Transportation (2013) 40:91–107
DOI 10.1007/s11116-012-9411-z

Calibrating a social-force-based pedestrian walking model based on maximum likelihood estimation

Moonsoo Ko · Taewan Kim · Keemin Sohn

Published online: 8 May 2012

Abstract Although various theories have been adopted to develop reliable pedestrian walking models, a limited effort has been made to calibrate them rigorously based on individual trajectories. Most researchers have validated their models by comparing observed and estimated traffic flow parameters such as speed, density, and flow rate, or replaced the validation by visual confirmation of some well-known phenomena such as channelization and platooning. The present study adopted maximum likelihood estimation to calibrate a social-force model based on the observed walking trajectories of pedestrians. The model was assumed to be made up of five components (i.e., inertia, desired direction, leader–follower relationship, collision avoidance, and random error), and their corresponding coefficients represented relative sensitivity. The model also included coefficients for individual-specific characteristics and for a distance-decay relationship between a pedestrian and his/her leaders or colliders. The calibration results varied with the two density levels adopted in the present study. In the case of high density, significant coefficient estimates were found with respect to both the leader–follower relationship and collision avoidance. Collision avoidance did not affect the pedestrian's walking behavior for the low-density case due to channelization. The distance limit was confirmed, within which a pedestrian is affected by neighbors. At the low-density level, by comparison with women, men were found to more actively follow leaders, and pedestrians walking in a party were found to be less sensitive to the motion of leaders at the high-density level.

Keywords Pedestrian walking behavior · Social force model · Calibration · Maximum likelihood estimation · Leader–follower relationship · Collision avoidance

M. Ko · T. Kim · K. Sohn (✉)
Department of Urban Engineering, Chung-Ang University, 221, Heukseok-dong, Dongjak-gu, Seoul 156-756, Korea
e-mail: kmsohn@cau.ac.kr

M. Ko
e-mail: kokyul@paran.com

T. Kim
e-mail: twkim@cau.ac.kr

Introduction

Recently, the importance of modeling pedestrian walking behaviors has been emphasized in the field of transportation studies. Pedestrian behavior was first examined in the context of how quickly and safely people could evacuate an emergent condition, and was then extended to the realm of normal situations. Many researchers have delved into the behavior of panicking people when an emergent event takes place in a closed space with limited exits (Xiaoping et al. 2010; Izquierdo et al. 2009; Fang et al. 2008; Guo and Huang 2008; Song et al. 2006). However, despite a number of devised models, there has been no way to validate such behavior based on observed trajectories, since it is almost impossible to synthesize an emergent condition and to contrive an experiment for human behavior. Shiwakoti et al. (2011) recently introduced a revolutionary calibration methodology by tracing panicking ants, and yielded a plausible model that can be directly applied to human behavior by simply scaling ant-based parameters.

As the non-motorized travel mode was spotlighted to accomplish sustainable transportation, many researchers have turned their eyes to an analysis of walking behaviors for normal conditions, which is necessary when establishing design guidelines for pedestrian-related facilities such as transit stations or terminals, walkways, and transportation plazas. Simulation studies have become the main mode for the analysis of pedestrian behavior under normal conditions. According to the taxonomy of Teknomo (2002), microscopic simulations for pedestrians can be categorized into three types based on both model structure and theory: cellular-based models, queuing-network models, and physical-force-based models. The queuing-network model, which was developed for the purpose of evacuation (Thompson and Marchant 1995; Watts 1987), was not a concern of the present study.

Blue and Adler (2001) adopted a cellular-automata model for pedestrians, which has already proven to be a good approximation for vehicular flow on a road (Maerivoet and De Moor 2005; Nagel 1998; Nagel and Rasmussen 1994). When broadly defined, the cellular-based pedestrian model encompasses a lattice-gas model (Isobe et al. 2004; Takimoto et al. 2002; Muramatsu and Nagatani 2000), a floors model (Burstedde et al. 2001), a pre-fixed probabilities model (Weifeng et al. 2003; Jian et al. 2005), and a dynamic-parameters model (Yue et al. 2010; Yamamoto et al. 2007). Pedestrian space is commonly divided into small fixed cells and applies intuitively accountable behavioral rules for changes in the state of each cell. This simple structure is known to be effective for reproducing collective phenomena observed in real pedestrian flow. Despite the advantage of simplicity, the calibration of the models is only partially based on real pedestrian movement data. The validation compared the models with observations for either collective results such as speed, density, and flow rate or their interrelationships. Moreover, the models were unable to account for the impact of individual characteristics on walking behavior—gender, age, whether to carry a bag, and whether to travel in party—although the characteristics were expected to differentiate pedestrian walking behavior to a great extent.

In the early stages of pedestrian study, Okazaki (1979) introduced a magnetic-force model that falls in the category of physical-force-based models. The model was developed to analyze pedestrian movement in an architectural space by borrowing a motion equation used for magnetic fields in general physics. The validation of the model, however, was done only by visual inspection after arbitrarily setting the magnetic load of a pedestrian. A more robust physical-force-based model was introduced by Helbing and Molár (1995) under the title of a social-force model. The model assumed

that a pedestrian is subjected to social forces that motivate him/her, which represent the sum of internal motivation, external impact, and random error. Afterwards, many researchers developed their own social-force models by modifying or adding elements to the original social-force model (Parisi et al. 2009; Mehran et al. 2009; Yu et al. 2005; Hoogendoorn 2002). Regarding calibration, they either derived the model parameters directly from intrinsic physical laws, or validated them by visually comparing estimated results of pedestrian movement with observed ones. Recently, as the calibration of pedestrian models has been considered an important pursuit, several researchers, such as Hoogendoorn and Daamen (2006) and several article authors included in a handbook edited by Klingsch et al. (2010), have calibrated social force-based models in a statistical manner.

The present study developed a social-force model by choosing five different force elements and by transforming the original continuous differential equation with respect to force into a discrete difference equation with respect to velocity. This simplification made it possible to devise a rigorous calibration methodology based on the MLE. The MLE is easy to apply and is based on a straightforward theory—once a model to be calibrated has an additive random term with a closed-form probability density function (PDF). The contribution of the present study is not confined to adopting such a rigorous methodology for the model calibration. Model coefficients that have proven to be intuitively accountable and statistically significant could offer information that is valuable in determining the intrinsic property of pedestrian behavior. For example, estimated coefficients answered many obscure questions: which of the social forces has maximal influence and which has minimal, how individual-specific characteristics change a pedestrian's walking behavior, to what degree pedestrians are affected by their neighbors, and whether a calibrated model is transferable across different levels of pedestrian density. While a noted effort to calibrate a pedestrian model within a framework of discrete choice was made by Robin et al. (2009) in a fully statistical manner based on the MLE, the calibration methodology adopted in the present study was simpler and similar to that used by Hoogendoorn and Daamen (2006). Differences between approaches of the present study and those by the two previous studies will be addressed in detail in the third section.

The next section reveals the details of a model that has been developed for calibration. The third section will introduce the data acquisition procedure and present descriptive statistics of the data, followed by a rigorous calibration methodology in the fourth section. Calibration results will be presented and discussed in the fifth section. Final conclusions will be drawn.

Model development

The social-force model developed in the present study chose five sources of social force: inertia to maintain direction, force toward destination, leader–follower relationships, collision avoidance, and random error. Inertia, the first term of the right-hand side of Eq. (1), was generated automatically in the process of transforming the original continuous differential equation with respect to force into the discrete difference equation with respect to velocity, which essentially cannot be seen as a social force, but will, for convenience, be referred to as a force in the present study.

$$V_i(t+\Delta t) = \beta_1 V_i(t) + \beta_2 \frac{\mathrm{dest_i} - \mathrm{S_i}(t)}{\|\mathrm{dest_i} - \mathrm{S_i}(t)\|} + \left(\beta_3 + \sum_{k=1}^{p} \delta_k X_{ki}\right) \sum_{j \neq i} \big(\mathrm{V_j}(t) - \mathrm{V_i}(t)\big) \mathrm{K}\big(\theta, \mathrm{d_{ij}}\big) \mathrm{w_{ij}}(t) + \left(\beta_4 + \sum_{k=1}^{p} \varphi_k X_{ki}\right) \sum_{j \neq i} \big(\mathrm{V_j}(t) - \mathrm{V_i}(t)\big) \mathrm{K}\big(\tau, \mathrm{d_{ij}}\big) \mathrm{y_{ij}}(t) + \mathrm{u_i}(t) \quad (1)$$

where $\mathrm{V_i}(t) = \big(\mathrm{v_{xi}}(t), \mathrm{v_{yi}}(t)\big)'$ is the velocity of pedestrian i at time interval t, Δt is the time interval, $\mathrm{S_i}(t) = (\mathrm{x_i}(t), \mathrm{y_i}(t))'$ is the location of pedestrian i at time interval t, $\mathrm{dest_i} = \big(\mathrm{dest_{xi}}, \mathrm{dest_{yi}}\big)'$ is the final destination of pedestrian i, which depicts his/her exit location from the test bed, X_{ki} is the dummy variable representing kth individual-specific characteristic of pedestrian i, P is the number of individual-specific characteristics considered, $\mathrm{K}(\theta, \mathrm{d_{ij}})[or \mathrm{K}(\tau, \mathrm{d_{ij}})]$ is the Kernel function reflecting the distant-decay relationship between pedestrians i and j, $\mathrm{d_{ij}}$ is the aerial distance between pedestrians i and j, $\mathrm{u_i}(t) = (\varepsilon_x, \varepsilon_y)'$ is the random fluctuation (=error) of pedestrian i's movement at time interval t, $\mathrm{w_{ij}}(t) = \begin{cases} 1 & \text{if pedestrian j is a leader of pedestrian i at a time interval t} \\ 0 & \text{Otherwise} \end{cases}$,

$\mathrm{y_{ij}}(t) = \begin{cases} 1 & \text{if pedestrian j is a collider against pedestrian i at a time interval t} \\ 0 & \text{Otherwise} \end{cases}$, and

$\beta_1, \ldots, \beta_4, \delta_1, \ldots, \delta_p, \varphi_1, \ldots, \varphi_p, \theta, \tau$ is the model coefficients to be calibrated.

The second term of the right-hand side of Eq. (1) denotes a pedestrian's inclination toward his/her final destination. In a physical concept, the force toward destination at desired speed (μ_{max}) is expressed by $\mu_{max} \frac{\mathrm{dest_i} - \mathrm{S_i}(t)}{\|\mathrm{dest_i} - \mathrm{S_i}(t)\|}$, but the present study assumed that, regardless of the physical law, the force might have a weight $\left(\beta_2'\right)$ reflecting its relative importance to other walking motivations on a empirical basis. For convenience, the parameter was estimated in an incorporated form [refer to Eq. (2)], and was then separated after estimation, given that μ_{max} was determined from the observed speed distribution.

$$\beta_2 = \beta_2' \mu_{max} \quad (2)$$

The third and fourth elements of the right-hand side of Eq. (1) stand for intentions to follow leaders and to avoid collision, respectively. At each time interval, the designation of leader or collider pedestrians among a pedestrian's neighbors was assigned depending on indicators $\mathrm{w_{ij}}(t)$ and $\mathrm{y_{ij}}(t)$ that were derived through the following procedure. The pseudo code is self-explanatory to the point that further interpretation was omitted.

If $\mathrm{V_i}(t) \cdot \big(\mathrm{S_j}(t) - \mathrm{S_i}(t)\big) \geq \|\mathrm{V_i}(t)\| \big\|\mathrm{S_j}(t) - \mathrm{S_i}(t)\big\| \cos\phi$ then

[Pedestrian j lies within pedestrian i's visual angle $2\phi(= 170^\circ)$]
If $\mathrm{V_i}(t) \cdot \mathrm{V_j}(t) \geq \|\mathrm{V_i}(t)\| \big\|\mathrm{V_j}(t)\big\| \cos\phi$ then

$\mathrm{w_{ij}}(t) = 1$ [Pedestrian j is a leader of pedestrian i]

Else

$\mathrm{y_{ij}}(t) = 1$ [Pedestrian j is a collider of pedestrian i]

End if

Else

$w_{ij}(t) = 0$, $y_{ij}(t) = 0$ [Pedestrian j is out of pedestrian i's vision]

End if

The leader and collider influence on pedestrians should be adjusted by the distance from the pedestrian to them. A kernel function with a coefficient (θ or τ) can accommodate such distance-decaying relationships. How to select a kernel function out of many candidates will be described later in the calibration methodology section.

A pedestrian's individual-specific characteristics were assumed to influence only his/her behavior as he/she followed leaders and avoided colliders, which differs from the other two behaviors that stem from internal motivation. It is plausible to hypothesize that a pedestrian's individual-specific characteristics are automatically involved in his/her behavior of keeping direction or orienting destination through his/her own specific speed level. Thus, the model was established such that the sensitivity only for the third and fourth terms of Eq. (1) could vary with individual-specific characteristics $\left(X_{1i}, \ldots, X_{pi}\right)$. The last term $\left[u_i = (\varepsilon_x, \varepsilon_y)^t\right]$ of the model represents the randomness of pedestrian movement, which makes it possible to estimate model coefficients in a rigorous way. Details of the error term will be discussed in the calibration methodology section.

Data acquisition

Pedestrian movement data for calibration was collected by shooting video, as utilized by many previous researchers (Robin et al. 2009; Mehran et al. 2009; Kretz et al. Kretz et al. 2006; Teknomo 2002). As Fig. 1 shows, the present study chose a crosswalk with a traffic signal as a test bed, the dimensions of which were 10 by 26 m. The signal cycle was tantamount to 150 (s), and the green phase was 30 (s).

Two video clips were taken during the green phases representing bi-directional pedestrian walking flow under low- and high-density conditions. What mattered the most in collecting data through videos was how to track the individual movement of each pedestrian. Pattern recognition technologies have been adopted by many researchers (Li et al. 2008; Antonini et al. 2006; Szarvas et al. 2005), but thus far their methodologies have not been satisfactory to allow the avoidance of a manual post-process, particularly when used under high-density conditions. The aim of the present study was not to develop an automatic methodology to track moving pedestrians from video sequences, but rather to establish a robust calibration methodology for a pedestrian behavior model. Thus, the necessary data were extracted manually from the video clips.

In order to discretize the continuous movements of pedestrians, the video clips were divided into screen sequences for every half-second frame. Each movement of pedestrians was then extracted from every two consecutive video sequences and recorded to a database. The total number of pedestrians traversing the crosswalk during the surveyed green phase were 53 and 105 for low- and high-density cases, respectively (see Fig. 1). With the exception of data that were contaminated in the course of the manual collection, 2,062 and 4,077 samples of half-second movement data were obtained for the two different cases. Although the green time was 60 time frames (=30 s), the number of total time frames was set as >60 (actually, 76 and 93 for the two different cases), so that pedestrians who did not complete their traversal till the green phase ended could be traced.

Fig. 1 Test bed for data acquisition

Affine matrices that were used to convert a screen coordinate system into an actual one were obtained based on a simple approach presented by Teknomo (2002). He established a regression model by setting real coordinates as dependent variables and screen coordinates as independent variables for several reference points designated in advance. Despite the use of a small number of reference points (=25) when using his approach, the present study resulted in very high R^2 values for both low- and high-density cases, as shown in Table 1. The location and angle of the camera varied for each case of low and high conditions because the two videos were shot separately on different days. All descriptive statistics below regarding pedestrian speed and density were derived based on the transformed real coordinate system.

Speed profiles can be seen in Fig. 2, in which the prior expectation that the pedestrian speed follows a log-normal distribution was confirmed by taking a logarithm of the original pedestrian speed. Figure 3 plots the mean values of speed and density computed for each time interval. The speed drop phenomenon was observed only when density level increased [larger than about 0.25 (persons/m^2)]. The reason that data points in the high-density case

Table 1 The results of coordinate transformation

Dependent variable	R^2
Low density	
Real X coordinate	0.985
Real Y coordinate	0.987
High density	
Real X coordinate	0.987
Real Y coordinate	0.991

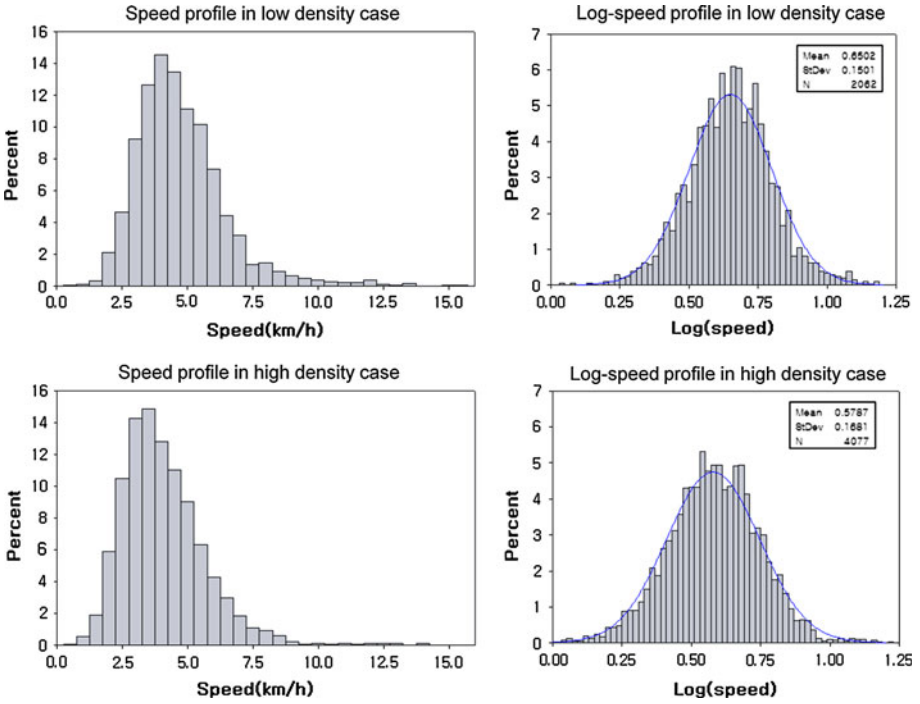

Fig. 2 Speed profiles of sample pedestrians

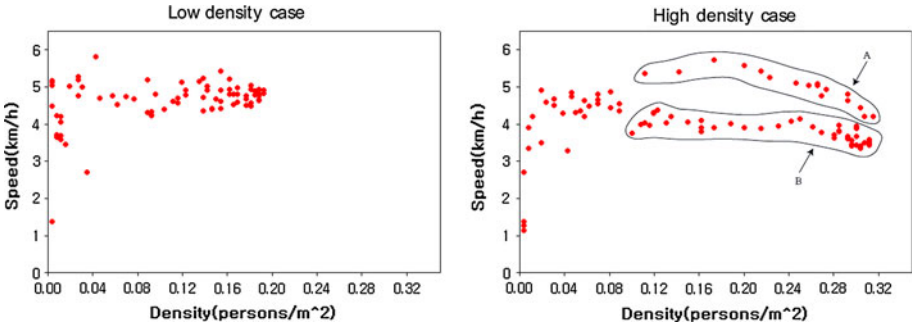

Fig. 3 Relationship between the density and speed observed for each time period

are clearly divided into groups "A" and "B" was confirmed to be due to interactions between pedestrians. That is, each data point in group "A" corresponded to time intervals prior to observing the interruption from counter flow.

A key contribution of the present study is that the individual properties of each pedestrian were surveyed and considered in the calibration analysis. Four kinds of individual-specific variables were taken into account: gender (X_{1i}), age (X_{2i}), whether to carry a bag (X_{3i}), and whether to travel in a party (X_{4i}). Among these properties, age was not easy to estimate by observation alone, so the only reliable distinction was between children and adults. Table 2 shows the results of the statistical test used to judge whether the mean

Table 2 Differences in pedestrian speed according to individual-specific variables

Variable	Mean speed (kph)	Standard deviation (kph)	t value	*p* vlaue
(a) Low-density case				
X_{1i}				
Male (=1)	4.81	1.76	1.80	0.04*
Female (=0)	4.67	1.68		
X_{1i}				
Grown-up (=1)	4.74	1.73	0.01	0.49
Children (=0)	4.74	1.35		
X_{3i}				
In party (=1)	4.91	1.26	1.64	0.05*
Alone (=0)	4.72	1.78		
X_{4i}				
With a bag (=1)	4.79	1.76	0.97	0.17
Without a bag (=0)	4.72	1.70		
(b) High-density case				
X_{1i}				
Male (=1)	4.23	1.72	5.23	0.00**
Female (=0)	3.97	1.51		
X_{1i}				
Grown-up (=1)	4.10	1.62	4.31	0.00**
Children (=0)	3.54	1.12		
X_{3i}				
In party (=1)	3.70	1.38	−6.54	0.00**
Alone (=0)	4.15	1.64		
X_{4i}				
With a bag (=1)	3.89	1.39	−7.14	0.00**
Without a bag (=0)	4.25	1.78		

* Significant at the 0.05 level

** Significant at the 0.001 level

speed was different for the two groups divided by the individual-specific characteristics. In the case of low density, the pedestrian speed proved to differ at the 5 % significance level only for two variables such as gender and whether to travel in a party. However, in the high-density case, significant differences in pedestrian speed were found for every individual-specific variable. The assumption of the previous section that a pedestrian's individual-specific characteristics are only associated with interaction forces with other pedestrians was verified after observing that they had a more statistically significant influence on pedestrian speed in the high-density case.

Calibration methodology

Prior to describing the calibration methodology adopted in the present study, the two prior studies that have previously dealt with the calibration of a pedestrian walking model in a

statistical manner must be addressed. Robin et al. (2009) developed a pedestrian model complete with a rigorous calibration methodology based on random utility theory and a discrete-choice model. They not only established a model and its calibration methodology, but also provided many meaningful and statistically significant results regarding pedestrian behavior, which could help other researchers to set up their own pedestrian models. In a broad sense, their model can be included in the category of cellular-based models, despite its unique merit as it adopted a context-dependent cellular system that varied according to situation. Thus, this model is superior to any other cell-based models developed with a fixed cellular system. Utilizing their model, however, requires considerable effort in setting up the context-dependent cellular system and computing the probability that each alternative cell will be chosen. It was burdensome for practitioners to restructure cell systems for every pedestrian at every time interval. A motivation of the present study was to find a simple model and its simple calibration methodology for a practical purpose. Thus, a social force model and a calibration methodology based on MLE were adopted to analyze pedestrian behaviors under normal conditions.

On the other hand, Hoogendoorn and Daamen (2006) has already dealt with a model based on social forces and a calibration procedure based on MLE. The calibration methodology adopted in the present study was essentially identical to that employed by Hoogendoorn and Daamen (2006), except it had more coefficients to be calibrated. The present model distinguished coefficients of a neighbor pedestrian's interferences according to whether he/she is a leader or a collider, and added coefficients for individual-specific characteristics, which could compensate the effort of Hoogendoorn and Daamen (2006) to verify inter-pedestrian differences in parameter values. In addition, the present study calibrated the coefficients separately for the two different density conditions. Details regarding the calibration methodology adopted in the present study were summarized as follows.

The inclusion of the kernel function deprives the linearity of the present model (Eq. 1), without which the calibration process would have been facilitated. MLE could be a simple way to estimate coefficients of the non-linear model. In general, MLE has been used in many transportation studies (Sohn and Kim 2010; Ben-Akiva et al. 2002; Liu and Mahmassani 2000). MLE enables the simultaneous estimation of all the coefficients, including the parameter in kernel function, and produces asymptotically efficient and unbiased estimates (Lehmann and Casella 1998).

The error's stochastic property is the key factor driving MLE. The random error term of Eq. (1) was assumed to have a bivariate normal probabilistic density function (PDF) with a zero mean and a variance–covariance matrix (C). Equation (4) shows a two-dimensional normal PDF.

$$f(u_i(t)) = \frac{1}{(2\pi)|C|^{1/2}} e^{-(u_i(t)' C^{-1} u_i(t))/2} \tag{4}$$

where $f(u_i(t))$ is the two-dimensional normal PDF for $u_i(t) = (\varepsilon_x, \varepsilon_y)'$, $C = COV[u_i(t)] = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix}$ is the variance–covariance matrix of $u_i(t)$.

The rationale for employing MLE is that the data are observed because they are most likely to appear. The likelihood of the present survey can be expressed by the product of PDF values for each error $[u_i(t)]$ of a pedestrian's movement at every time interval. The likelihood function is expressed with model coefficients $(\beta_1, \ldots, \beta_4, \delta_1, \ldots, \delta_4, \varphi_1, \ldots, \varphi_4, \theta, \tau)$ and

variance–covariance elements $(\sigma_x, \sigma_y, \sigma_{xy})$, given that six kinds of observed data, $\left[V_i(t), S_i(t), dest_i, d_{ij}, w_{ij}(t), \text{and } y_{ij}(t)\right]$, are available, as follows:

$$L\left(\beta_1, \ldots, \beta_4, \delta_1, \ldots, \delta_4, \varphi_1, \ldots, \varphi_4, \theta, \tau, \sigma_x, \sigma_y, \sigma_{xy} | V_i(t), S_i(t), dest_i, d_{ij}, w_{ij}(t), y_{ij}(t)\right) = \prod_i \prod_t f(u_i(t)) \tag{5}$$

$u_i(t)$ can be expressed by the difference between the observed and estimated velocities of each pedestrian's movement, as follows:

$$\begin{aligned} u(t) = {} & V_i(t+\Delta t) - \beta_1 V_i(t) + \beta_2 \frac{dest_i - S_i(t)}{\|dest_i - S_i(t)\|} \\ & - \left(\beta_3 + \sum_{k=1}^{p} \delta_k X_{ki}\right) \sum_{j \neq i} \left(V_j(t) - V_i(t)\right) K(\theta, d_{ij}) w_{ij}(t) \\ & - \left(\beta_4 + \sum_{k=1}^{p} \varphi_k X_{ki}\right) \sum_{j \neq i} \left(V_j(t) - V_i(t)\right) K(\tau, d_{ij}) y_{ij}(t) \end{aligned} \tag{6}$$

The maximum likelihood estimates of model parameters are obtained such that Eq. (5) is maximized. To facilitate the computation, the logarithm is taken on both sides of Eq. (5). Readers who are interested in the computational details can refer to (Lehmann and Casella 1998).

$K(\theta, d_{ij})$ [or $K(\tau, d_{ij})$] is the kernel function that reflects how far a pedestrian can be affected by his/her neighbor pedestrians, and θ (or τ) is the parameter associated with the distance limit (=bandwidth). It was postulated in the model development section that the influence of neighbor pedestrians on a pedestrian of interest decays as the distance between them increases. Several typical kernel functions that Wand and Jones (1995) had suggested were tested for the present study. As a result, the Gaussian kernel was adopted since its mathematical differentiability can facilitate the calibration process. However, the Gaussian kernel function is unbounded by distance (d_{ij}), so in theory every neighbor is considered to affect a reference pedestrian. However, the bandwidth (h) can be approximately set as $2\sqrt{1/2\theta}$ representing two standard deviations of the Gaussian kernel, beyond which the neighborhood effect is drastically damped to a negligible level. Equation (7) expresses the Gaussian Kernel with the relationship between kernel parameter and bandwidth.

$$K(\theta, d_{ij}) = e^{-\theta d_{ij}^2} \quad \text{where}, h \cong 2\sqrt{1/2\theta} \tag{7}$$

The triangular, quadratic, and exponential kernels were also tested, which, respectively, reduced the neighborhood influence linearly, quadratically, and exponentially proportional to the distance between pedestrians. Unfortunately, the adoption of the first two kernels led to failure in the calibration procedure, which might be due to the function discontinuity around the bandwidth value. The exponential kernel brought out the success in calibration, but was excluded because the Gaussian kernel was a better fit.

Discussion of calibration results

It should be noted that not every model specification can be successfully calibrated using given sample data. The final two models were obtained after repeatedly testing a number of

plausible specifications within the model framework developed above. Moreover, to a great extent, the success in calibration depended on how to choose an initial solution. To facilitate the convergence of the log-likelihood function, an initial solution was obtained from a preliminary regression analysis, given that the kernel parameter giving rise to nonlinearity was fixed as an arbitrary value.

While being tested, the variance–covariance matrix (C) of error term was reduced to $(\sigma^2 I)$, as shown in Eq. (8). This reflects that both errors with respect to the lateral and vertical components of a pedestrian's velocity would have nothing to do with each other $(\sigma_{xy} = 0)$, and their fluctuation ranges would be the same $\left(\sigma_x^2 = \sigma_y^2 = \sigma^2\right)$. Any other combinations of the matrix elements did not lead to success in model calibration.

$$C = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix} = \sigma^2 I = \sigma^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{8}$$

The final models were found to share a kernel coefficient for the leader–follower relationship and collision avoidance (i.e., $\theta = \tau$), which means that a pedestrian's distance ranges for the neighborhood effect would be identical regardless of whether his/her neighbor is a leader or a collider. Regarding parameters for individual-specific variables, the gender parameter (δ_1) was left in the final model for the low-density case, while the parameter (δ_3) of whether to travel in a party survived in the final model for the high-density cases. Equations (9) and (10) denote the final model specifications for low- and high-density cases, respectively.

$$\begin{aligned} V_i(t+\Delta t) = {} & 0.758 V_i(t) + 0.160 \frac{dest_i - S_i(t)}{\|dest_i - S_i(t)\|} \\ & + (0.062 + 0.038 X_{1i}) \sum_{j \neq i} \left(V_j(t) - V_i(t)\right) e^{-0.064 d_{ij}^2} w_{ij}(t) \\ & + 0.001 \sum_{j \neq i} \left(V_j(t) - V_i(t)\right) e^{-0.064 d_{ij}^2} y_{ij}(t) + u_i(t) \end{aligned} \tag{9}$$

$$\begin{aligned} V_i(t+\Delta t) = {} & 0.760 V_i(t) + 0.137 \frac{dest_i - S_i(t)}{\|dest_i - S_i(t)\|} \\ & + (0.107 - 0.054 X_{3i}) \sum_{j \neq i} \left(V_j(t) - V_i(t)\right) e^{-0.134 d_{ij}^2} w_{ij}(t) \\ & + 0.007 \sum_{j \neq i} \left(V_j(t) - V_i(t)\right) e^{-0.134 d_{ij}^2} y_{ij}(t) + u_i(t) \end{aligned} \tag{10}$$

Prior to discussing the calibration results, the model performance was measured by comparing observed and estimated velocities of pedestrians. Figure 4 compares observed and estimated speeds, which were obtained by computing the mathematical norm of the velocity vector. For both cases of low and high density, Regression coefficients (1.0098, 1.0139) were very close to unity, but relatively low R^2 values (0.567, 0.5443) were derived. The low R^2 values imply that a considerable amount of variables, which were unknown but play a crucial role in accounting for pedestrian behavior, were omitted in the model. To overcome this complication, further studies should develop a more delicate model structure and excavate new potential elements affecting the behavior of pedestrians.

However, the model's accuracy with regard to the direction of velocity was at an acceptable level. The difference in angle between observed and estimated velocity vectors was computed to confirm the model performance. As shown in Fig. 5, for both cases, the

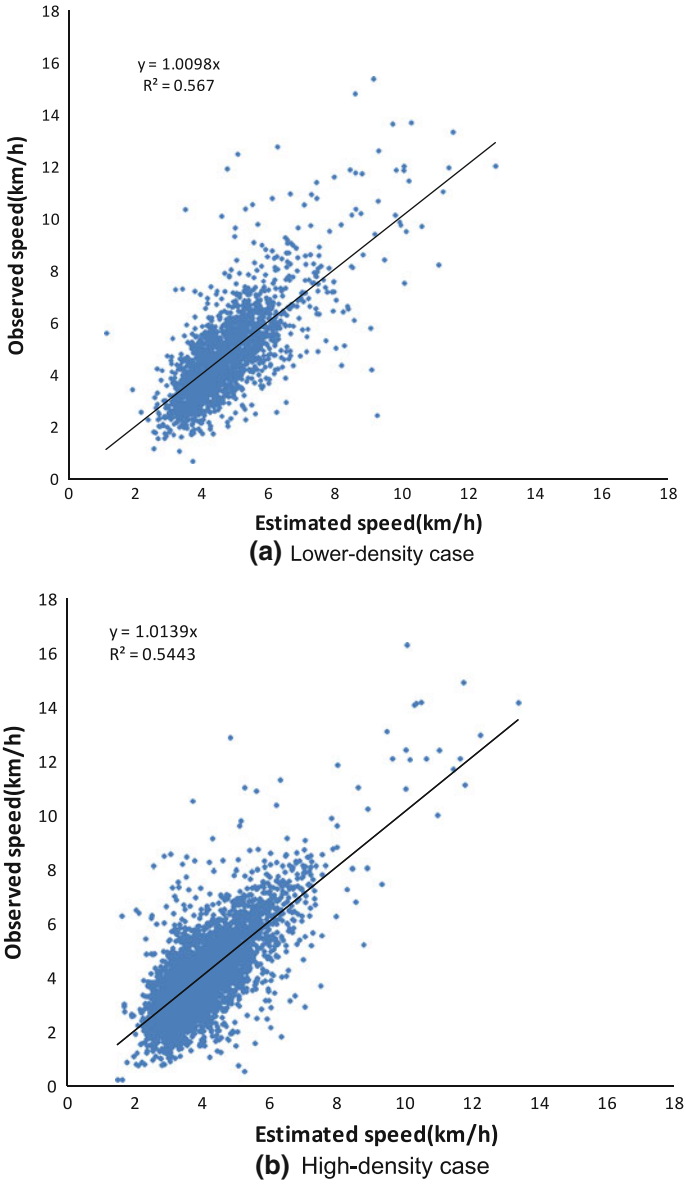

Fig. 4 Comparison of observed and estimated speeds

distribution of angles between observed and estimated velocity vectors was not so skewed to the right. More than 80 % (97 %) of angles were <10° (20°) for the lower-density case, while more than 72 % (93 %) fell below the range for the high-density case. This implies that the model has good qualification in reproducing the real direction of pedestrian movement. The model performance, however, slightly deteriorated under the high-density condition where conflicts among pedestrians grew.

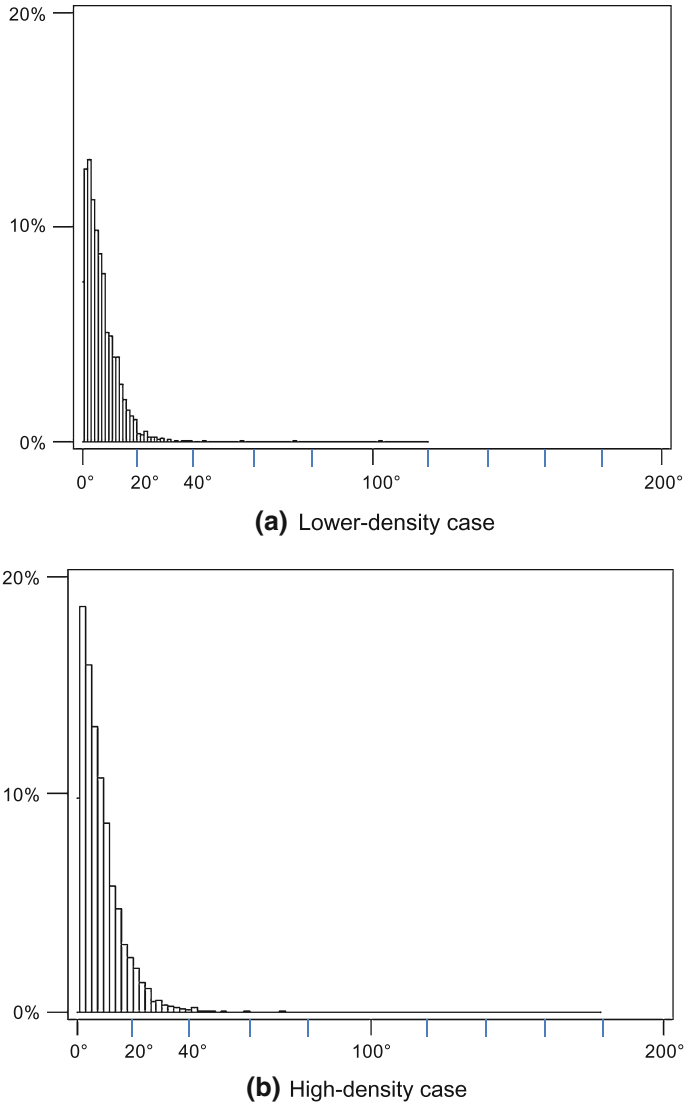

Fig. 5 Distribution of angles between observed and estimated velocities

One of the main objectives of the present study was to investigate what the calibration results imply. The calibrated estimates with their statistical properties are shown in Table 3. For both cases of low and high density, the coefficient for inertia had the biggest influence on accelerating or decelerating, and the magnitude of the influence was a constant value of ~ 0.76. This means that most of the driving force for pedestrians stemmed from the inertia. The net value (β_2') of the coefficient for force toward destination was computed by Eq. (2) under the assumption that the maximum speed (μ_{max}) of pedestrians would be 6 (km/h). The coefficient (0.490) for the high-density case was smaller than that (0.533) for the low-density case. That is, in the high-density case, pedestrians were less likely to walk straight to their own destinations since they frequently conflicted with

Table 3 Calibration results of the final model specifications

Variables	Coefficients	Estimates	Standard error	t value	p value
(a) Low-density case					
Inertia	β_1	0.758	0.012	63.401	0.000**
Toward destination	$\beta_2(\beta_2')$	0.160 (0.533)	0.008	19.736	0.000**
Leader–follower	β_3	0.062	0.010	6.335	0.000**
Collision avoidance	β_4	0.001	0.002	0.397	0.691
Gender w.r.t Leader–follower	δ_1	0.038	0.011	3.486	0.001*
Distance-decaying	θ(h)	0.064 (5.590)	0.012	5.234	0.000**
Random fluctuation	σ	0.017	0.000	45.397	0.000**
Log-likelihood	2591.72				
Number of samples	2061				
(b) High-density case					
Inertia	β_1	0.760	0.009	86.36	0.000**
Toward destination	$\beta_2(\beta_2')$	0.137 (0.490)	0.006	24.623	0.000**
Leader–follower	β_3	0.107	0.010	10.567	0.000**
Collision avoidance	β_4	0.007	0.002	4.918	0.000**
Whether to travel in party w.r.t leader–follower	δ_3	−0.054	0.011	−4.855	0.000**
Distance-decaying	θ(h)	0.134 (3.863)	0.017	8.059	0.000**
Random fluctuation	σ	0.016	0.000	63.85	0.000**
Log-likelihood	2736.83				
Number of samples	4077				

* Significant at the 0.05 level

** Significant at the 0.001 level

colliders in their way. This interpretation was also consistent with the statistical insignificance regarding the coefficient for collision avoidance in the low-density case. This also reflects that a pedestrian is unlikely to be interrupted by other pedestrians where counter flows are channelized as in the low-density case in the present study. The varying influence of colliders according to density level was confirmed from another result that suggested larger changes in pedestrians' walking angles in the high-density case than in the low-density case. The walking angle changes $>10°$ exceeded 34 % in the high-density case, while the changes never exceeded >23.6 % in the low-density case.

The leader–follower relationship happened to have a larger influence on a pedestrian's acceleration or deceleration than collision avoidance in the high-density case. This was opposite to prior expectations, and might have been due to a specific condition of the chosen test bed where channelized counter flows reduced the possibility of collision. Regarding leader–follower relationships and collision avoidance, the maximum distance by which pedestrians can be affected by other pedestrians was derived from the calibration results. Based on the approximation formula $(2\sqrt{1/2\theta})$, the maximum bandwidth (h) of distance-decaying kernel function was computed from the estimated coefficient (θ). As a result, a follower accelerated or decelerated in response to the speed of his/her leaders or colliders within a range of up to 5.590 and 3.863 (m) for low- and high-density cases, respectively. There was no doubt that the distance limit was shorter in the high-density case, because a pedestrian must be close to his/her neighbors under jammed conditions.

Regarding the effect of individual-specific characteristics, the calibration results were not consistent with the prior expectations that they would play a decisive role in describing pedestrian behavior. Including some variables of individual-specific characteristics into the model even made the calibration process infeasible. Only one coefficient for each case of low and high density yielded a feasible model specification and was found to be statistically significant. It is natural that a man was more responsive to his leaders than a woman in the low-density case. It also was not surprising that pedestrians walking in a party were less adaptive to their leaders in the high-density case.

From the above results, it was apparent that the estimated coefficients differed between high- and low-density cases. This connotes that a single model specification could not cover all conditions of walking density. Although further tests for diverse density levels must be conducted, it is clear that the previous efforts to set up a single model specification should be seriously reconsidered. Multi-regime analysis by Johansson and Helbing (2008) that was adopted to identify the behavior of people in an emergent situation might be applied for studies of normal walking conditions.

Conclusions

The present study suggested a social-force-based model for pedestrian walking behavior and calibrated its coefficients in a rigorous methodology. The model was assumed to be made up of five components (i.e., inertia, desired direction, leader–follower relationship, collision avoidance, and random error). The calibration was conducted on a statistical basis using MLE. The calibrated model proved to reproduce the observed velocity of pedestrians at an acceptable level. Several useful facts were discovered from the calibration results.

While the impact of inertia was a constant for both low- and high-density cases, the estimated coefficient of force toward destination showed the difference between the two density levels. As expected, the influence of force toward destination weakened as the density in pedestrians grew. Impacts of a leader–follower relationship and collision avoidance were also apparently different according to density level. In particular, the coefficient for collision avoidance in the low-density case had no statistical significance. It turned out that the leader–follower relationship has a larger influence on a pedestrian's acceleration or deceleration than collision avoidance in the case of high density. It was proven that a follower accelerates or decelerates in response to the speed of his/her leaders within a range of up to 5.590 and 3.863 (m) for low- and high-density cases, respectively. Only one coefficient with respect to individual-specific characteristics was found to be statistically significant for each case of low and high densities; gender in the low-density case and whether to travel in a party for the high-density case. Consequently, it was apparent that the calibration results varied with the two density levels hypothesized in the present study. Further studies, however, need to be conducted to support the argument by investigating a greater number of density conditions.

References

Antonini, G., Mrtinez, S.V., Bierlaire, M., Thiran, J.P.: Behavioral priors for detection and tracking of pedestrians in video sequences. Int. J. Comput. Vision **69**, 159–180 (2006)

Ben-Akiva, M., Walker, J., Bernardino, A.T., Gopinath, D., Morikawa, T., Polydoropoulou, A.: Integration of choice and latent variable models. In: Mahmassani, H.S. (ed.) Perpetual Motion: Travel Behaviour Research Opportunities and Application Challenges, pp. 431–470. Elsevier, Amsterdam (2002)

Blue, V.J., Adler, J.L.: Cellular automata microsimulation for modeling bi-directional pedestrian walkways. Transp. Res. B **35**, 293–312 (2001)
Burstedde, C., Klauck, K., Schadschneider, A., Zittartz, J.: Simulation of pedestrian dynamics using a two-dimensional cellular automaton. Phys. A **295**, 507–525 (2001)
Fang, G., Kwok, N.M., Ha, Q.P.: Swarm interaction-based simulation of occupant evacuation. In: Proceeding of Pacific-Asia Workshop on Computational Intelligence and Industrial Application, pp. 329–333 (2008)
Guo, R.Y., Huang, H.J.: A mobile lattice gas model for simulating pedestrian evacuation. Phys. A **387**, 580–586 (2008)
Helbing, D., Molár, P.: Social force model for pedestrian dynamics. Phys. Rev. E **51**, 4282–4286 (1995)
Hoogendoorn, S.P.: Walker behaviour modelling by differential games. In: Proceedings of Computational Physics of Transport and Interface Dynamics. Springer, Berlin (2002)
Hoogendoorn, S.P., Daamen, W.: Microscopic parameter identification of pedestrian models and implications for pedestrian flow modeling. Transp. Res. Rec. **1982**, 57–64 (2006)
Isobe, M., Adachi, T., Nagatani, T.: Experiment and simulation of pedestrian counter flow. Phys. A **336**, 638–650 (2004)
Izquierdo, J., Montalvo, R.P., Fuertes, V.S.: Forecasting pedestrian evacuation times by using swarm intelligence. Phys. A **388**, 1213–1220 (2009)
Jian, L., Lizhong, Y., Daoliang, Z.: Simulation of bi-direction pedestrian movement in corridor. Phys. A **354**, 619–628 (2005)
Johansson, A., Helbing, D.: Analysis of Empirical Trajectory Data of Pedestrians. Handbook of Pedestrian and Evacuation Dynamics 2008, Part 1, pp. 203–214, (2010)
Klingsch, W.W.F., Rogsch, C., Schadschneider, A., Schreckenberg, M.: Pedestrian and Evacuation Dynamics 2008. Springer, Heidelberg (2010)
Kretz, T., Grünebohm, A., Schreckenberg, M.: Experimental study of pedestrian flow through a bottleneck. J. Stat. Mech. P10014 (2006)
Lehmann, E.L., Casella, G.: Theory of Point Estimation, 2nd edn. Springer, Berlin (1998)
Li, Z., Tang, Q.L., Sang, N.: Improved mean shift algorithm for occlusion pedestrian tracking. Electron. Lett. **44**, 622–623 (2008)
Liu, Y.H., Mahmassani, H.S.: Global maximum likelihood estimation procedure for multinomial probit (MNP) model parameters. Transp. Res. B **34**, 419–449 (2000)
Maerivoet, S., De Moor, B.: Cellular automata models of road traffic. Phys. Rep. **419**, 1–64 (2005)
Mehran, R., Oyama, A., Shah, M.: Abnormal crowd behavior detection using social force model. In: Proceeding of IEEE Conference on Computer Vision and Pattern Recognition, pp. 935–942 (2009)
Muramatsu, M., Nagatani, T.: Jamming transition of pedestrian traffic at a crossing with open boundaries. Phys. A **275**, 281–291 (2000)
Nagel, K.: From particle hopping models to traffic flow theory. Transp. Res. Rec. **1644**, 1–9 (1998)
Nagel, K., Rasmussen, S.: Traffic at the edge of chaos. In: Artificial Life IV: Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems, pp. 222–225 (1994)
Okazaki, S.: A study of pedestrian movement in architectural space, Part I: pedestrian movement by the application of magnetic models. Trans. A.I.J. **283**, 111–119 (1979)
Parisi, D.R., Gilman, M., Moldovan, H.: A modification of the social force model can reproduce experimental data of pedestrian flows in normal conditions. Phys. A **388**, 3600–3608 (2009)
Robin, T., Antonini, G., Bierlair, M., Cruz, J.: Specification, estimation and validation of a pedestrian walking behavior model. Transp. Res. B **43**, 36–56 (2009)
Shiwakoti, N., Sarvi, M., Rose, G., Burd, M.: Animal dynamics based approach for modeling pedestrian crowd egress under panic conditions. Transp. Res. B Methodol. **45**(9), 1433–1449 (2011)
Sohn, K., Kim, D.: Zonal centrality measures and the neighborhood effect. Transp. Res. A **44**(9), 733–743 (2010)
Song, W., Xu, X., Wang, B.H., Ni, S.: Simulation of evacuation processes using a multi-grid model for pedestrian dynamics. Phys. A **363**, 492–500 (2006)
Szarvas, M., Yoshizawa, A., Yamamoto, M., Ogata, J.: Pedestrian detection with convolutional neural networks. In: IEEE Proceedings on Intelligent Vehicles Symposium, pp. 224–229 (2005)
Takimoto, K., Tajima, Y., Nagatani, T.: Pattern formation and jamming transition in pedestrian counter flow. Phys. A **308**, 460–470 (2002)
Teknomo, K.: Microscopic pedestrian flow characteristics: Development of an image processing data collection and simulation model. Ph.D. Dissertation, Tohoku University, Japan (2002)
Thompson, P.A., Marchant, E.W.: A computer model the evacuation of large building population. Fire Saf. J. **24**, 138–148 (1995)

Wand, M.P., Jones, M.C.: Kernel Smoothing, Monographs on Statistics and Applied Probability. Chapman & Hall/CRC, New York (1995)
Watts, J.M.: Computer models for evacuation analysis. Fire Saf. J. **12**, 237–245 (1987)
Weifeng, F., Lizhong, Y., Weicheng, F.: Simulation of bi-direction pedestrian movement using a cellular automata model. Phys. A **321**, 633–640 (2003)
Xiaoping, Z., Wei, L., Chao, G.: Simulation of evacuation process in a square with a partition wall using a cellular automation model for pedestrian dynamics. Phys. A **389**, 2177–2188 (2010)
Yamamoto, K., Kokubo, S., Nishinari, K.: Simulation for pedestrian dynamics by real-coded cellular automata (RCA). Phys. A **379**, 654–660 (2007)
Yu, W.J., Chen, R., Dong, L.Y., Dai, S.Q.: Centrifugal force model for pedestrian dynamics. Phys. Rev. E **72**, 016002 (2005)
Yue, H., Guan, H., Zhang, J., Shao, C.: Study on bi-direction pedestrian flow using cellular automata simulation. Phys. A **389**, 527–539 (2010)

Author Biographies

Moonsoo Ko is a graduate student in the Department of Urban Engineering, Chung-Ang University, Seoul, Korea. He is preparing a dissertation for a master's degree based on the present study.

Taewan Kim is an Associate Professor in the Department of Urban Engineering at Chung-Ang University in Seoul, Korea. He obtained his Ph.D. degree from the University of California, Davis in 2003. His research interests cover traffic flow theory, traffic operation, ITS, and urban transportation planning.

Keemin Sohn is an Associate Professor in the Department of Urban Engineering at Chung-Ang University in Seoul, Korea. He obtained the degree of Doctor of Engineering from Seoul National University, Korea. His research interests cover urban transportation planning and its interactions with policies directed at reducing car use.