仿真定量评估. 同时引入失真情况的判断和处理规则, 使得评估方法在仿真失真情况下能够保持准确性. 实验结果表明, 本文提出的算法及规则能有效地实现人群仿真模型的定量评估并给出模型参数选择的指导.

关键词: 虚拟人群; 复杂系统; 仿真模型评估; 分布估计; 熵度量

引用格式: 刘杨, 王雷, 盛捷. 基于熵度量的人群模型评估方法. 计算机系统应用, 2021, 30(11): 342-347. <http://www.c-s-a.org.cn/1003-3254/8149.html>

Evaluation Method of Crowd Model Based on Entropy Metric

LIU Yang, WANG Lei, SHENG Jie

(School of Information Science and Technology, University of Science and Technology of China, Hefei 230027, China)

Abstract: Crowd model evaluation is a key issue in virtual crowd simulation. Most studies use the error between the individual simulation trajectory and the real trajectory to evaluate the crowd model. However, crowd behavior is essentially a complex random system, and simple trajectory comparison cannot adequately reflect model capabilities. In this study, the model evaluation method based on entropy metric is used to achieve the accurate quantitative evaluation of crowd simulation by estimating the error distribution between the real crowd state and the simulated crowd state. In addition, the judgment and processing rules for distortion are introduced so that the evaluation method can be accurate under the simulated distortion. The experimental results show that the algorithm and rules proposed in this study can fully realize the quantitative evaluation of the crowd simulation model and provide guidance for the selection of model parameters.

Key words: virtual crowd; complex system; simulation model evaluation; distribution estimation; entropy metric

近年来, 随着虚拟人群仿真技术在公共安全、城市规划、智能交通等领域的广泛应用, 人群仿真模型的评估方法成为研究人员关注的问题. 人群是一个复杂系统^[1], 个体行为具有随机性, 同时个体之间又会互相影响. 众所周知, 描述复杂系统时, 不能只依赖确定的微观信息, 而应该同时考虑系统宏观特征, 所以, 评估人群仿真模型的主要挑战在于复杂系统间相似性的

全面度量.

当前模型评估算法研究主要分为定性和定量两类. 定性评估方法主要包括视觉比较^[2]和感知实验^[3]等, 但是这类方法依赖主观因素, 随着模型复杂度提升, 难以得到有效的评估结果. 定量方法分为两类: 主要有数据驱动评估方法和非参估计评估方法. 其中, 数据驱动的方法基于人群微观信息和低级特征的比较来评估模

① 基金项目: 中国科学技术大学预研基金 (YZ2101900004)

Foundation item: Preliminary Study Fund of University of Science and Technology of China (YZ2101900004)

收稿时间: 2021-01-24; 修改时间: 2021-02-23; 采用时间: 2021-03-03; csa 在线出版时间: 2021-10-22

型,如 Pettré等^[4]通过对比轨迹间的误差来评估模型准确性, Seyfried等^[5]利用速度和密度相似程度对于模型进行评估.但是数据驱动评估方法缺少对于人群运动随机性和环境噪声的综合考虑,所以评估指标缺乏全面性和鲁棒性.非参估计方法则是通过微观信息和人群低级特征进行人群运动分布的估计.如 Wang等^[6]通过建模人群路径模式来对比评估模型效果, Guy等^[7]通过估计人群运动分布,来对于误差建模分析仿真模型的效果.但是,这类方法侧重从宏观角度分析人群高级特征,过程中忽略了微观现象的处理,使得在一些仿真情况下评估效果并不理想.

针对上述问题,本文在相关研究的基础上,应用熵度量模型评估方法,提出了一种有效的人群仿真模型的定量评估算法,在宏观分析人群运动特征的同时,添加了典型微观问题的判断与处理,能够全面地对于人群仿真模型进行精确的定量评估.

1 熵度量概述

人群具有复杂系统的特性,对于某一场景下 t 时刻的人群状态,可以用符号 s_t 表示,其包含人群随时间变化的各项信息. s_t 的维度取决于 t 时刻场景中人群的数量以及人群状态信息,如位置,速度等.通常采用抽象函数 f 表示人群的演变规则,那么人群状态随时间的演变可以表示为:

$$s_{t+1} = f(s_t) \quad (1)$$

这与大多数基于连续空间的人群仿真方法类似:仿真从某一时刻人群状态开始,根据人群演变规则的建模,来得到下一时刻人群的状态信息.可以将人群仿真模型用抽象函数表示为 \hat{f} .那么,人群仿真的目标就是期望 \hat{f} 能够近似 f ,即:

$$\hat{f}(s_t) \approx f(s_t) = s_{t+1} \quad (2)$$

基于上述分析,人群仿真模型的评估问题可以被定义为:给定真实人群状态 s_t ,如何衡量不同模型 \hat{f} 预测的人群状态与真实人群状态 s_{t+1} 的接近程度.

模型评估问题主要存在两点挑战:(1)人群属于复杂系统,状态变化具有随机性,某一时刻人群的真实状态 s_t 并不唯一.(2)人群真实状态信息获取困难.常见的人群数据来源于视频数据集,但是视频数据集不能完整反映人群真实状态.

熵度量方法从信息论角度出发解决了上述问题.

与数据驱动方法直接对比人群信息不同,熵度量方法认为真实人群状态和仿真人群状态分别服从某种分布,通过估计两者间的误差分布来评估模型.即在式(2)的基础上,定义单一时间步上的模型误差为 e_t ,可得:

$$s_{t+1} = f(s_t) = \hat{f}(s_t) + e_t \quad (3)$$

假设整个仿真时间段上的 e_t 符合误差分布 ε ,即 $e_t \in \varepsilon$.熵度量方法通过衡量误差的分布给出模型的评估结果.

在估计真实人群状态分布时,熵度量方法将真实人群数据集作为观测数据,记为 z_1, z_2, \dots, z_t .假设真实人群状态 s_t 和 z_t 的关系由函数 h 给出,则有:

$$z_t = h(s_t) + q_t \quad (4)$$

其中, q_t 表示数据中的噪声.熵度量方法采用卡尔曼滤波的相关方法估计每个时间步的真实人群状态分布,将仿真任务分解.然后利用估计的真实人群状态分布来进行熵度量评估.

可以看到,熵度量通过估计状态分布从宏观角度解决了人群仿真模型的评估问题,但是,上述的熵度量方法仍然存在局限性.完全依赖宏观角度的评价,使得熵度量方法在评估模型时忽略了一些可能的微观失真情况.

微观失真情况的研究来源于多种人群仿真实验的观察.在一些仿真任务中,当仿真模型或者模型参数的选择不合适时,仿真过程会出现失真的情况,如个体抖动和异常静止.图1给出了失真情况的示意.图中是个体穿越的不同大小路口的仿真运动轨迹,ped1, ped2, ped3自上而下,ped4自下而上.图1(a)中路口为4m时,个体轨迹正常,当图1(b)中路口变为2m时,个体轨迹出现了失真情况,ped1与ped4出现了抖动,ped2和ped3出现了异常静止.

失真情况会严重影响仿真的效果,甚至导致仿真过程无法正常进行,如图1中的异常静止现象使得个体无法到达目的位置.而原始熵度量方法对于失真情况并不敏感,因为将仿真任务分解的步骤淡化了失真情况对于评估结果的影响.所以,本文创新性地的人群状态估计的过程中添加了失真情况的判断与处理规则,并对于失真情况进行合理惩罚,使得熵度量方法在宏观评价模型时能够合理综合微观信息.同时,本文方法对于失真情况的程度也能给出准确的评价.

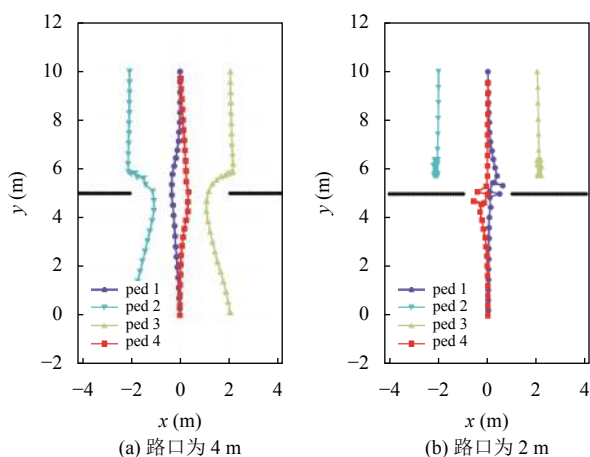


图1 仿真失真情况示意

2 评估算法描述

本文提出的评估算法主要分为如下几步: 首先, 根据人群的观测数据估计真实人群状态信息, 将仿真任务分解为单个时间步. 在这个过程中, 采用失真判断规则来判断失真情况并记录惩罚. 其次, 对于每个时间步的任务进行仿真以估算误差分布的影响. 然后, 通过期望极大算法 (EM) 算法^[8] 最大程度接近估计的真实人群状态得到误差分布的熵度量大小. 最后, 将误差分布的熵度量结果与失真情况的惩罚进行综合, 最终的结果反映了仿真模型重现真实人群状态的能力.

在估计真实人群状态时, 本文采用了集合卡尔曼平滑 (EnKS) 的方法^[9]. EnKS 大致分为两个步骤, 首先根据仿真模型得到下一个时间步的状态 $\hat{f}(s_t)$, 然后利用观测数据来修正预测状态. 计算时, 在每个时间步上对于给定的仿真模型进行 n 次仿真, 每次仿真过程中使用高斯模型进行噪声的建模. 在先前相关研究中, Karamouzas 等人^[10] 的工作证明了选用高斯模型建模噪声的合理性. 此外, 为了在分步估计状态时保留微观失真情况的信息, 对于待评估模型会先行进行仿真以获得完整的仿真数据作为微观参照数据. 算法 1 给出了伪代码实现.

从算法 1 可以看出, 算法中添加了微观失真情况的判断. 本文方法主要关注两种失真情况: 抖动和异常静止. 失真情况借助仿真数据的上下文速度信息以及人群观测数据的状态信息进行判断. 仿真人群的抖动主要来源于个体速度方向的反复剧烈变化, 因此对于每个时间步的仿真个体, 算法重点考察其相邻时间步

的速度信息. 抖动情况的判定规则为: 如果仿真个体在当前时间步与前后时间步的速度夹角均大于 90° , 将其判定为发生了抖动, 然后将此步的误差记为:

$$d(x_k, \bar{z}_k) = \bar{z}_{k+1} - x_k \quad (5)$$

即用下一时间步模型预测的观测状态信息与此步参照信息的差作为惩罚, 等同于认为此步仿真失效, 对于严重影响仿真效果的失真情况给出足够的惩罚. 异常静止的惩罚与上述抖动情况相同, 不同之处在于判断规则. 异常静止的判断规则为: 以人群观测数据的速度信息作为参照, 如果仿真轨迹在当前时间步以及前后时间步均出现速度异常为零的情况, 将其判定为发生了异常静止.

算法 1. 包含失真判断的人群状态估计算法

输入: 人群观测数据 z_1, \dots, z_t , 人群仿真模型 \hat{f} , 误差估计方差 E , 完整参照数据 x_1, \dots, x_t , 失真惩罚函数 d
输出: 真实人群状态估计 s_1, \dots, s_t , 失真情况惩罚值 l_d

```

1. foreach  $k \in 1, \dots, t$  do
2.   foreach  $i \in 1, \dots, n$  do //状态预测
3.      $\hat{s}_k^{(i)} = \hat{f}(s_{k-1}^{(i)}) + e_{k-1}^{(i)}$ ,  $e_{k-1}^{(i)} \in \mathcal{E}$ 
4.      $\hat{z}_k^{(i)} = h(\hat{s}_k^{(i)}) + q_k^{(i)}$ 
5.      $\bar{z}_k = \frac{1}{n} \sum_{i=1}^n \hat{z}_k^{(i)}$ 
6.     if 判定失真:
7.        $l_d += d(x_k, \bar{z}_k)$  //记录失真惩罚
8.        $Z_k = \frac{1}{n} \sum_{i=1}^n (\hat{z}_k^{(i)} - \bar{z}_k)(\hat{z}_k^{(i)} - \bar{z}_k)^T$ 
9.     foreach  $j \in 1, \dots, k$  do //状态修正
10.       $\bar{s}_j = \frac{1}{n} \sum_{i=1}^n \hat{s}_j^{(i)}$ 
11.       $\Sigma_j = \frac{1}{n} \sum_{i=1}^n (\hat{s}_j^{(i)} - \bar{s}_j)(\hat{s}_j^{(i)} - \bar{s}_j)^T$ 
12.      foreach  $i \in 1, \dots, n$  do
13.         $\hat{s}_j^{(i)} = \hat{s}_j^{(i)} + \Sigma_j Z_k^{-1} (z_k - \bar{z}_k^{(i)})$ 
    
```

在通过算法 1 得到真实人群状态分布后, 可以估计噪声方差 E , 本文采用最大似然估计的方法, 来最大化仿真数据重现真实人群状态的可能性. 伪代码如算法 2 所示.

值得注意的是, 算法 1 与算法 2 出现了循环依赖问题: 估计人群真实状态, 需要误差估计方差, 而估计误差方差, 又需要知道真实人群状态. 为了解决此问题, 本文采用了 EM 算法, 通过迭代的方式交替使用算法 1 和算法 2, 直至结果收敛. 关于 EM 算法的收敛性证明可以参考 McLachlan 等人^[11] 的研究.

算法 2. 误差分布方差估计算法

输入: 真实人群状态估计 s_1, \dots, s_t , 人群仿真模型 \hat{f} , 失真情况惩罚值 l_d
输出: 误差估计方差 E

```

1.  $E = 0$ 
    
```



```

2. foreach k∈0,⋯,t-1 do
3.   foreach i∈1,⋯,n do
4.     foreach j∈1,⋯,m do
5.       E+= $(\hat{s}_{k+1}^{(i)}[j]-\hat{f}(s_k^{(i)})[j])(\hat{s}_{k+1}^{(i)}[j]-\hat{f}(s_k^{(i)})[j])^T$ 
6. E+= $I_d I_d^T$ 
7. E|=tmn
    
```

最终,得到收敛的误差方差后,利用下式测算误差分布大小:

$$e(\varepsilon) = \frac{1}{2} \ln((2\pi e)^d \det(E)) \quad (6)$$

其中, d 是个体状态维数. 式 (6) 给出了熵度量的最终结果.

3 实验分析

本文选取 SFM (Social Force Model) 模型^[12]、RVO (Reciprocal Velocity Obstacles) 模型^[13]、以及 ORCA (Optimal Reciprocal Collision Avoidance) 模型^[14] 检验上述熵度量评估方法. 这 3 种模型是目前应用广泛的经典人群仿真模型, 有很多变体和拓展应用.

为了全面地分析评估方法的效果, 实验选取了几种典型的仿真场景. 这些场景代表不同人群运动情况, 分别为: 1) 交换场景: 反映了两个个体交换彼此的位置, 记为 swap-1. 同时还有 swap-1 的镜像场景, 用 swap-2 表示; 2) 街道场景: 反映了街道上购物的人群运动, 记为 street; 3) 狭窄路口场景: 反映了相向通过狭窄路口的人群运动, 用 passage 表示.

作为对比, 本文选取两种基于数据驱动的评估方法. 分别为: 1) L2 距离方法^[15]. 通过计算模型预测轨迹与视频数据集轨迹的距离误差来评估仿真模型; 2) 最近邻距离方法^[16]. 通过计算个体与最近邻个体的距离来评估仿真模型.

实验一对比了本文方法与上述方法的评估效果. 在 swap-1, swap-2, street 三种场景下, 分别使用社会力模型, RVO 模型以及 ORCA 模型进行人群仿真实验, 然后利用本文方法与上述两种方法对于仿真效果进行评估. 实验结果如表 1 所示, 3 种评估方法都是结果越小表示模型的仿真效果越好. 在实验一中, swap-1 与 swap-2 为镜像场景. 对于个体交换的场景, 原始路径与镜像路径都应该是个体正确的避让选择, 所以模型在 swap-1 和 swap-2 上的评估结果应该尽可能接近.

从表 1 中可以看出, L2 距离方法在镜像场景的表现较差, 相对于 swap-1 场景, swap-2 场景下的评估结

果明显升高, 同时不同模型的差异性也有所下降. 说明 L2 距离方法局限于固定的状态信息, 不能宏观地评估模型的能力. 而最近邻距离方法由于只考查个体与最近个体的距离, 所以在镜像实验下保持了结果的一致性. 但是在较为复杂的 street 场景下的评估效果急剧下降, 这可能与评估依据较为单一有关. 本文提出的熵度量方法在镜像场景中保持了较好的相似性以及区分度. 这说明熵度量方法在一定程度上反映了仿真模型捕获真实人群行为模式的能力. 同时, 熵度量方法在 street 场景下也保持了较好的评估效果. 评估结果区分度明显而且与另外两种经典方法给出的模型排名一致.

表 1 评估方法对比结果

方法	SFM	RVO	ORCA
swap-1 (L2)	0.77	0.53	0.18
swap-2 (L2)	0.83	0.70	0.38
street (L2)	0.41	0.58	0.17
swap-1 (最近邻)	0.57	0.45	0.16
swap-2 (最近邻)	0.57	0.45	0.16
street (最近邻)	0.77	0.98	0.75
swap-1 (本文)	0.98	0.59	0.18
swap-2 (本文)	1.03	0.63	0.24
street (本文)	1.99	2.36	1.09

为了反映本文方法与先前熵度量研究的区别, 实验二针对微观失真情况下的模型评估进行了实验. 在 passage 场景下, 选取了两种不同参数的社会力模型进行仿真, 模型分别记为 SFM-1 和 SFM-2. 两种模型在仿真过程中都出现了不同程度的失真现象. 图 2 给出了仿真实验的人群轨迹. 最后, 分别采用 Guy^[7] 的熵度量方法与本文方法对于仿真结果进行评估, 实验结果见表 2.

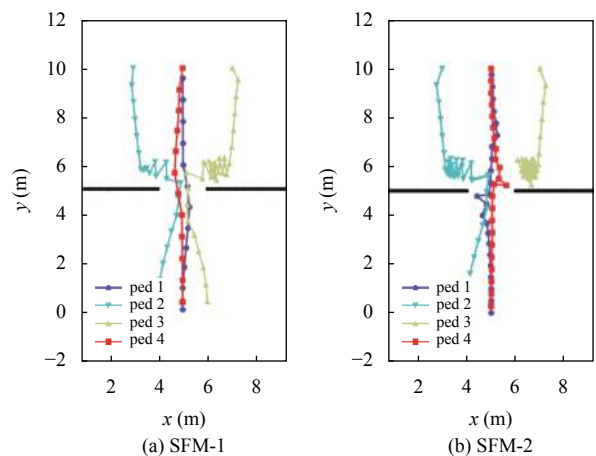


图 2 失真情况评估实验轨迹示意

表2 失真情况评估结果

熵度量	SFM-1	SFM-2
原始熵度量	0.79	0.96
本文熵度量	1.27	2.47

从图2可以看出,实验选取的两种不同参数的社会力模型都出现了明显的失真情况,但是两者的失真程度不同. SFM-1中的个体在出现抖动情况后,很快脱离了异常并到达了目标点,而 SFM-2中的个体抖动情况严重,甚至出现了静止的情况,导致部分个体并没有成功到达目标点.表2中,与原始熵度量方法的结果相比,在添加了失真情况的判断与处理后,本文方法能更为显著的区分两种模型参数组合的优劣.

此外,本文方法还可以对于模型的参数选择进行指导,实验三中给出了验证.在street场景下,参照Xue等^[17]的研究选取了ORCA模型的3种参数组合,分别对应不同规模人群仿真任务的推荐模型参数.表3中给出了3种参数组合的具体信息,Min, Max和Default分别对应小规模人群仿真,大型人群仿真以及中等人群仿真的推荐参数.实验采用上述3组参数进行仿真,并使用本文方法进行评估.实验结果如表4所示.

表3 3种模型参数组合

模型参数	Min	Max	Default
NeighborDist (m)	3	30	15
MaxNeighbors (n/a)	1	100	10
TimeHorizon (s)	1	30	10
Radius (m)	0.3	2.0	1.0
PrefVelocity (m/s)	1.2	2.2	1.45

表4 3种参数组合的评估结果

方法	ORCA-Min	ORCA-Max	ORCA-Default
本文方法	1.43	2.62	2.10

从表4可以看出,本文方法的评估结果表明,ORCA-Min的参数组合的仿真效果最好,ORCA-Default效果次之,而ORCA-Max的效果最差.这符合文献[17]中的模型参数分析,证明了本文方法在模型参数选择时能提供有效的指导.

4 结论与展望

本文针对人群仿真模型难以精确评估的问题,应用熵度量评估方法,采用合理的途径、规则及算法,实现了精确的人群仿真模型定量评估,在宏观对比

人群特征的同时引入了微观失真情况的判断与惩罚规则.通过实验验证,本文提出的算法可以对于人群仿真模型进行合理而准确的评估,同时也可以为模型参数的选择给出有价值的指导.下一步将考虑更为精确的微观信息参照方法以及完整的模型评估框架开发.

参考文献

- Bellomo N, Piccoli B, Tosin A. Modeling crowd dynamics from a complex system viewpoint. *Mathematical Models and Methods in Applied Sciences*, 2012, 22(S2): 1230004. [doi: 10.1142/S0218202512300049]
- Kim S, Guy SJ, Manocha D. Velocity-based modeling of physical interactions in multi-agent simulations. *Proceedings of the 12th ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. New York: ACM, 2013. 125–133.
- Ennis C, Peters C, O'Sullivan C. Perceptual effects of scene context and viewpoint for virtual pedestrian crowds. *ACM Transactions on Applied Perception*, 2011, 8(2): 1–22. [doi: 10.1145/1870076.1870078]
- Petré J, Ondřej J, Olivier AH, *et al.* Experiment-based modeling, simulation and validation of interactions between virtual walkers. *Proceedings of the 2009 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. New York: ACM, 2009. 189–198.
- Seyfried A, Steffen B, Winkens A, *et al.* Empirical data for pedestrian flow through bottlenecks. *Traffic and Granular Flow'07*. Berlin Heidelberg: Springer, 2009. 189–199.
- Wang H, Ondřej J, O'Sullivan C. Trending paths: A new semantic-level metric for comparing simulated and real crowd data. *IEEE Transactions on Visualization and Computer Graphics*, 2017, 23(5): 1454–1464. [doi: 10.1109/TVCG.2016.2642963]
- Guy SJ, van den Berg J, Liu WX, *et al.* A statistical similarity measure for aggregate crowd dynamics. *ACM Transactions on Graphics (TOG)*, 2012, 31(6): 1–11. [doi: 10.1145/2366145.2366209]
- Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 1977, 39(1): 1–22. [doi: 10.1111/j.2517-6161.1977.tb01600.x]
- Evensen G. The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, 2003, 53(4): 343–367. [doi: 10.1007/s10236-003-0036-9]

- 10 Karamouzas I, Sohre N, Hu R, *et al.* Crowd space: A predictive crowd analysis technique. *ACM Transactions on Graphics (TOG)*, 2019, 37(6): 1–14. [doi: [10.1145/3272127.3275079](https://doi.org/10.1145/3272127.3275079)]
- 11 McLachlan GJ, Krishnan T. *The EM Algorithm and Extensions*. New York: John Wiley & Sons, 2007.
- 12 Helbing D, Molnár P. Social force model for pedestrian dynamics. *Physical Review E*, 1995, 51(5): 4282–4286. [doi: [10.1103/PhysRevE.51.4282](https://doi.org/10.1103/PhysRevE.51.4282)]
- 13 van den Berg J, Lin M, Manocha D. Reciprocal velocity obstacles for real-time multi-agent navigation. 2008 IEEE International Conference on Robotics and Automation. Pasadena: IEEE, 2008. 1928–1935.
- 14 van den Berg J, Guy SJ, Lin M, *et al.* Reciprocal n -body collision avoidance. In: Pradalier C, Siegwart R, Hirzinger G, eds. *Robotics Research*. Berlin Heidelberg: Springer, 2011.3–19. [doi: [10.1007/978-3-642-19457-3_1](https://doi.org/10.1007/978-3-642-19457-3_1)]
- 15 Lerner A, Chrysanthou Y, Shamir A, *et al.* Context-dependent crowd evaluation. *Computer Graphics Forum*, 2010, 29(7): 2197–2206. [doi: [10.1111/j.1467-8659.2010.01808.x](https://doi.org/10.1111/j.1467-8659.2010.01808.x)]
- 16 Charalambous P, Karamouzas I, Guy SJ, *et al.* A data-driven framework for visual crowd analysis. *Computer Graphics Forum*, 2014, 33(7): 41–50. [doi: [10.1111/cgf.12472](https://doi.org/10.1111/cgf.12472)]
- 17 Xue ZX, Dong Q, Fan XT, *et al.* Fuzzy logic-based model that incorporates personality traits for heterogeneous pedestrians. *Symmetry*, 2017, 9(10): 239. [doi: [10.3390/sym9100239](https://doi.org/10.3390/sym9100239)]