Evaluation of Crowd Model Suitability for Mobile Robot Simulation*

Rio Nishida[1] and Yuka Kato[2]

Abstract— **Studies on autonomous mobile robot navigation frequently employ crowd models within simulations to study robot movement in crowded environments. While numerous crowd models exist, each possesses distinct strengths and weaknesses, making no single model universally applicable. Consequently, selecting an appropriate model for a specific scenario presents a significant challenge. To address this issue, our research focuses on a method to classify simulation environments into categories and select a suitable crowd model based on this classification. As a key component of this approach, this paper proposes a methodology to evaluate the suitability of crowd models. This involves comparing pedestrian movement trajectories extracted from real-world datasets with those generated by crowd simulations conducted in equivalent virtual environments. Specifically, we represent the variations within sets of movement trajectories as probability distributions. The similarity between the distribution derived from the real-world dataset and the distribution generated by the simulation serves as the evaluation metric. Furthermore, we present evaluation experiments using real-world datasets and a crowd simulator to validate the effectiveness of the proposed evaluation approach.**

I. INTRODUCTION

Studies on autonomous mobile robot navigation frequently employ crowd models within simulations to study robot movement in crowded environments [1]. Crowd models simulate the behavior and movement of large groups of people. They are being actively studied and developed for various applications beyond robotics, such as computer graphics, games, and spatial layout. However, no single crowd model is universally applicable to all scenarios, making it difficult to select the suitable model for a given environment [2]–[4].

To address this problem, we are currently studying a method to classify simulation environments into multiple categories and select a suitable crowd model based on the classification results [5], [6]. Achieving this requires metrics and methods to evaluate crowd model suitability. To meet this need, our previous work has proposed methods for comparing real-world pedestrian trajectory datasets against simulation results from equivalent environments. These comparison methods utilize both analyses of visual trajectory shapes and Dynamic Time Warping (DTW), a technique for time-series data comparison [7]. Regarding DTW, we extracted one representative trajectory each from the dataset and simulation results, and calculated the similarity between these two trajectories. However, challenges remained as it was difficult to select appropriate representative trajectories, and it was challenging to properly express the movement tendencies of crowds, which consist of collections of multiple trajectories.

To overcome this issue, we propose a novel quantitative metric for evaluating overall crowd movement tendencies. In this approach, we represent the degree of variation among multiple trajectories as a probability distribution, and use the similarity between distributions derived from datasets and simulation results as the evaluation metric. Specifically, we employ Gaussian Mixture Models (GMMs) and discrete distributions based on sample values to represent the probability distributions and compare the similarity between them. Kullback-Leibler (KL) divergence is utilized for this comparison. Furthermore, we conduct evaluation experiments using datasets and crowd simulators to validate the effectiveness of the proposed method. The contributions of this paper are:

- Propose a novel quantitative metric for evaluating overall crowd movement tendencies.
- Validate the effectiveness of the proposed method by conducting evaluation experiments using real-world datasets.

II. RELATED WORK

A. Crowd Model for Robot Simulation

Crowd models are techniques for modeling the behavior of numerous individuals or the movement of objects, utilized in various fields such as urban traffic management and disaster evacuation guidance. They are generally classified into two types: macroscopic simulations, which treat the crowd as a single large flow, and microscopic simulations, which treat it as a collection of individuals. However, models integrated into robot simulations are typically microscopic simulations. Microscopic simulations are often further categorized into three types: Force-based models, Velocity Object (VO)-based models, and vision-based models.

In Force-based models, agents use distance-based functions to avoid collisions with surrounding agents and obstacles. A typical example is the Social Force Model (SFM) [8], [9]. In VO-based models, agents compute collision-free velocities using optimization or sampling techniques. A well-known example is Optimal Reciprocal Collision Avoidance (ORCA) [10]. In vision-based models, agents generate their behavior based on visual information, such as camera images. One example is Vision-Based Navigation (VBN) [11]. Each of these models has its own strengths and weaknesses, and no single model is universally suitable for all scenarios.

*This work was partially supported by JSPS KAKENHI (23K11087, 23K11073), the Telecommunications Advancement Foundation, the Cooperative Research Project Program of RIEC, Tohoku University

[1]Rio Nishida was with Graduate School of Science, Tokyo Woman's Christian University, Tokyo, Japan. d23m203@cis.twcu.ac.jp

[2]Yuka Kato is with Division of Mathematical Sciences, Tokyo Woman's Christian University, Tokyo, Japan. yuka@cis.twcu.ac.jp

B. Crowd model Evaluation

Consequently, studies investigating how to select an appropriate crowd model for a given scenario have been conducted [2], and they require evaluation metrics to determine the suitability of these models. Such metrics are generally divided into qualitative metrics (which assess visualized results like agent behavior and trajectories) and quantitative metrics (which evaluate simulation accuracy numerically) [12].

Common quantitative metrics include Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). For example, one study uses MAE and RMSE for crowd density estimation [13]. Research proposing novel quantitative metrics also exists; for instance, one study introduced the Quality Function (*QF*), a quantitative index designed to evaluate the naturalness of pedestrian trajectories [14]. This index is constructed by listing trajectory features that potentially affect crowd animation quality and mapping them to a single quality score.

Although it would be possible to utilize these metrics in this paper, evaluation metrics should be chosen based on the specific evaluation objectives. The aforementioned metrics primarily aim to evaluate factors such as naturalness in crowd animations, making their focus different from that of this paper, which targets applications within robot simulators.

C. Previous Work

In our previous work, we proposed a method for comparing pedestrian trajectory datasets with the results of crowd simulations conducted in equivalent environments, using both the visual shapes of the trajectories and DTW [7]. DTW is a technique for calculating the similarity between two time-series data sequences, which may differ in length or variation. It achieves this by finding corresponding points between the two sequences and computing a cumulative cost based on a problem-specific cost function. Our prior study proposed calculating DTW by selecting a single representative trajectory from the dataset and another from the simulation results. This representative trajectory was chosen from all trajectories moving in the same general direction as the one exhibiting the median value (or position) at the point of crossing a predefined line segment within the area.

However, the movement tendency of a crowd in simulations is inherently characterized by a set of multiple trajectories spread over a certain range. Representing this collective behavior with only a single, specific trajectory, as done in our previous work, hindered appropriate evaluation. This paper attempts to overcome this limitation.

III. Method

A. Outline

In the proposed evaluation method, we assess the suitability of crowd models by comparing data from real-world datasets with results obtained from crowd simulations. Here, sets of trajectories are represented using a specific feature representation, and the similarity between these representations is compared. Specifically, sets of trajectories, which are inherently distributed as data spread within the target environment, are represented as probability distributions. The process flow is outlined below:

1) Prepare the target dataset.
2) Perform crowd simulation in an environment similar to that of the dataset.
3) Generate feature representations corresponding to both the dataset and the results of the crowd simulation.
4) Evaluate the crowd model by comparing the similarity between the feature representations.

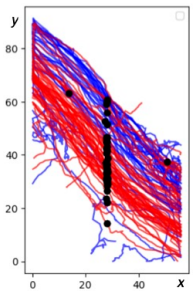

Fig. 1: Image of the one-dimensional data acquisition. Data points used are indicated by solid black dots. For each trajectory, the observation data point closest to the midpoint line segment ($x = 28$) is selected. Leftward-moving ones are shown in red, while rightward ones are shown in blue.

B. Feature Representation Generation Method

In the proposed method, a set of movement trajectories is represented as a probability distribution. From both the dataset and the simulation results, it is possible to obtain the frequency counts of locations where pedestrians are present within the observation time (i.e., the number of pedestrians who stayed at or passed through each location). This frequency distribution is modeled as a probability distribution. Here, two types of models are used: (i) Gaussian Mixture Model (GMM); (ii) a discrete distribution based on sample values (Discrete Model). Additionally, two types of data dimensionality are employed: (i) two-dimensional data, using data from the entire area; (ii) one-dimensional data, using only the data corresponding to passage through a line segment at the midpoint of the area. An image of the one-dimensional data acquisition is shown in Fig. 1.

Note that the GMM is employed to account for the direction of pedestrian movement. It is generated by first calculating the mean and variance of sample values for each movement direction to create individual Gaussian distributions. A weighted sum of these generated distributions then forms the GMM. The discrete distribution, on the other hand, is generated by aggregating the frequency counts per grid cell. The grid size is assumed to be specified beforehand.

The GMM is defined by the following equation:

$$p(\boldsymbol{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\boldsymbol{x}|\mu_k, \Sigma_k), \quad \sum_{k=1}^{K} \pi_k = 1 \tag{1}$$

where $\mathcal{N}(\cdot)$ denotes the Gaussian distribution, $\boldsymbol{x}$ is the feature vector, k is the index for the direction of pedestrian movement, K is the total number of directions, μ_k, Σ_k are

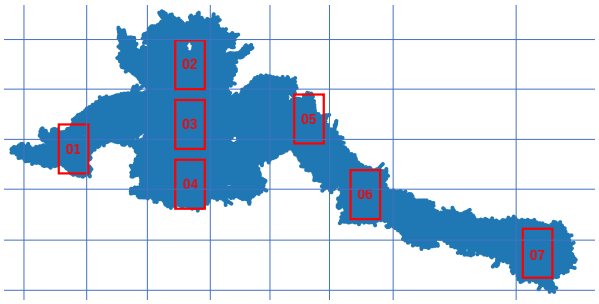

Fig. 2: Overall shape of the shopping mall and the candidate areas selected as the target environment. The data within Area 06 are used in this experiment [7].

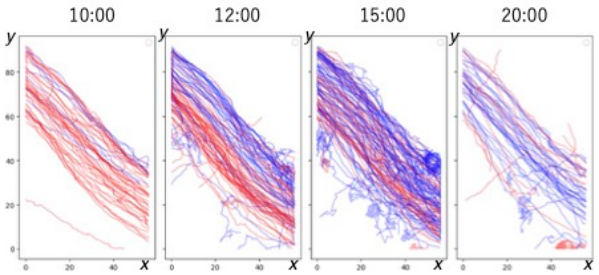

Fig. 3: Visualization result of the trajectories (data in the dataset) used in the experiment. Leftward-moving trajectories are displayed in red, and rightward-moving trajectories are displayed in blue. This data serves as the ground truth.

the mean and covariance of data belonging to movement direction k, and π_k represents the mixture weight corresponding to movement direction k, satisfying $0 \leq \pi_k \leq 1$.

Subsequently, each feature representation will be denoted by the following symbols:

- **GMM1**: Gaussian Mixture Model with 1D data
- **GMM2**: Gaussian Mixture Model with 2D data
- **DM1**: Discrete Model with 1D data
- **DM2**: Discrete Model with 2D data

C. Similarity Comparison Method

We consider a method to compare the similarity between datasets using the feature representations defined in the previous section. This method utilizes the Kullback-Leibler (KL) divergence. KL divergence is a measure of the information loss between two probability distributions; lower values indicate higher similarity between the distributions. Specifically, the similarity is calculated using the following formulas: for the GMM cases (**GMM1**, **GMM2**) is

$$\mathcal{D}_{\mathrm{KL}}(p||q) = -\int p(\boldsymbol{x}) \ln \frac{q(\boldsymbol{x})}{p(\boldsymbol{x})} dx \tag{2}$$

and for the discrete distribution cases (**DM1**, **DM2**) is

$$\mathcal{D}_{\mathrm{KL}}(P||Q) = -\sum_{s \in \mathrm{M}} P(s) \ln \frac{Q(s)}{P(s)}. \tag{3}$$

Here, $q(\boldsymbol{x})$ and $p(\boldsymbol{x})$ are the GMMs generated from the dataset and the simulation results. $Q(s)$ and $P(s)$ are the discrete distributions generated from the dataset and the simulation results.

IV. EXPERIMETAL SETUP

A. Outline

We verify the effectiveness of the proposed measure through experiments using a pedestrian movement trajectory dataset. Here, the experimental scenario is based on the data acquisition environment of the dataset, and we conduct crowd simulations designed to replicate this environment. Evaluation metrics are then calculated from the obtained results. For the crowd models, similar to the study [7], we employ the SFM and an Agent-Based Model (ABM). The validity of the evaluation using our proposed method is assessed by comparing our results with those from the study [7], which involved visual evaluation and DTW.

B. Dataset

We use the ATC pedestrian tracking dataset [15] collected at a shopping mall in Japan[1]. The dataset values are:

- Date: Wed. and Sun. from 10/24/2012 to 11/29/2013
- Items: time [ms], person ID, position (x, y, z) [mm], velocity [mm/s], motion angle [rad], facing angle [rad]

Fig. 2 shows the overall shape of the shopping mall (visualized via a scatter plot of pedestrian positions) and candidate areas for the target environment. From the data within Area 06, trajectories observed on 10/28/2012 were extracted for use in this experiment. Here, data from the 3-minute period between 12:01 and 12:03 were extracted. Fig. 3 shows a visualization of these trajectories. Leftward-moving ones are displayed in red, and rightward ones are displayed in blue. This data serves as the ground truth.

C. Crowd Model

1) SFM: In the SFM, each agent is modeled as a particle. It is assumed that physical and psychological repulsive forces act between agents, and between agents and walls (obstacles). The agent's direction of movement is determined by the resultant force, which is the sum of these repulsive forces and an attractive force toward the destination. The change in velocity at time t is calculated using the following equation of motion:

$$m_i \frac{d\boldsymbol{v}_i(t)}{dt} = m_i \frac{\boldsymbol{v}_i^0(t) - \boldsymbol{v}_i(t)}{\tau_i} + \sum_{j(\neq i)} \boldsymbol{f}_{ij} + \sum_w \boldsymbol{f}_{iw} \tag{4}$$

where m_i is the mass of agent i, $\boldsymbol{v}_i^0(t)$ is the desired velocity, $\boldsymbol{v}_i(t)$ is the actual velocity, and τ_i is some duration. Each agent attempts to maintain distance from other agents j and walls w, and is subject to the following interaction forces $\boldsymbol{f}_{ij}$ (between agents) and $\boldsymbol{f}_{iw}$ (between agents and walls):

$$\begin{aligned} \boldsymbol{f}_{ij} &= \left\{ A_i \exp\left(\frac{r_{ij} - d_{ij}}{B_i} \right) + kg(r_{ij} - d_{ij}) \right\} \boldsymbol{n}_{ij} \\ &\quad + \kappa g(r_{ij} - d_{ij}) \Delta v_{ji}^t \boldsymbol{t}_{ij}, \end{aligned} \tag{5}$$

$$\begin{aligned} \boldsymbol{f}_{iw} &= \left\{ A_i \exp\left(\frac{r_{iw} - d_{iw}}{B_i} \right) + kg(r_{iw} - d_{iw}) \right\} \boldsymbol{n}_{iw} \\ &\quad - \kappa g(r_{iw} - d_{iw}) (\boldsymbol{v}_i \cdot \boldsymbol{t}_{iw}) \boldsymbol{t}_{iw} \end{aligned} \tag{6}$$

[1]Datasets from the ATC shopping center: https://dil.atr.jp/crest2010_HRI/ATC_dataset/

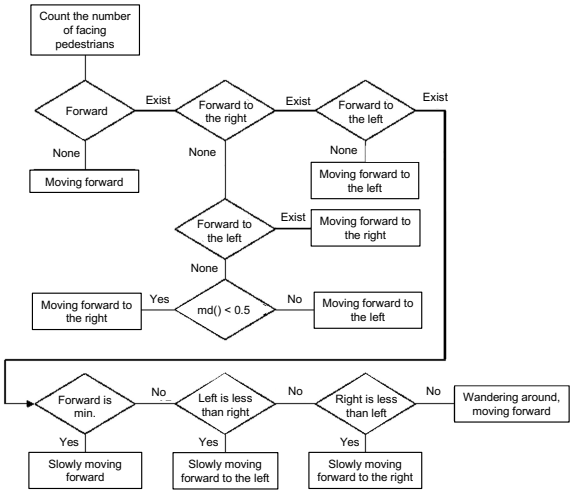

Fig. 4: Pedestrian behavior rule in ABM. Reconfigured by the authors based on materials from cited in the footnote [7].

where A_i and B_i are constants, r_{ij} is the sum of the radii of the bodies of agent i and j, d_{ij} is the distance between the agents' center of mass, $\boldsymbol{n}_{ij}$ is the unit vector pointing from agent j to i, k is a constant, and $g(x)$ is the function that returns 0 if the agents do not contact with each other and returns x otherwise. $\kappa g(r_{ij} - d_{ij})\Delta v_{ji}^t \boldsymbol{t}_{ij}$ represents the friction where κ is a constant, Δv_{ji}^t is the tangential velocity difference, and $\boldsymbol{t}_{ij}$ is the unit vector orthogonal to $\boldsymbol{n}_{ij}$.

In this model, the values of A_i and B_i determine the magnitude of the SFM potential and the range over which interactions occur. In this paper, for simplicity, we redefine these as constants v_0 and σ, respectively, which are independent of agent i. We then conduct experiments using two types of models based on different values for these constants. Each model is defined as follows:

- SFM1: $\sigma = 2.1,\ \ v_0 = 0.3$
- SFM2: $\sigma = 10.5,\ \ v_0 = 0.9$

We use the Deep Social Force simulator[2] to implement the model [16].

2) ABM: In an ABM, each pedestrian generates actions according to its own rules and interactions with its surroundings. The behavior of the entire system is formed based on the interaction between pedestrians. In this paper, we use the behavioral rule of commuters at a terminal station[3]. Fig. 4 shows the pedestrian behavioral rule in ABM.

First, each agent randomly selects its destination coordinates from points within a specified range and sets the direction toward these coordinates as its "front". Next, the agent checks three directions: straight ahead ('front'), 45 degrees to the left-front, and 45 degrees to the right-front. In each of these directions, it counts the number of oncoming agents present within a circle of radius r. The agent then advances a distance d in the direction with the fewest oncoming agents. After advancing, it resets its "front" to face the target coordinates again. By repeating this sequence of actions, the agent moves towards its target coordinates.

In the ABM, we conduct experiments using two model variations by changing the parameter r, which is the radius used for counting agents. Each model is defined as follows:

- ABM1: $r = 1.0$
- ABM2: $r = 2.0$

We use artisoc4[4] to implement the model.

D. Simulation Setup

Based on the target environment from the dataset, the simulation environment for all models was configured as shown in Table I. The number of agents matches the count observed during the period in the original dataset. Initial coordinates correspond to the agents' entry points into Area 06, and destination coordinates correspond to their exit points. In this experiment, almost all agents enter the area by crossing the line $x = 0$ and exit by crossing $x = 5,600$. The area boundaries were defined as $0 \leq x \leq 5,600,\ 0 \leq y \leq 9,600$.

Note that when using discrete distributions for feature representation, the grid size for discretization must be set beforehand. Here, we set the number of grid divisions to 10 and 20 for **DM1**, and 5 and 10 for **DM2**. Denoting these configurations as **DM1-S** (Small), **DM1-L** (Large), **DM2-S**, and **DM2-L**, the resulting size per grid cell is as follows:

- **DM1-S**: 960 mm
- **DM1-L**: 480 mm
- **DM2-S**: $1,920$ mm $\times$ $1,120$ mm
- **DM2-L**: 960 mm $\times$ 560 mm

Additionally, for comparison using DTW, following the approach in [7], we select a representative trajectory from the leftward-moving agents. Specifically, we consider agent data points near $x = 2800$ and choose the trajectory of the agent whose y-coordinate is the median at that approximate x-location.

V. EXPERIMENTAL RESULTS

First, as a baseline comparison, we present the results of a visual evaluation, showing visualizations of the simulation results for each model at a specific time with the visualized data of the dataset in Fig. 5. Leftward-moving trajectories are shown in red and rightward ones in blue. Additionally, the model deemed to best reproduce the ground truth is highlighted with a blue frame. It is evident that the most suitable model can vary by time, even within the same area.

From these, data corresponding to 12:00 was selected, and the KL divergence was calculated for six feature representations: **GMM1**, **GMM2**, **DM1-S**, **DM2-S**, **DM1-L**, and

TABLE I: Simulation setup.

Dir.	No.	Initial coord. [mm]	Dest. coord. [mm]
Right	52	$(0,\ 8000 \pm 10)$	$(5600,\ 2500 \pm 15)$
Left	51	$(5600,\ 1500 \pm 15)$	$(0,\ 6750 \pm 12.5)$

[2]Deep Social Force: https://github.com/svenkreiss/socialforce

[3]MAS community: https://mas.kke.co.jp/teachingmaterials/

[4]artisoc4: https://mas.kke.co.jp/artisoc4/

TABLE II: Result of calculating KL divergence for each feature representation.

	Visual eval.	DTW	**GMM1**	**GMM2**	**DM1-S**	**DM2-S**	**DM1-L**	**DM2-L**
SFM1	**1**	2.96	0.39	0.22	2.15	0.33	2.22	1.11
SFM2	**1**	**2.92**	**0.11**	**0.10**	1.88	**0.23**	1.99	**0.51**
ABM1	3	7.55	0.21	0.13	**1.07**	0.25	**1.32**	0.77
ABM2	2	5.23	0.41	0.25	2.03	0.38	2.24	2.24

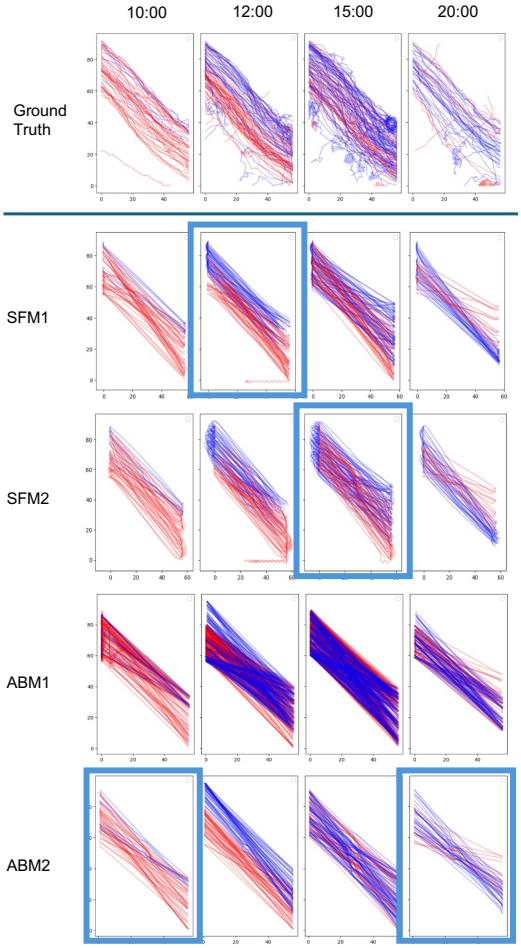

Fig. 5: Visual evaluation results. The results of the simulation for each model at each time are shown. The model that seems to reproduce the ground truth best at each time is highlighted with a blue frame.

DM2-L. The results are presented in Table II. For comparison, the table also includes the suitability ranking derived from the visual evaluation (based on Fig. 5, with lower ranks indicating higher suitability) and DTW. Furthermore, Figs. 6 through 9 show visualizations of the feature representations for **GMM1**, **GMM2**, **DM1-S**, and **DM2-S**.

First, comparing the **GMM1** and **GMM2** results, both indicate high suitability for the SFM2 model. This suggests low dependence on the specific GMM feature type used here. However, as the experimental data consists mainly of unidirectional trajectories, it's considered that high simulation reproducibility can be achieved even with simpler representations. Results might differ with more complex movement patterns. Figs. 6 and 7 show that all simulation models yield distributions with larger means and smaller variances. This indicates the simulations might not fully capture the randomness of real pedestrian behavior.

Next, we compare the results for the discrete methods. For all models evaluated, the KL divergence values are higher for the 'L' versions than the 'S' versions. This suggests that smaller grid cells yield better suitability scores in this specific experimental context. However, the optimal grid size likely depends on factors like area size, crowd density, and data volume, warranting further detailed investigation. Additionally, Figs. 8 and 9 show that ABM1, despite appearing least suitable visually in Fig. 5, exhibits strong overall distributional agreement with the ground truth, which is corroborated by its low KL divergence scores in these cases.

Overall, while the discrete distribution representations appear capable of capturing complex shapes, their effectiveness as an evaluation metric seems lower in this paper. This highlights the advantage of the GMM-based approach. Looking solely at the presented DTW values, this baseline method seems to align well with the visual assessment rankings, potentially appearing superior to the proposed KL divergence metrics. However, as noted in the study [7], the DTW approach relying on a single representative trajectory suffers from the drawback of high sensitivity to the trajectory selection process. In contrast, the proposed methods account for the overall distribution and dispersion of multiple trajectories, making them potentially more robust and effective for evaluating the suitability of crowd simulation models.

VI. CONCLUSION

In this paper, we proposed a method for evaluating the suitability of crowd simulation models for a given scenario. The approach involves representing the dispersion of multiple pedestrian trajectories as probability distributions and using the similarity between the distributions derived from a dataset and the simulation results as an evaluation metric. Specifically, we propose using GMM and discrete distributions of sampled values to represent these probability distributions, and employing KL divergence to measure the similarity between them. Furthermore, experimental results using a real dataset confirmed that although the model suitability ranking did not perfectly match the ranking from visual assessment, our approach enables an objective evaluation that considers the entire pedestrian trajectories.

Future work will involve incorporating pedestrian velocity and orientation information into the feature representations. We also plan to conduct further evaluation experiments using different crowd models and datasets.

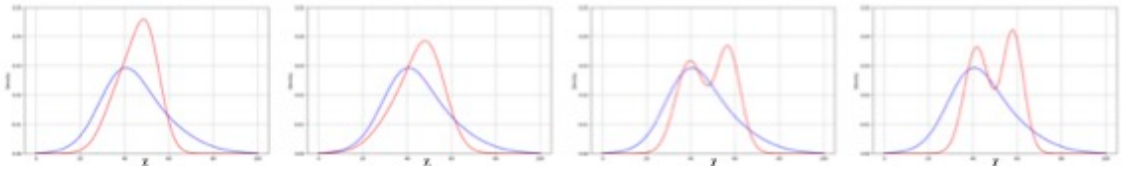

Fig. 6: Feature representation of **GMM1**. From left to right: SFM1, SFM2, ABM1, and ABM2. The blue line represents the ground truth, and the red line represents the simulation results.

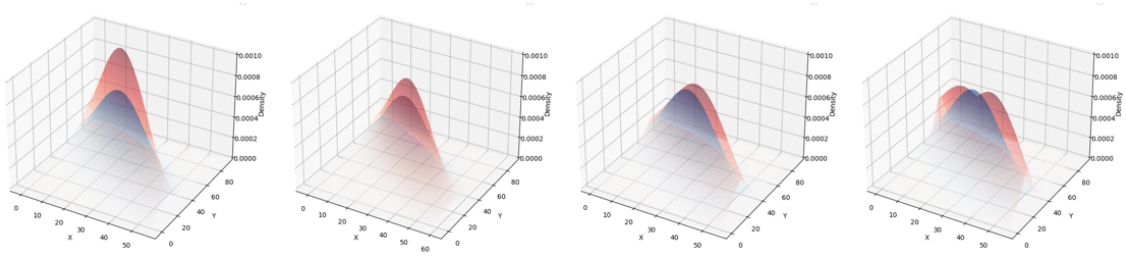

Fig. 7: Feature representation of **GMM2** (SFM1, SFM2, ABM1, and ABM2). Blue: ground truth, red: simulation results.

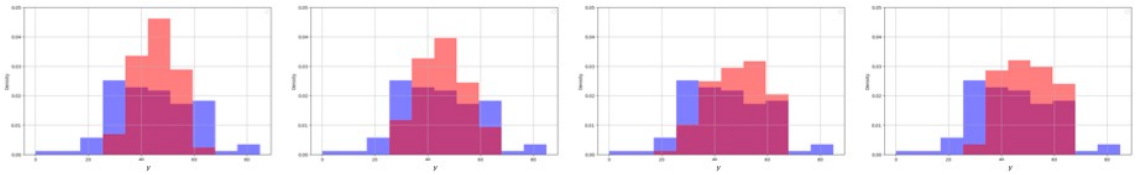

Fig. 8: Feature representation of **DM1-S** (SFM1, SFM2, ABM1, and ABM2). Blue: ground truth, red: simulation results.

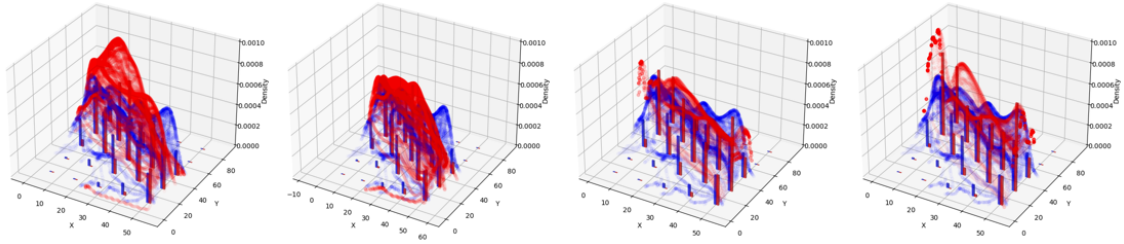

Fig. 9: Feature representation of **DM2-S** (SFM1, SFM2, ABM1, and ABM2). Blue: ground truth, red: simulation results.

REFERENCES

[1] K. Amano, A. Komori, S. Nakazawa, and Y. Kato, "Impact of environment on navigation performance for autonomous mobile robots in crowds," in *Proc. IEEE/SICE International Symposium on System Integration (SII 2023)*, 2023, pp. 794–799.

[2] I. Karamouzas, N. Sohre, R. Hu, and S. J. Guy, "Crowd space: A predictive crowd analysis technique," *ACM Trans. on Graphics*, vol. 37, no. 6, pp. 186.1–186.14, 2018.

[3] T. Fraichard and V. Levesy, "From crowd simulation to robot navigation in crowds," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 729–735, 2020.

[4] M. Tanaka and Y. Kato, "Evaluation of crowd models under various environments for robot navigation simulator," in *Proc. IEEE/SICE International Symposium on System Integration (SII 2024)*, 2024.

[5] S. Nakazawa and Y. Kato, "Environment classification method using autoencoder to select appropriate crowd model for robot simulation," in *Proc. IEEE International Conference on Automation Science and Engineering (CASE 2024)*, 2024, pp. 1877–1882.

[6] ——, "Evaluation of an environment classification method for optimal crowd model selection in autonomous mobile robot simulations," in *Proc. IEEE/SICE International Symposium on System Integration (SII 2025)*, 2025, pp. 491–496.

[7] R. Nishida and Y. Kato, "A crowd model evaluation method for autonomous mobile robot simulator," in *Proc. International Conference on Human System Interaction (HSI 2024)*, 2024, pp. 1–6.

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

[9] D. Helbing, I. Farkas, and T. Vicsek, "Simulation dynamics features of escape panic," *Nature*, vol. 407, no. 6803, pp. 487–490, 2000.

[10] J. Berg, S. Guy, M. Lin, and D. Manocha, "Reciprocal n-body collision avoidance," in *Robotics Research*. Springer, 2011, pp. 3–19.

[11] T. Dutra, R. Marques, *et al.*, "Gradient-based steering for vision-based crowd simulation algorithms," *Computer Graphics Forum*, vol. 36, no. 2, pp. 337–348, 2017.

[12] R. Dupre and V. Argyriou, "A human and group behavior simulation evaluation framework utilizing composition and video analysis," *Computer Animation and Virtual Worlds*, vol. 30, no. 1, p. e1844, 2019.

[13] W. Liu, M. Salzmann, and P. Fua, "Counting people by estimating people flows," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 11, pp. 8151–8166, 2021.

[14] B. C. Daniel, R. Marques, L. Hoyet, J. Pettré, and J. Blat, "A perceptually-validated metric for crowd trajectory quality evaluation," *Proc. of ACM Computer Graphics and Interactive Techniques*, vol. 4, no. 3, pp. 1–18, 2021.

[15] D. Brscic, T. Kanda, T. Ikeda, and T. Miyashita, "Person position and body direction tracking in large public spaces using 3D range sensors," *IEEE Trans. on Human-Machine Systems*, vol. 43, no. 11, pp. 522–534, 2013.

[16] S. Kreiss, "Deep social force," in *arXiv:2109.12081*, 2021.