Construction of A Route Choice Model for Application to A Pedestrian Flow Simulation

Ryo Nishida
Graduate School of Information Sciences
Tohoku University
Sendai, Japan
ryo.nishida.t4@dc.tohoku.ac.jp

Masaki Onishi
Artificial Intelligence Research Center
National Institute of Advanced Industrial Science and Technology
Tsukuba, Japan
onishi-masaki@aist.go.jp

Koichi Hashimoto
Graduate School of Information Sciences
Tohoku University
Sendai, Japan
koichi@m.tohoku.ac.jp

***Abstract*—In a crowded environment, it is necessary to safely guide the pedestrians. To archive optimum guidance control, we need to know the manner in which pedestrians move. Therefore, several types of studies have been conducted to propose pedestrian flow simulations for reproducing the pedestrian movement under varying situations. In one of these studies, a simulation was performed by assuming that all pedestrians will follow a guidance control. However, there are pedestrians who do not follow the guidance control because they move based on their own intentions. Therefore, in this study, we model a route choice by considering human intention.**

***Index Terms*—route choice, real crowd data**

I. Introduction

Large scale pedestrian flow simulations are used to evaluate the designs of public facilities, such as stations and shopping malls, or to evaluate the guidance control of pedestrian movements at large events. Examples of large-scale events include fireworks displays, music concerts, and sports games. For example, hundreds of thousands of people visit to the Kanmon Straits Fireworks Festival in Kitakyushu, Japan, every year. Therefore, approximately tens of thousands of people move from the fireworks display venue to the nearest train station toward the end of the event to return home. During this movement, an appropriate set of navigation instructions is required to reduce congestion. At the Kanmon Straits Fireworks Festival, people move using the routes that are indicated by the black lines in Fig. 1. Further, at a total of 10 guidance points composed of two branch points and eight pause points, as depicted in Fig. 2, security guards guide the route and control the movement of people. The congestion is reduced by dividing the pedestrians into each route via ”forward/detour” guidance at the 2 branch points. In addition, at the 8 pause points, prevent large numbers of pedestrians from flowing into the train station at the same time by stopping the progress of the pedestrian for a certain amount time via ”progress/stop” guidance.

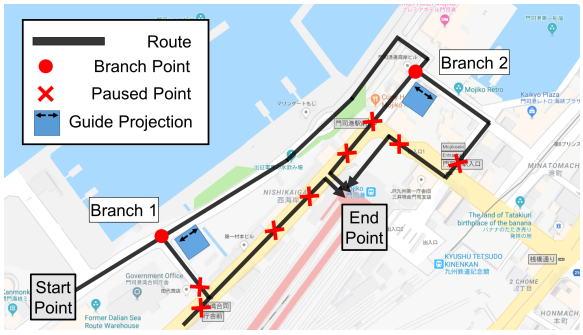


Fig. 1: Map of the area around the fireworks display venue and the train station. The black lines indicate the routes on which the pedestrians move, and the red circles indicate the branch points, and red crosses indicate the pause points. The blue squares indicate buildings with projected guidance information, as depicted in Fig. 2.

(a) Guide to move forward (b) Guide to move detour

Fig. 2: Example of guide projections

It is not realistic to collect tens of thousands of people for conducting an experiment to evaluate how smoothly people can move when the guidance control is changed. However, a pedestrian flow simulation that can reproduce the movement of a people makes it possible to evaluate the pedestrian movements while changing the

978-1-5386-9151-9/19/$31.00 ©2019 IEEE

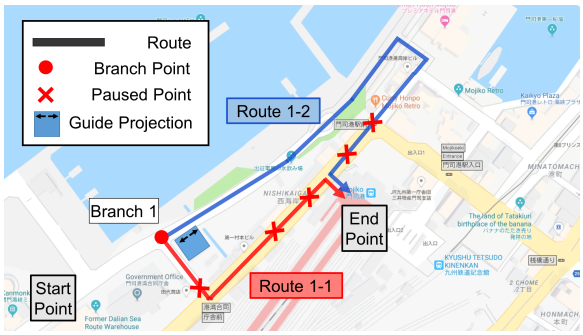
conditions several times. Therefore, studies have been conducted to propose pedestrian flow simulations. Some of these studies perform simulations by assuming that people follow the guidance control; however, there are people who do not follow this guidance because they move with based on their own intentions. Therefore, we formulate a route choice model by considering the intention of people for application to a pedestrian flow simulation.

In the following section, we describe the previous studies. In Section 3, we explain the procedure for modelling the route choice. In Section 4, we describe about a experiment constructing a route choice model using actual pedestrian data measured via LiDAR. Finally, in Section 5, we summarize the results of this paper.

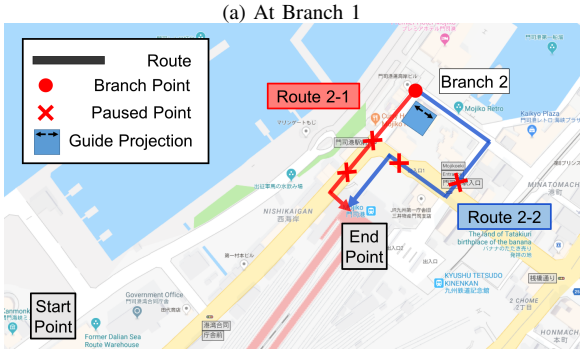
II. Previous Work

In recent years, many studies of pedestrian flow simulations using real pedestrian behavior data have been conducted. These data-driven methods tend to use video data captured by cameras. For example, Aniket et al. extracted pedestrian trajectories from the video data and estimated the state of the pedestrians and learn the parameters of a motion model. Then, they used these information to simulate the movement of pedestrians in a virtual environment [1], [2]. Further, Musse et al. clustered the trajectories that automatically extracted from the video data and generated a velocity field in each cluster. Agents in the simulation then moved based on the calculated velocity [3].

Some studies use data assimilation methods and video data to improve their reproducibility. Data assimilation can improve the simulation results by statistically or dynamically combining information obtained from both the actual measurement and the simulation. For example, Ward et al. used an ensemble Kalman filter as a method of data assimilation and use video data recorded by cameras installed in a city centr to track pedestrian movements and count the number of people [4]. Moreover, Sanish et al. simulated the dynamics of occupants and estimated their real-time spatial distribution in a building using data assimilation [5]. Other than those, Shigenaka et al. reproduced the movement of people using GPS data in addition to camera data and a data assimilation method [6]. They also studied people's movements at the Kanmon Straits Fireworks Festival with the guidance method is the same as that described in Section 1. Because it is difficult to measure the movements of all the pedestrians over the wide area of the entire venue, they measured the flow of pedestrians in a certain area using cameras and traced the entire paths of a few people using GPS. Then, the entire pedestrian flow was



(a) At Branch 1



(b) At Branch 2

Fig. 3: Route choices at the branches.

reproduced from these data using a data assimilation method. In their study, the pedestrians' proportion of division at the branch points is a fixed value, and it is assumed that all pedestrians follow the guidance control. In practice, however, the proportion of division at the branch points changes from moment to moment, and not all pedestrians follow the guidance control. Therefore, we model a route choice to reproduce more human-like movement such as not following guidance control.

III. Modelling of Route Choice

In this study's situation, the pedestrian considers two routes as choices at the first branch point as depicted in Fig. 3(a), and at the second branch point as depicted in Fig. 3(b). As depicted in Fig. 3(a), the route of the red line selected at Branch 1 is denoted Route 1-1 and the route of the blue line is denoted Route 1-2. Similarly, as depicted in Fig. 3(b), the route of the red line selected at Branch 2 is denoted Route 2-1 and the route of the blue line is denoted Route 2-2.

A. Random Utility Theory

A route choice is regarded as a discrete choice that selects one option from multiple options. For discrete choices, the random utility theory [7] is usually used. This assumes that a person chooses the option that provides the maximum utility. The choice behavior often

includes uncertain factors; therefore, the utility fluctuates stochastically. This is why it is called random utility theory. In this paper, we define the utility function, as the following equations for a route and assume that pedestrians choose the route that maximizes utility.

$$U_{in} = V_{in} + \epsilon_{in}, \tag{1}$$

$$= \sum_{k} \beta_{ik} x_{ink} + \epsilon_{in}. \tag{2}$$

Here, the utility U_{in} for a certain route n of a pedestrian i consists of the deterministic term V_{in} and the random term ϵ_{in}. The deterministic term is expressed by a linear sum obtained by multiplying the observable variables x_{ink} of size k related to the route choice by a weight β_{ik}. The random term includes terms such as unobserved variables, factors that can be considered other than the deterministic term, the error of a function of the linear sum of the deterministic term, and the measurement error of the variables of the deterministic term, and so on.

B. Variables of the Deterministic Term

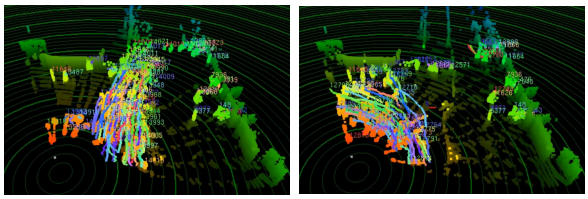
This section describes the variables $\boldsymbol{x}$ of the deterministic term of the utility function. Senevarante et al. collected questionnaires from $2,685$ people and investigated the reasons why they choose certain routes [8]. According to their survey, the number of people who cited the reason "short distance" was the highest and the other reasons are listed in descending order of number of respondents cited: "always use", "attractive", "only usable" , "few intersections", and "no congestion". We applied these to the situation at the fireworks display.

In this paper, we assume that many of the people who visited the fireworks festival came from outside the region; therefore, we did not consider the second reason, "always use". Because pedestrian were given two routes as choices at the two branch points, the reason "only usable" was also not considered. Because there are no intersections to consider in the route used, we also did not consider the reason that "few intersections". In addition to these, the pedestrians in our case want to go home as soon as possible; therefore, waiting until a route is available according to the guidance control at a branch point is definitely a factor related to the route choice. Further, in order to consider the influence of other pedestrians, we considered the number of people selected for each route to be a possible reason. Therefore, the factors "distance", "attraction", "congestion", "wait", and "number of other people who selected that route" were used as variables in the deterministic term.

We now describe the method for determining each variable of the deterministic term.

- Distance
  Here, we use the actual map values. The unit is km.
- Attraction
  At Branch 1, fireworks can be viewed between 19:50 and 20:40 while waiting at the entrance of Route 1-1; therefore, it is assumed that Route 1-1 is attractive while fireworks is setting off and a wait is occurs until Route 1-1 is open. Further, because there are stalls related to the festivals on Route 1-2, it is assumed that this route is attractive until 22:00 when the stall close. Routes 2-1 and Route 2-2 from Branch 2 do not have any attractive features. This term is 1 if the route is attractive and 0 if not.
- Congestion
  For the degree of congestion, we used the density of the pedestrians on the route. At the Kanmon Straits Fireworks Festival in 2018, we measured the flow of pedestrians using LiDAR at the two branch points as depicted in the Fig. 4. Compared to RGB-D cameras, LiDAR has the advantages that it is not affected by light changes due to time of day or environments, and it can measure a wider range. As a method for calculating the number of pedestrians on each route, we used the pedestrian flow measurement method proposed by Onishi et al [12] . Fig. 4 shows the analysis results of the data measured by LiDAR at the branch point. Dividing the number of pedestrians on the route measured every second by the area of the route, the density of pedestrians on the route can be calculated. The unit is [number of people/m^2].
- Wait
  As depicted in Fig. 2, security guards guide people at near the entrance of Route 1-1 at Branch 1 and near the entrance of Route 2-1 at Branch 2. Using measurement data via LiDAR, we judged whether or not a pedestrian was moving at the guidance position at a certain time, and the presence or absence of a sait was determined. When a wait is occurs due to the guidance control of the security guards at entrance of Routes 1-1 or 2-1, this term is set to 1, otherwise it is set to 0.
- Number of others who selected that route
  Using the measurement data via LiDAR, we calculated the route choice number per second. In this paper, the number of people who choose each route in a five-second period is used as a variable.

Using the above discussion, the deterministic term is defined as in a following equation. TABLE I shows an example of data that were actually used.



(a) Guide to move forward (b) Guide to move detour

Fig. 4: Measurement of the pedestrian movement with LiDAR at a branch point. These figures correspond to the scenes as depicted in Fig. 2.

TABLE I: Example of the variables of the deterministic term.

place	time	choices	distance	attract	density	wait	other
Branch 1	19:30:00	Route 1-1	0.325	0	0.437	1	5
		Route 1-2	0.575	1	0.220	0	0
	21:00:00	Route 1-1	0.325	0	0.693	1	0
		Route 1-2	0.575	1	0.347	0	3
	22:30:00	Route 1-1	0.325	0	0.015	0	0
		Route 1-2	0.575	0	0.0	0	0
Branch 2	19:30:00	Route 2-1	0.135	0	0.227	0	0
		Route 2-2	0.325	0	0.134	0	1
	21:00:00	Route 2-1	0.135	0	0.32	1	0
		Route 2-2	0.325	0	0.205	0	10
	22:30:00	Route 2-1	0.135	0	0.093	0	0
		Route 2-2	0.325	0	0.046	0	2

$$V_{in} = \beta_1 \times distance_{in} + \beta_2 \times attract_{in} + \beta_3 \times density_{in} + \beta_4 \times wait_{in} + \beta_5 \times other_{in}. \quad (3)$$

C. Route Choice Model

Route choice models using random utility theory can be classified into a Multinomial Probit Model (MNP) [9], which assume a normal distribution in the random term, and a Multinomial Logit Model (MNL) [10], which assume a gumbel distribution. The cumulative distribution function and the probability density function of the gumbel distribution, which is a probability distribution similar to the normal distribution, are given by equation (4) and (5), respectively. In this study, we use MNL, which is popular and easy to use, and a Mixed Multinomial Logit Model (Mixed MNL) [11] as the route choice model. Mixed MNL assumes that the weights of the variables of the deterministic term follow a probability distribution and that the weights differ among individuals. In this study, we assume that the weights of the variables of the deterministic term follow a normal distribution.

$$F(\epsilon) = \exp\left[-\exp\{-\lambda(\epsilon-\alpha)\}\right], \quad (4)$$

$$f(\epsilon) = \lambda \exp\{\lambda(\epsilon-\alpha)\}\exp[-\exp\{\lambda(\epsilon-\alpha)\}. \quad (5)$$

Next, the estimation of the weights of the variables of the deterministic term of the route choice model is described. First is the case of MNL. Given the parameters of the gumbel distribution as $\alpha = 0$ and $\lambda = 1$, the probability that a pedestrian i chooses a route n is given by

$$P_i(n) = Pr[U_{in} > U_{im}, \text{for} \forall n, m \neq n], \quad (6)$$

$$= \frac{\exp\left(V_{in}(\beta_{ik})\right)}{\sum_M \exp\left(V_{im}(\beta_{ik})\right)}, \quad (7)$$

where k represents the number of variables in the deterministic term, and in this study $k = 1, \cdots, 5$, and M indicates the number of choices. From the above, the likelihood function and the log-likelihood function are

$$L(\boldsymbol{\beta}) = \prod_i \prod_n P_i(n)^{y_{in}}, \quad (8)$$

$$\ln L(\boldsymbol{\beta}) = \sum_i \sum_n y_{in} P_i(n), \quad (9)$$

respectively, where y_{in} is 1 when a pedestrian i selects a route n and 0 otherwise. The parameters that maximize the log-likelihood function are estimated.

Next, the parameter estimation of the weights of Mixed MNL is described. Assuming that the weight β_{ik} follows a normal distribution with an average μ_k and a standard deviation σ_k and setting the parameters of the gumbel distribution to $\alpha = 0$ and $\lambda = 1$, the probability that a pedestrian i chooses a route n is given by

$$P_i(n) = \int \left[\frac{\exp\left(V_{in}(\beta_{ik})\right)}{\sum_M \exp\left(V_{im}(\beta_{ik})\right)} f(\beta_{ik}|\mu_k, \sigma_k)\right] d\boldsymbol{\beta}. \quad (10)$$

In the case of Mixed MNL, because the log-likelihood function expressed by equation (9) includes an integral form and cannot be analytically obtained, it is calculated via a numerical simulation. We generate R random numbers that follow the normal distribution via Halton extraction and use the parameters $\boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}, \cdots, \boldsymbol{\beta}^{(R)}$ to calculate the log-likelihood function. Then, the parameters that maximize this log-likelihood function are estimated.

IV. Experiments and Results

This paper deals with the movements of tens of thousands of people from the venue at the Kanmon Straits Fireworks festival to the train station. We estimated the weights of the route choice model using the measurement data at the Kanmon Straits Fireworks Festival in 2018. Both Branch 1 and Branch 2 were treated the same when estimating the weight and evaluating the accuracy of the route choice model. In other words, we assumed the situation was a route choice with two routes at a branch point. A total of 29,746 route choices

were observed. We estimated the weights using $23,796$ observations: 80% of the total data. To consider the difference in the number of parameters based on the pseudo decision coefficient [10] depicted in equation (11), we used the degree of freedom adjusted pseudo decision coefficient depicted in equation (12).

$$\rho^2 = 1 - \frac{\ln L(\hat{\boldsymbol{\beta}})}{\ln L(0)}, \tag{11}$$

$$\bar{\rho}^2 = 1 - \frac{\ln L(\hat{\boldsymbol{\beta}}) - H}{\ln L(0)}, \tag{12}$$

where $\ln L(\hat{\boldsymbol{\beta}})$ is the log-likelihood value when using the estimated parameters, $\ln L(0)$ represents the log-likelihood value obtained when all parameters are 0, and H is the estimated number of parameters. We confirmed the accuracy of route choice using the $5,950$ data: remaining 20% of the total data. The utility function was calculated using the estimated weight, and the coincidence ratio between the route with the highest utility and the route actually selected was used as the evaluation value.

TABLE Ⅱ shows the results of the weight estimation of MNL. TABLE Ⅲ shows results of the parameter estimation of Mixed MNL for the Halton extraction times $R = 1,000$. In the tables, $***$ and $**$ represent significance levels of 0.001 and 0.01, respectively. The degree of freedom adjusted pseudo decision coefficients of MNL and Mixed MNL were 0.229 and 0.260, respectively, and the accuracy of the route choices were 65.4% and 66.7%, respectively. Therefore, Mixed MNL performs better than MNL and is used as the route choice model. Route choices accuracy of 66.7% is not very high compared to a general binary classification problem; however, considering that it deals with uncertain factors, it is thought that this level of accuracy is acceptable for a route choice model.

At Branch 1, when Route 1-1 is closed by the guidance control, there are several pedestrians waiting for Route 1-1 to be open as depicted in Fig. 5. Using CrowdWalk [13] that is a pedestrian flow simulation,

TABLE II: Estimated weights of MNL

coefficient	estimate	z-value
distance	-1.739	-12.895 ***
attract	0.116	3.106 **
density	-0.291	-2.865 **
wait	-0.4808	-14.240 ***
other	0.204	64.004 ***
$\bar{\rho}^2$	0.229	
accuracy	65.4%	

TABLE III: Estimated parameters of Mixed MNL

coefficient	estimate	z-value
distance.mu	-29.412	-9.864 ***
attract.mu	0.786	0.776
density.mu	-1.699	-0.915
wait.mu	-10.055	-11.990 ***
other.mu	6.401	96.614 ***
distance.std	96.597	-21.352 ***
attract.std	54.543	20.646 ***
density.std	3.286	0.258
wait.std	2.206	0.470
other.std	4.708	47.806 ***
$\bar{\rho}^2$	0.260	
accuracy	66.7%	

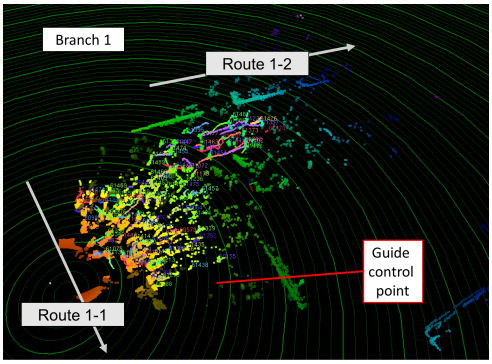
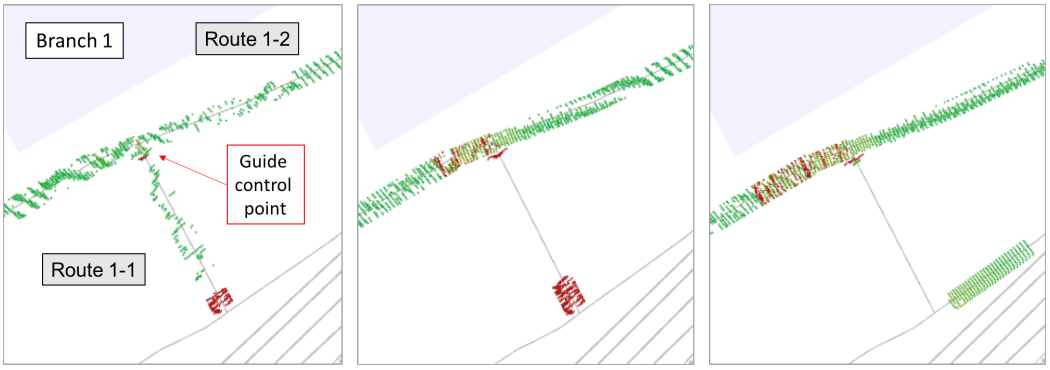


Fig. 5: Measurement of the pedestrian movement with LiDAR at Branch 1. There are several pedestrians waiting at the entrance of Route 1-1 for Route 1-1 to be open.

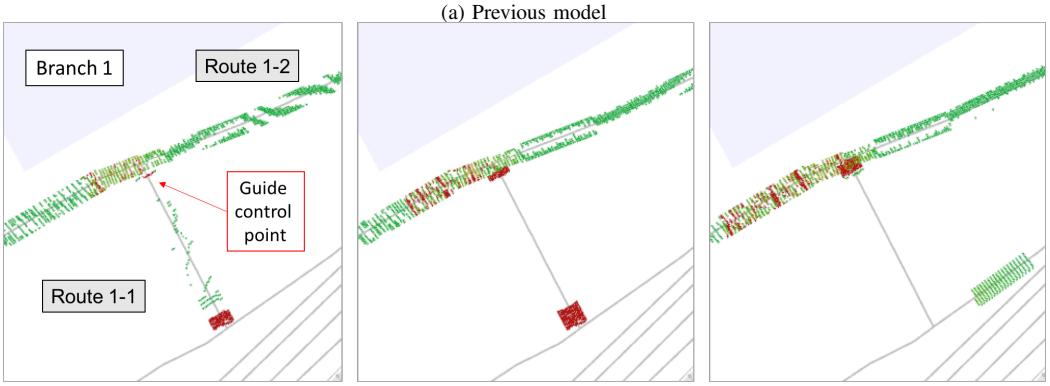
we confirmed whether human-like movement can be expressed with Mixed MNL. The pedestrians are generated, and at Branch 1, the pedestrians select a route according to Mixed MNL. As depicted in Fig. 6, we can see that the pedestrians who want to select a route with a shorter distance even if there is a waiting time, wait at the guidance control points can be reproduced. Therefore, it is obvious that the proposed model can reproduce a human-like behavior, and it is considered that a pedestrian flow simulation with the proposed route choice model can reproduce more realistic pedestrian flow.

V. Conclusions and Future Work

In this study, we have constructed a route choice model using actual data using the movement of tens of thousands of people. The future works are improving the performance of the route choice model; therefore, we will attempt to add factors related to the route choice and to improve the measurement accuracy. We believe that, if the performance of the route choice model improves, we will be able to simulate a pedestrian flow that is close



(a) Previous model



(b) Proposed model

Fig. 6: The examples of the simulation results at Branch 1. The green and red dots represent the pedestrians; pedestrians that moving are green, and pedestrians in traffic jams are represented in red. (Left) the state when the Route 1-1 start to be closed. (Middle) the state after 2 min. (Right) the state when the Route 1-1 is opened 4 min after the route is closed. we can see that the pedestrians not following the guidance, can be reproduced by proposed model.

to reality. In addition, an optimization of the route plan using the proposed simulation model can be used for evaluating the design of public facilities and planning guidance plan.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number JP16H06536.

REFERENCES

[1] B. Aniket, K. Sujeong, and M. Dinesh, "Online parameter learning for data-driven crowd simulation and content generation," Computers & Graphics, vol. 55, pp. 68–79, 2016.

[2] B. Aniket, K. Sujeong, and M. Dinesh, "Interactive Crowd-Behavior Learning for Surveillance and Training," IEEE Computer Graphics and Applications, vol. 36, no. 6, pp. 37–45, 2016.

[3] S. R. Musse, C. R. Jung, J. C. S. Jacques, and A. Braun, "Using computer vision to simulate the motion of virtual agents," Compute. Animation and Virtual Worlds, vol. 18, no. 2, pp. 83–93, 2007.

[4] J. A. Ward, A. J. Evans, and N. S. Malleson, "Dynamic calibration of agent-based models using data assimilation," Royal Society open science, vol. 3, no. 4, 2016.

[5] S. Rai and X. Hu, "Building occupancy simulation and data assimilation using a graph-based agent-oriented model," Physica A: Statistical Mechanics and its Applications, vol. 502, pp. 270–287, 2018.

[6] S. Shigenaka, M. Onishi, T. Yamashita, and I. Noda, "Estimating Pedestrian Flowin Crowded Situations with Data Assimilation," The Thirty-Third AAAI Conference on Artificial Intelligence workshop on Games and Simulations for Artificial Intelligence, 2019, unpublished.

[7] C. F. Manski, "The structure of random utility models," Theory and Decision, vol. 8, no. 3, pp. 229–254, 1978.

[8] P. N. Seneviratne and J. F. Morrall, "Analysis of factors affecting the choice of route of pedestrians," Transportation Planning and Technology, vol. 10, no. 2, pp. 147–159, 1985.

[9] L. L. Thurstone, "A law of comparative judgement," Psychological Review, vol. 34, pp. 273–286, 1927.

[10] D. McFadden, "Conditional Logit Analysis of Qualitative Choice Behaviour," In Frontiers in Econometrics, P. Zarembka (Ed.). Academic Press New York, pp. 105–142, 1973.

[11] D. McFadden and K. Train, "Mixed MNL models of discrete choice," Journal of Applied Econometrics, vol. 15, no. 5, pp. 447–470, 2000.

[12] M. Onishi, "Analysis and Visualization of Large-Scale Pedestrian Flow in Normal and Disaster Situations," ITE Transactions on Media Technology and Applications, vol. 3, no. 3, pp. 170–183, 2015.

[13] T. Yamashita, T. Okada, and I. Noda, "Implementation of Simulation Environment for Exhaustive Analysis of Huge-Scale Pedestrian Flow," SICE Journal of Control, Measurement, and System Integration, vol. 6, no. 2, pp. 137–146, 2013.