Estimation of People Flow in a City Using Particle Filter with Person Trip Survey, Building Data and Traffic Count

Toshikazu Nakamura, Yoshihide Sekimoto, Ryosuke Shibasaki

Center for Spatial Information Science, University of Tokyo, Japan
ki_ki_gu@csis.u-tokyo.ac.jp (T. Nakamura); sekimoto@csis.u-tokyo.ac.jp (Y. Sekimoto); shiba@csis.u-tokyo.ac.jp (R. Shibasaki)

Abstract

Recently, people flow information in a city has become more necessary to maintain roads, prevent traffic accidents, and activate the city. Although some surveys have been conducted to obtain such information, it is necessary to estimate whole people flow by combining such measurement data because measuring whole people flow is difficult due to cost. Based on these backgrounds, the objective of this study is to estimate whole people flow from Person Trip Survey (PTS) more accurately by some different kinds of measurement data using data assimilation techniques. In the proposed algorithm, we construct a people flow estimation model from PTS and building data, and correct the estimation by traffic count using data assimilation technique with Particle Filters. For validation, we apply it to Nagakute-shi in Japan, an area of about 21.5 of square kilometers and 54,000 residents, and more accurate people flow can be estimated.

1. Introduction

Nowadays people flow information in a city level has become more important for surveillance, activity recognition, traffic accident avoidance and traffic flow analysis. For example, we can detect roads where few people pass through by grasping people flow information and the information about the roads is useful to set cameras to monitor them to prevent incidents. Moreover, prediction of pedestrians who run into a street lets vehicles avoid traffic accidents. In this situation, for measuring people flow, many surveys such as person trip survey have been conducted and many systems have been developed using various measurement instruments such as cameras, integrated circuit (IC) card, global positioning systems (GPS) and laser range scanners. For example, we can count the number of passengers getting on and off by IC card, obtain people trajectories by GPS, and track and detect people in a restricted area by cameras or laser scanner sensors. However, in spite of the developments, measurement of the whole people flow in a city is difficult because of cost and maintenance. Therefore, it is necessary to estimate whole people flow. In addition, combining these fragmented observations is important to make a more accurate estimation.

Based on these backgrounds, the objective of this study is to estimate people flow by combining some different kinds of measurement data, based on data assimilation techniques. In the proposed algorithm, a people flow estimation model is constructed by Person Trip Survey and building data, and then the estimation of the model is corrected by traffic count by data assimilation technique with Particle Filters. For validation, the proposed method is applied to Nagakute-shi in Japan, an area of about 21.5 of square kilometers and 54,000 residents, and more accurate people flow can be estimated compared with only using the people flow estimation model constructed from Person Trip Survey.

2. Overview

2.1. Reconstruction of people flow using Person Trip Survey

Person Trip Survey (PTS), a questionnaire transportation survey, is often conducted to analyze people movements for urban planning. In this survey, some households are extracted as samples and they are required to answer their information, individual attributes, and trip information often in a day.

Trip is a person's movement from origin to destination with purpose. For example, if a man went to his office in the morning, left the office for a supermarket in the evening after working, and went back to his home, he had three trips in the day. Trip consists of some sub-trips, which is a movement separated by transportation mode. For example, if a man moved from his home to the nearest station of his home by walk, moved from the station to the nearest station of his office by train, and arrived at his office by walk, the trip contains three sub-trips.

Many studies for reconstruction of people flow from PTS have been developed to estimate the whole people flow in a city. For example, Sekimoto et al. proposed a data process system for the reconstruction of spatio-temporal positions of large numbers of people from PTS at high-resolution time intervals of a minute, based on fragmentary spatio-temporal location information of person trip survey (Sekimoto et al., 2011).

As PTS is a sampling survey, how to magnify samples is important to grasp the whole people flow. Many studies expand the survey data using the magnification value calculated by population data to analyze urban transportation system and predict traffic demand. It may be useful in an aggregation level such as transportation mode analysis and od demand analysis. However, in a trajectory level, the reconstruction using the magnification value is not accurate because the reconstruction let many people move as similarly as the sampled person did in the survey. More accurate estimation in a trajectory level is required to analyze the movements in a city for surveillance and traffic accident avoidance.

2.2. Route Estimation

Many studies use the shortest path as a route estimation of a movement from origin to destination. Dijkstra algorithm is well-known logic to create the shortest path. However, as human activity has become more complicated, the diversity of peoples' route choices has increased. Hence, the shortest path is not adequate and many route choice algorithms have been developed to tackle this problem. Some are geometric algorithms and some involve use of the logit model.

There are two important steps in route estimation. One is to generate a route choice set and the other is to select the best-fitting route. Although the former is focused on in this study, some algorithms such as those based on the logit model address the two steps at once (Dial, 1971). These algorithms can generate some routes probabilistically. However, they have a number of disadvantages. For example, the parameter tuning is cumbersome, the computational cost is high, and many similar routes are generated.

Some algorithms separate the problem into a generation phase and a selection phase. There are many algorithms to generate route choice sets. The K-shortest Path algorithm is a simple and well-known method that generates the first k shortest paths. Link penalty and link elimination are two popular techniques that iteratively generate a shortest path by giving penalties to or eliminating links used as the shortest path (De la Barra et al., 1993). The labeling approach generates many types of route such as the shortest path, the minimum cost path, the minimum time path, and the maximum use of expressways path (Ben-Akiva et al., 1984). Some algorithms based on GSP algorithm have been developed. GSP algorithm is an algorithm that searches the shortest path passing through a given gateway (Lombard, K. et al., 1993). However, some critical shortcomings are indicated (Akgum, V., et al., 2000). The main one is that some of the gateway paths may contain loops. In recent years, some studies have been developed to generate routes using some movement data such as GPS.

In this paper, link penalty logic is used to simply generate some routes. Although this method do not explain the real route choice activities because penalty varies much depending on an origin and destination, routes generated by it are sufficient in this study because they are evaluated by some real observation data such as traffic count.

2.3. Data assimilation

Data assimilation is the process of combining observations and simulation models. It is often used in weather forecasting. More accurate estimation can be operated by selecting the best one fitting to observations among the candidates that are given probabilistically by the simulation model.

Some techniques have been developed to assimilate observation data into simulation models. Kalman filter, one of the simplest techniques, can be applied to linear dynamic simulation models and all error terms and measurements that have a Gaussian distribution (Kalman, 1960). Extended Kalman filter has been developed for nonlinear simulation models. Since it adapted Taylor series expansions for nonlinear models to be applied Kalman filter, the restriction that the models can be differentiable is remained. Ensemble Kalman filter is a recursive filter suitable for problems with a large number of variables (Evensen, 1994). It can be applied to simulation models that cannot be differentiated, but makes the assumption that all probability distributions involved are Gaussian. Particle filter is a sequential monte carlo method as same as Ensemble Kalman filter, but it does not require that probability distributions are Gaussian (Kitagawa, 1996, Gordon et al., 1993). Hence, it is suitable for the nonlinear simulation models that include various probability distributions.

As trajectories of people movements are not linear and have many kinds of probability distributions, we use particle filter to assimilate observation data into the people flow estimation model.

3. Approaches

3.1. Whole framework of the proposed method

Three phases are in the framework: person creation phase, particle generation phase, and data assimilation phase (Figure 1).

First of all, we create persons by magnifying PTS samples who move in the target area. Magnification rate is calculated by population in the target area and sample persons' attribute data. For example, when the sampling rate is 3 percent, 33 persons are created on average. After samples are magnified, the created persons have some movements in a day based on PTS. Trips in PTS use some zones as origin and destination, each of which contains many buildings. Therefore, there are many candidate buildings as origin and destination building. Hence, we randomly give a building to each created person as home and disperse the destination of each trip so that many created persons do not move from the same origin and destination building. Home locations and destinations are dispersed with building data such as home, office, school and station locations. Each person departs each home as the origin at the first movement and then departs at the previous destination at the other movements.

Secondly, in the particle generation phase, some routes are generated by link penalty algorithm and the same numbers of particles, each of which has each route, are generated.

Finally, in the data assimilation phase, weights of generated particles are updated by the observation data such as traffic count at some selected roads and the probabilities of particles are calculated fitted best to the observation data. Then, the whole people flow is estimated by selecting particles using their weights.

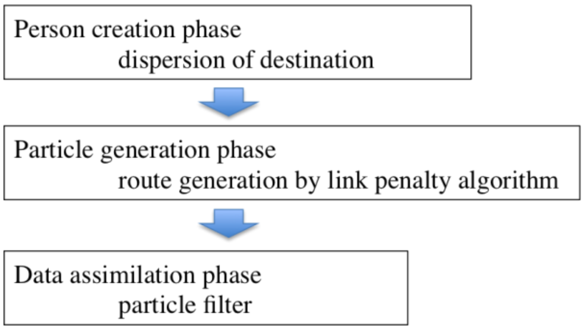

Fig. 1. Flow chart of the framework

3.2. People flow simulation model

In this section, two operations are explained: how to disperse destinations and how to generate some routes.

We use building data to disperse destinations. PTS has purpose information in each trip such as going to office, going back home, going to school, and shopping. On the other hand, building data has a type such as home, office, school, station, shopping store, and restaurant. Therefore, candidates of destination buildings can be narrowed down by choosing buildings, types of which are related to trip purpose. Buildings are divided into three types in this study because some types of buildings cannot determine trip purpose. Three types are home, school, and business place. In PTS, the origin and destination is aggregated into some zones. Hence, building data is allocated to the zones and the destination is dispersed in the same zone.

Some routes are generated by link penalty algorithm (Figure 2). At first, the shortest path is generated by a shortest path generation algorithm such as Dijkstra algorithm. Then, the cost of links in the shortest path is magnified at the given rate. By this operation, links in the shortest path are less likely chosen when the next shortest path is generated. As a result, by iterating the generation of the shortest path search and the magnification of the cost of links in the shortest path, some routes that are usually not the same each other can be generated. This route generation algorithm has some disadvantages is used in this study in spite of some disadvantages because we focus on combining simulation models and observation data and it is simple enough to apply.

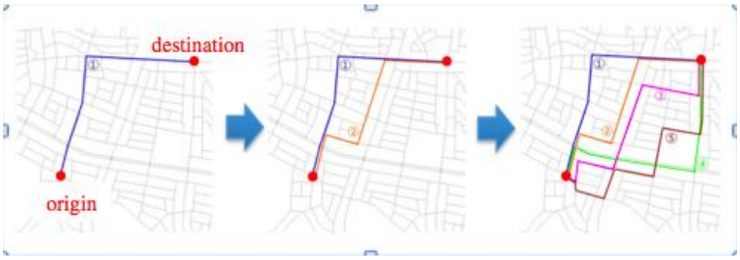

Fig. 2. Route generation by link penalty algorithm

3.3. Particle filter

Particle filter consists of two models: dynamic model and measurement model. The dynamic model is used to express how the state changes over time. The dynamic model is defined as following:

$$P(x_t|x_{t-1}) \tag{1}$$

where x_t is the estimated value at time t and a first order Markov process. This equation describes that x_t is given stochastically using the previous state x_{t-1}. The measurement model is used to describe the probability distribution of the state after some observation data are obtained. The measurement model is defined as following:

$$P(y_t|x_t) \tag{2}$$

where y_t is an observation at time t and each y_t only depends on x_t. This equation describes that y_t has probability distribution depending on x_t. In particle filter, the conditional probability is represented by a set of weighted samples (particles). It can be expressed by Figure 3. In Figure 3, $S_{t,k}^{(n)}$ is n-th samples of object k at time t and $p_{t,k}^{(n)}$ is probability of samples (weights). At first, samples are generated and xt+1 is predicted by equation 1. Then, weights of samples are updated by the observation data at time t+1 and the measurement model. Finally, the state at time t+1 is estimated and it is reflected the dynamic model, the measurement mode, and the observation data.

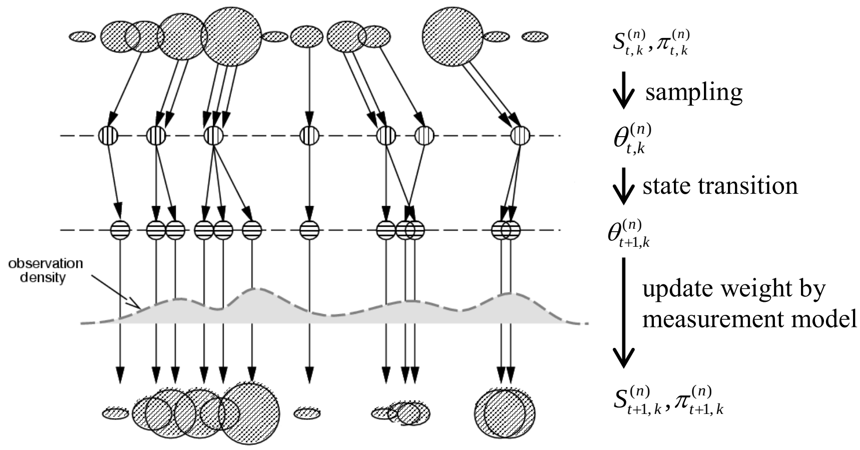

Fig. 3. Update step in particle filter

3.4. Application of particle filter

In this study, an algorithm using particle filter is proposed for data assimilation of people flow estimation and observation data. At first, particles are generated by the proposed people flow estimation model. Each of them has a weight that is initialized as *1/n* (*n* is the total number of particles that a person has). For example, when a person moves from an origin to destination at time t, some particles each of which has a different route from the origin to destination are generated. If 10 particles are generated, weight of each of them is 0.1. Particles are generated for all people in the target city. Then the weights of particles are updated when some observation data are obtained. When an observation *n(t)* is obtained at time t, weights of particles are updated by the following equation:

$$\hat{w}(i) = \frac{n(t)}{\sum_{j=1}^{n} w(j)} \times w(i) \tag{3}$$

where *w(i)* is the weight of the i-th particle that is observed between time t-1 and time t. After weights are updated, the total of weights of each person's particles may not equal to 1. Therefore, finally, weights of each person's particles are normalized and the weights are regarded as the probabilities of particles.

The updated weights of particles may not be fitted to the observation data because they are normalized by each person. Therefore, the update phase should be repeated so that weights of particles satisfy both that they are fitted to the observation data and that they are normalized by each person. In the convergence phase, the condition of the convergence is expressed as the following equation:

$$\left| \frac{w_n - w_{n-1}}{w_n} \right| < 0.01 \tag{4}$$

where w_n is the weight of a particle at n-th iteration.

4. Experimental studies

4.1. Data

The proposed method is applied to Nagakute-shi in Japan, an area of about 21.5 of square kilometers and 54,000 residents. Nagakute-shi is one of residential areas in Japan and residents often take vehicles to move because railway network is much less than Tokyo. Of course, there are many pedestrians and people who take bicycles such as students going to school or going back to home and people going shopping. Therefore, the avoidance of traffic accident with vehicles and pedestrians or bicycles is one of the most important issues in the city. Based on these backgrounds, the flow of people who move by walk or bicycle is estimated in this study.

Person trip survey has been conducted in Chukyo metropolitan area five times. In these surveys, People Flow Project[1] provides the processed survey data in 2001. The processed data are spatio-temporal positions by a minute instead of trip information. Table 1 shows main items in the processed data. As a preparation, the trip data that are by walk or bicycle are extracted from main items in Table 1. As a result, 1060 persons and 2292 trips are extracted.

[1] http://pflow.csis.u-tokyo.ac.jp/index.html

Table 1. Main items of the processed data provided by People Flow Project

Person id	Trip id	Sub-trip id
Date	Longitude	Latitude
Purpose	Transportation mode	Magnification factor

Figure 4 shows the origin and destination of the extracted person trip survey data. Red markers are the origin and destination, and the black lines connect them. Color strength of black lines shows the number of trips between the connected origin and destination. As shown in this figure, the origin and destination are biased because person trip survey deals with the location by zones.

Figure 5 shows the building locations in the part of the target area. Yellow markers are the locations of home, orange markers are the ones of office and commercial building, and blue markers are the ones of school. Black lines are borders of zones. As the destinations are aggregated, they are dispersed using the building data and the borders of zones. In this study, home, school and business place are used as three types of building in terms of the accuracy of building types. For example, separating offices and shopping stores is difficult because of the cases that a person works at a shopping store and the other person going there to shop something and that offices and shopping stores are often in the same building.

Table 2 shows the correspondence table between purpose and building type of destination. Since buildings can be divided into only three types due to the accuracy, purposes are also divided into three ones. For example when purpose is “go to school”, destinations are dispersed into buildings of which type is “school” in the same zone. Thus, destinations are dispersed in a building level.

DRM (Digital Road Map), one of the most accuracy and detail road network data in Japan, is used as a road network. In addition, as the observation data for the data assimilation, traffic count surveys are conducted at 12 locations shown in Figure 6. Hourly traffic count is used for data assimilation in this study.

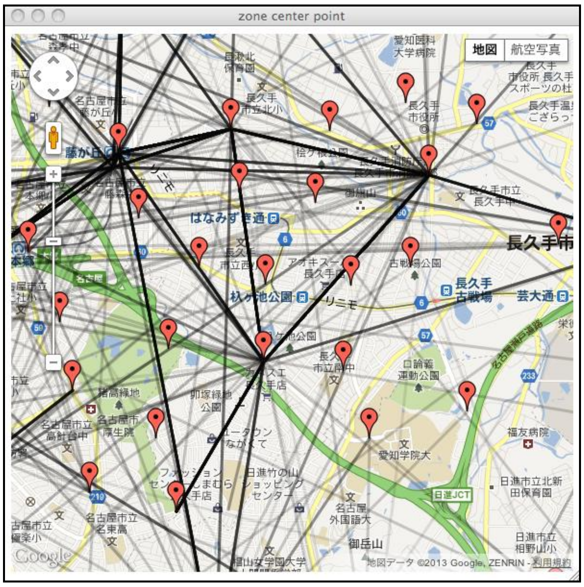

Fig. 4. Origin and destination of extracted person trip survey data

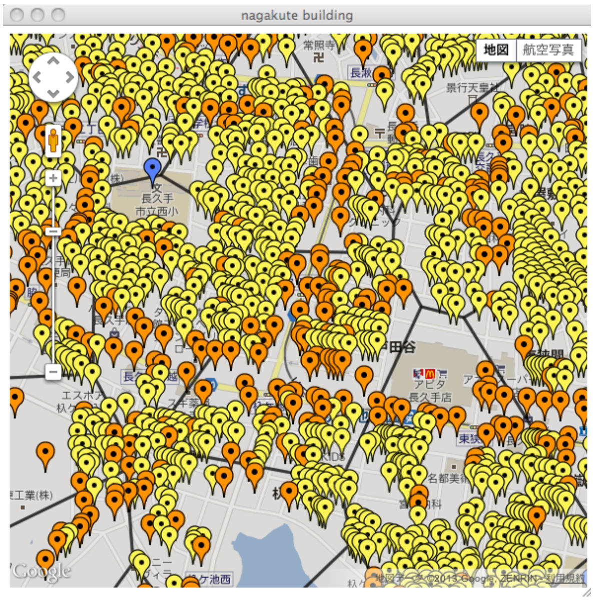

Fig. 5. Building locations in the part of the target area

Table 2. Correspondence table between purpose and building type of destination

Purpose	Building type of destination
Go to school	School
Go back to home	Home
Other purpose	Business place

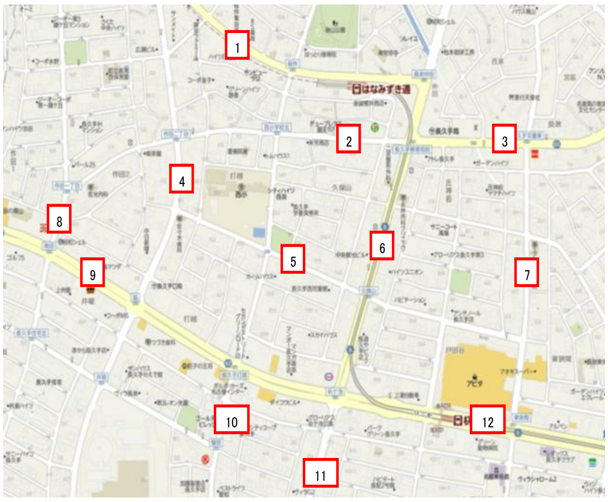

Fig. 6. Locations where traffic counts surveys are conducted

4.2. Result

For validation, three estimations are compared. One is the estimation of the proposed method, which contains the data assimilation and the dispersion of the destinations. Second is the estimation with data assimilation, but without the dispersion of the destinations, and the other is the estimation without both data assimilation and the dispersion.

Table 3 shows the result of the aggregation of the number of people who pass through the observation points and matching ratio between the observation data and the estimation. Matching ratio means the percentile of the estimation divided by the observation value. The estimations with data assimilation, the right two estimations, are more accurate than the estimation without data assimilation, the left one, in terms of the matching ratio. Especially, the estimation without data assimilation is completely out of the observation data at observation points, 2, 3, 5, 9, 10, 11, and 12. On the other hand, much improvement can be seen in the two estimations with data assimilation and most of matching ratios are better than the ones of the estimation without data assimilation. As for the dispersion of destinations, the estimation with it is more accurate at most of the observation points, especially, 2, 3, 5.

Table 3. Result of estimation and matching ratio

Observation point No.	Observation data	estimation					
	walk and bicycle	without both data assimilation and dispersion of destinations		with data assimilation and without dispersion of destinations		with both data assimilation and dispersion of destinations	
		estimation	matching ratio	estimation	matching ratio	estimation	matching ratio
1	1,470	1512	**102.9**	1032	**70.2**	1060	**72.1**
2	421	2,154	**511.6**	803	**190.7**	392	**93.1**
3	906	2,623	**289.5**	1,342	**148.1**	1,113	**122.8**
4	1,021	632	**61.9**	822	**80.5**	685	**67.1**
5	449	0	**0.0**	124	**27.6**	259	**57.7**
6	2,021	2457	**121.6**	1458	**72.1**	1735	**85.8**
7	724	4161	**574.7**	669	**92.4**	664	**91.7**
8	2,341	1876	**80.1**	1455	**62.2**	1408	**60.1**
9	3,363	126	**3.7**	2121	**63.1**	2278	**67.7**
10	847	28	**3.3**	597	**70.5**	563	**66.5**
11	945	2238	**236.8**	941	**99.6**	869	**92.0**
12	2,696	126	**4.7**	1550	**57.5**	1834	**68.0**
total	**17,204**	**17,933**	-	**12,914**	-	**12,860**	-

Figure 7 shows hourly RMSE (root mean square error) of the observation data and the estimation. It can be shown that the estimation without data assimilation is much worse than the estimations with data assimilation. The estimation with both data assimilation and dispersion of destinations is little better than the one with data assimilation and without dispersion. Comparing these two estimations, RMSE is almost the same at 9, 15, and 17 o'clock. At such time, RMSE is not improved because children often move by walk or bicycle and person trip survey often miss the movements of children. Actually, many children are observed at the time at observation points, 5 and 9, where an elementary school is near. Other data for estimation model is required to estimate these movements. For example, GPS data can be an additional data to analyze such movements.

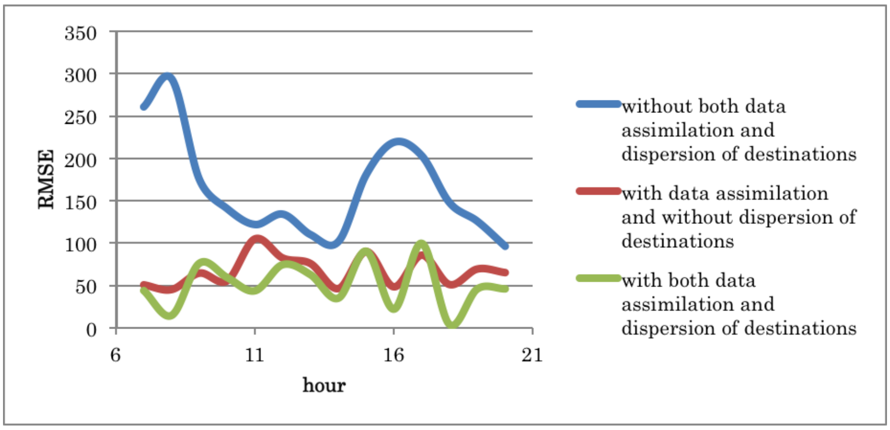

Fig. 7. Hourly RMSE of the observation data and the estimation

5. Conclusions and future works

In this study, a method that estimates people flow using particle filter has been proposed. By the proposed method with person trip survey, building data, and traffic count, people flow fitting to the observation data can be estimated in a city.

The experimental results show that the estimation of people flow becomes more accurate with data assimilation and dispersion of destinations. At most of hours and observation points, the estimation of people flow fits better to the observation data, traffic count data, with data assimilation and dispersion of destinations using building data.

However, some errors have remained in this study. Movements likely by children cannot be estimated accurately because person trip survey is a questionnaire style and such movements are not perfectly recorded. In order to tackle this problem, combining other movement data such as GPS data is worth considering. Moreover, as building types, only three types, home, school and business place, are taken into account in this study because some business places are not obviously separated. This problem should be also improved so that the construction of more accurate people flow simulation model results in better estimation. To summarize points that should be improved, the improvement of the accuracy of the simulation model is important.

In the future works, the online estimation of people flow with real-time observation is desired for surveillance and traffic accident avoidance. It requires the reduction of calculation cost and the more robust estimation system. As calculation cost and accuracy are generally trade-off, to find the most efficient point is a important task.

References

Akgun, V., Erkut, E., and Batta, R. (2000) On finding dissimilar paths, European Journal of Operational research 121, pp. 232-246.

Ben-Akiva, M., M.J. Bergman, A.J. Daly, and R. Ramaswamy (1984) Modelling Inter Urban Route Choice Behaviour. Proceedings of the 9th International Symposium on Transportation and Traffic Theory, VNU Press, Utrecht, pp. 299–330.

De la Barra, T., B. Perez, and J. Anez (1993) Multidimensional Path Search and Assignment. Proceedings of the 21st PTRC Summer Meeting, pp. 307–319.

Dial, R.B. (1971) A probabilistic multipath traffic assignment model which obviates path enumeration. Transportation Research, 5(2), pp. 83–111.

DRM (Digital Road Map) http://www.drm.jp/english/drm/e_index.htm

Evensen, G. (1994) Sequential data assimilation with a non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, Journal of Geophysical Research, 99, 10143–10162.

Gordon, N. J., Salmond, D. J. and Smith, A. F. M. (1993) Novel approach to non-linear/non-Gaussian Bayesian state estimation, IEE Proceedings F, 140, 107–113.

Kalman, R. E. (1960) A new approach to linear filtering and prediction problems, Journal of Basic Engineering, 82, 35–45.

Kitagawa, G. (1996) Monte Carlo filter and smoother for non-Gaussian nonlinear state space models, Journal of Computational and Graphical Statistics, 5, 1–25.

Lombard, K., and Church, R.L. (1993) The gateway shortest path problem: Generating alternative routes for a corridor location problem. Geographical Systems 1, pp. 25-45.

Y. Sekimoto, R. Shibasaki, H. Kanasugi, T. Usui, Y. Shimazaki (2011) PFLOW: Reconstruction of people flow by recycling large-scale fragmentary social survey data, IEEE Pervasive Computing 10(4), pp.27-35.