

# Travel path and transport mode identification method using “less-frequently-detected” position data

T Shimizu<sup>1,5</sup>, T Yamaguchi<sup>2</sup>, H Ai<sup>3</sup>, J Kawase<sup>4</sup> and Y Katagiri<sup>3</sup>

<sup>1</sup> Professor, Department of Tourism Science, Tokyo Metropolitan University, Japan

<sup>2</sup> Hokkaido Railway Company, Japan

<sup>3</sup> Assistant Professor, Department of Tourism Science, Tokyo Metropolitan University

<sup>4</sup> Graduate Student, Department of Tourism Science, Tokyo Metropolitan University

E-mail: t-sim@tmu.ac.jp

**Abstract.** This study aims to seek method on travel path and transport mode identification in case positions of travellers are detected in low frequency. The survey in which ten test travellers with GPS logger move around Tokyo city centre was conducted. Travel path datasets of each traveller in which position data are selected every five minutes are processed from our survey data. Coverage index analysis based on the buffer analysis using GIS software is conducted. The condition and possibility to identify a path and a transport mode used are discussed.

## 1. Introduction

In the tourism research field, GPS tracking data has recently been considered as a powerful tool for identifying locations and travel paths of travellers. Shimizu obtained travel path data of foreign rental car drivers using GPS loggers in Central Hokkaido region, Japan, in order to assess the availability of rental car travel by foreigners [1]. McKercher *et al* used GPS data loggers to analyze the difference between behavioural pattern of a first visitor and that of a repeating visitor in Hong Kong [2]. Hallo *et al* examined GPS technology to track walking path of nature-based tourists in Virginia [3]. Yabe *et al* reviewed analytical methods on tourists' activities and behaviours using GPS technology [4]. Even after these studies, methods to apply GPS technology for understanding travellers' behaviour in tourism areas have not been established.

The studies above gave dedicated GPS loggers to sample travellers. In these cases, it is impossible to obtain data from an unspecified number of samples. Mobile phone companies have recently obtained a huge volume of position data of their users through GPS equipped mobile phones in order to utilize to various marketing analyses. The use of such data in tourism industry has also been anticipated.

From position and its time stamp data obtained by GPS technology, the location of an activity including the information of time spent, and travel path between locations of activity can be estimated effectively. This study pays attention to the latter, especially how we identify a transport mode. Zenji *et al* and Nakayama *et al* developed a method on transport mode identification using GPS logger [5][6]. In these studies, positions were detected in high frequency, every five or ten seconds. If we use a dataset provided by mobile phone companies, positions may not be detected frequently (e.g. every five minutes) due to the limitation of communication capacity. In this case, methods proposed in these studies cannot be applied effectively.

This study aims to seek a method on travel path and transport mode identification in case positions of travellers are detected in low frequency. Ono *et al* already studied trip pattern estimation method using less-frequently-detected GPS data [7]. However, the transport mode was not yet identified.

<sup>5</sup> To whom any correspondence should be addressed.



The outline of the survey is explained in Chapter 2. Data processing and analytical method are explained in Chapter 3. Some results of the analysis are explained in Chapter 4. Chapter 5 contains the conclusions.

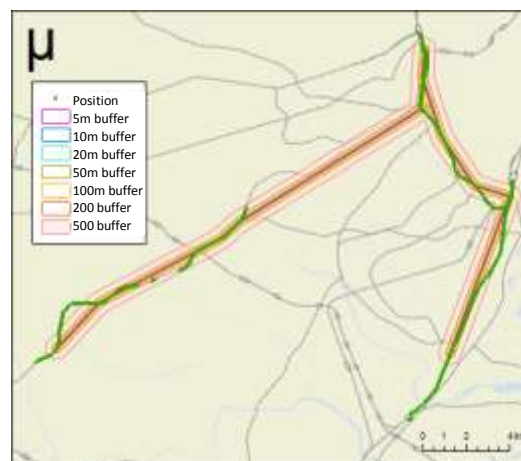
## 2. Survey

The survey in which ten test travellers with GPS logger (Black Gold 1300, Qstarz International Co. Ltd.) move around Tokyo city centre was conducted in December 2012 and January 2013. Each traveller was asked to visit designated zones by designated transport modes and, to take a tour in the zones for several hours. The GPS logger obtained position data every one second (hereinafter this is called as “one second travel path”). 14 zones and three transport modes, walking, railway (surface and underground) and bus were selected. A test traveller was asked to report its travel path and transport modes used in the designated map. We can exactly identify a travel path and transport modes by this map.

## 3. Data processing and analytical method

30 different travel path datasets of each traveller in which position data are selected every five minutes (hereinafter this is called as “five minutes travel path”) were processed. Each one second travel path and five minutes travel path were separated on the basis of transport mode.

In the GIS software (ArcGIS), seven buffer widths (5m, 10m, 20m, 50m, 100m, 200m and 500m) were created for each five minutes travel path (Figure 1). Here, one evaluation index, coverage, is introduced. Coverage means the content percentage of position data of one second travel path in the focused buffer width. If coverage in one five minutes travel path of one transport mode is close to 100%, the travel path and the transport mode used can definitely be identified.



**Figure 1.** Buffer analysis in GIS software

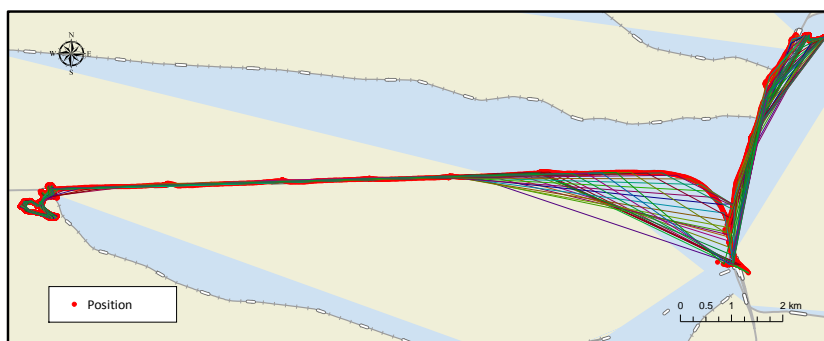
## 4. Results

### 4.1. Coverage analysis by traveller

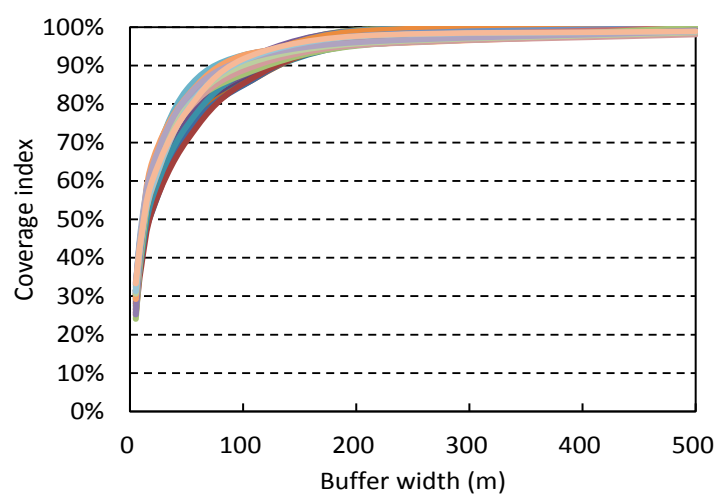
Figure 2 shows the distribution of 30 five minutes travel paths by one traveller. This traveller starts in Kichijyoji zone and takes a tour in Kichijyoji and Ikebukuro zones and move between zones by surface railway. There is a long curve section on surface railway and travel paths at this section vary each other. However, we can guess by travel speed information between detected positions that this traveller may use railway. Figure 3 shows the coverage index by buffer width. It is obvious that about 90% position data of the one second travel path are included in 100m buffer width.

Figure 4 shows the distribution of 30 five minutes travel paths by another traveller. This traveller starts in Aoba ward, Yokohama and takes a tour in Shinjuku and Kawasaki zones and move between zones by surface and underground railway. Figure 5 shows the coverage index by buffer width. Compared with the case of previous traveller, less than 80% position data of the one second travel

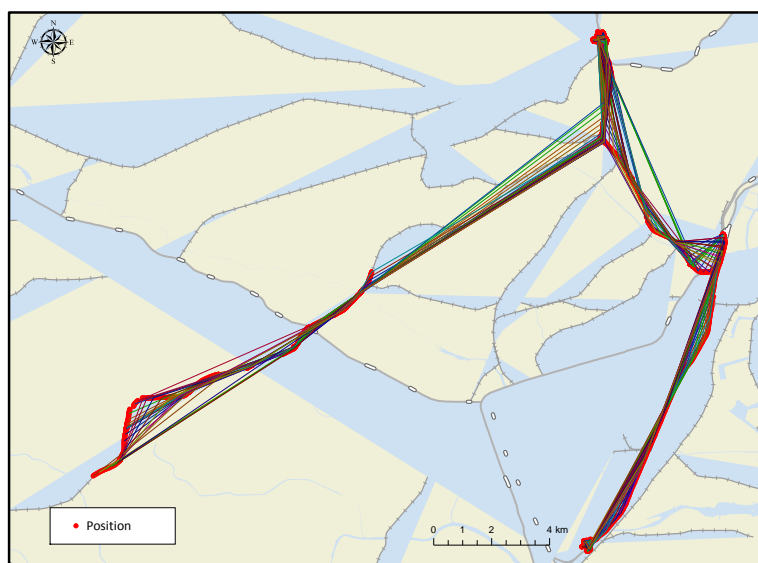
path are included in 100m buffer width. Even if the buffer width is 500m, the coverage ranges from 85% to 95%.



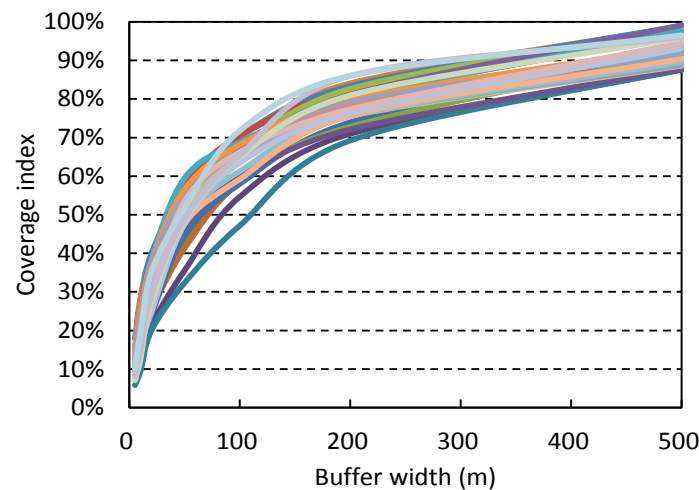
**Figure 2.** 30 five minutes travel paths by traveller A.



**Figure 3.** Coverage index by buffer width (traveller A).



**Figure 4.** 30 five minutes travel paths by traveller J.



**Figure 5.** Coverage index by buffer width (traveller J).

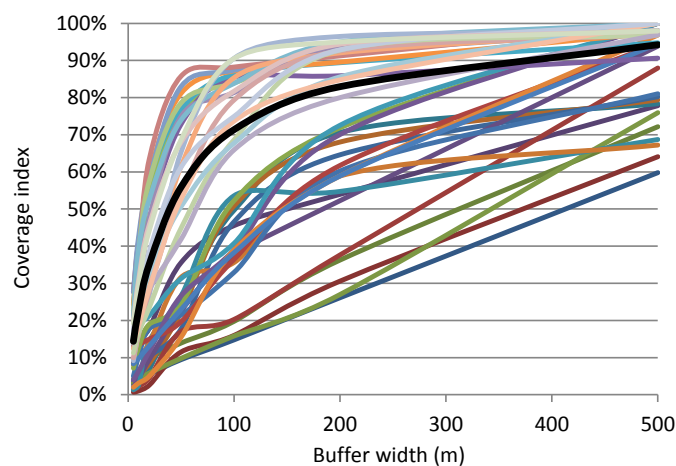
#### 4.2. Coverage analysis by transport mode

Figure 6, 7 and 8 shows the coverage index by buffer width in bus, railway and walking respectively. Black solid line means the average of all five minute travel paths of all travellers.

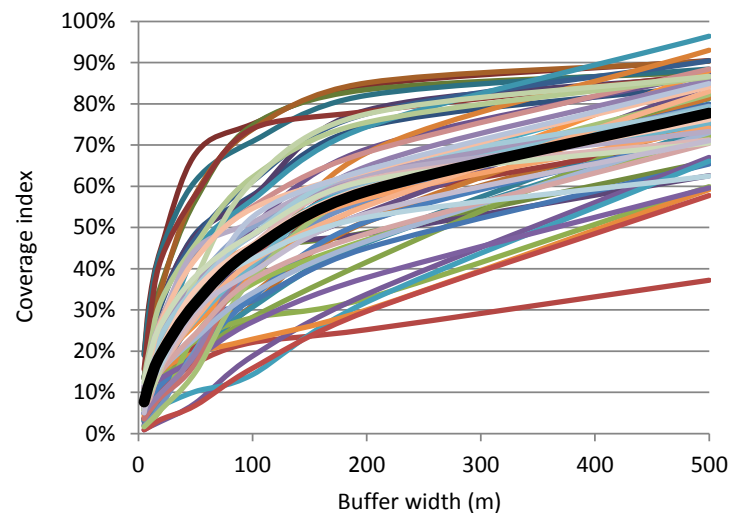
Averagely, 70% position data of the one second travel path by bus are included in 100m buffer width. While, there are many cases in which coverage index is less than 50%. These less coverage indexes are mainly caused by cases in which a bus route is tortuous. If we consider a speed between positions, we can distinguish from walking. However, it is hard to distinguish bus from taxi or passenger car in reality.

In average, only 40% position data of the one second travel path by railway are included in 100m buffer width. Besides, coverage index varies according to shape of a route. However, speed between positions by railway should be larger. In the end, the use of railway and its path can be identified by speed information even if coverage index is smaller.

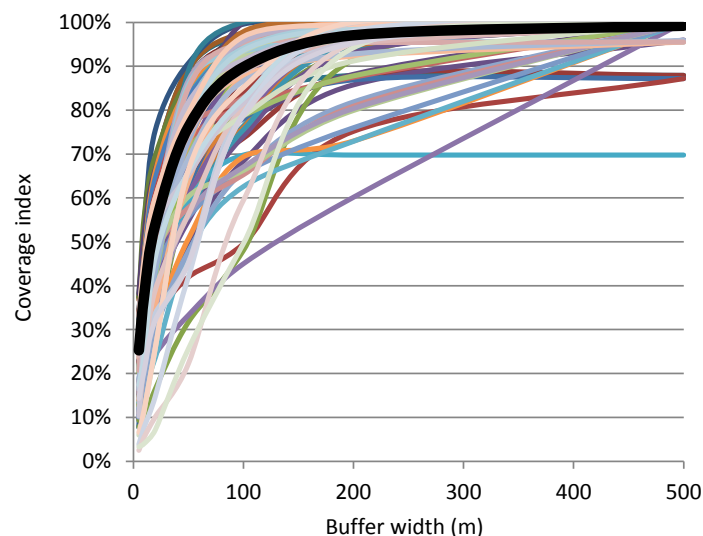
On the contrary, more than 85% position data of the one second travel path by walking are included in 100m buffer width, averagely. Lower coverage index in some cases is caused by the effect of high-rise building. Despite this higher coverage index, it is hard to identify the exact travel path.



**Figure 6.** Coverage index by buffer width (bus).



**Figure 7.** Coverage index by buffer width (railway).



**Figure 8.** Coverage index by buffer width (railway).

## 5. Conclusions

This study proposed method on travel path and transport mode identification in case positions of travellers are detected in low frequency. Through the original survey and the proposal of coverage index, the condition and possibility to identify a path and a transport mode used were discussed. In further study, the condition of the identification should be more clarified.

## References

- [1] Shimizu T 2010 Promotion of use of rental car for inbound tourist *Traffic Engineering* **45** 20–23 (in Japanese)
- [2] McKercher B, Shoval N, Ng E and Birenboim A 2012 First and repeat visitor behaviour: GPS tracking and GIS analysis in Hong Kong *Tourism Geographies* **14** 147-161
- [3] Hallo J C, Beeco J A, Goetcheus C, McGee J, McGehee N G and Norman W C 2012 GPS as a method for assessing spatial and temporal use distributions of nature-based tourists *Journal of Travel Research* **51** 591-606
- [4] Yabe N, Arima T, Okamura Y and Kadono A 2010 An agenda on tourist activity survey using GPS and investigation of analysis methods *The International Journal of Tourism Science*. **3** 17-30 (in Japanese)

- [5] Zenji T, Horiguchi R, Akahane H and Komiya T 2005 Development of a method for estimating transportation modes with handy GPS equipment *Proceedings of the 4th ITS Symposium*. 6 pages (in Japanese)
- [6] Nakayama H, Shimizu T and Nakajima H 2012 A methodology to estimate tourists's travel mode using GPS data *TER-12-66 IEE Japan*. 19-22
- [7] Ono N, Sekimoto Y, Nakamura T, Horanont T and Shibasaki R 2012 Estimation of routes using long-term GPS data in Tokyo *Proceedings of the 21st conference of GIS Association of Japan*. 4 pages (in Japanese)