PEDESTRIAN FACILITY USAGE MONITORING USING MULTIPLE SOURCES OF DATA

P. WEPULANON [a], W. H. K. LAM [b] and A. SUMALEE [c]
[a,b,c] *Department of Civil and Environmental Engineering,*
The Hong Kong Polytechnic University, China
Email: w.piyanit@gmail.com

ABSTRACT

This paper aims to investigate the possibility of developing facility usage monitoring system using multiple sources of information. The system is able to provide fundamental information for analyzing pedestrian usage behavior including pedestrian flows, speeds, and densities. In particular, MAC address data are considered as complementary data sources which could passively obtain pedestrian walking time from individuals. A number of data processing steps are introduced to tackle the particular challenges of pedestrian travel time estimation using MAC address data. The system performance is evaluated using a case study on The Hong Kong Polytechnic University. Practicality of the system is evaluated in terms of estimation accuracy and the penetration rate of MAC address data. The results show that pedestrian travel time can be estimated from the passive MAC address data with acceptable accuracy.

Keywords: facility usage monitoring, pedestrian travel time estimation, MAC address data, multiple data sources, data filtering

1. INTRODUCTION

An initial step for pedestrian facility development is to understand pedestrian usage behavior. In terms of pedestrian traffic conditions, a level-of-service standard is generally used to describe the traffic state of a pedestrian facility. The level of services can be determined based on three fundamental parameters: pedestrian flows, walking speeds, and pedestrian densities. The parameters are crucial for facility management, and developing pedestrian models for future planning.

To derive such pedestrian characteristics, a traditional method is conducting surveys. As the manual survey is labor-intensive, video recording equipment has been employed for data collection. Several image processing techniques have been developed to extract pedestrian characteristics from the video data. First, pedestrian flows are obtained when each pedestrian crossed a predetermined section in the video image. Second, pedestrian density profiles can be extensively extracted by counting the number of pedestrian in a predetermined area at each time interval such as 5-second intervals. The last parameter is average walking speed or pedestrian travel time. Automatic travel time extraction is more challenging as individual pedestrians need to be matched across predetermined sections. The pedestrian re-identification involves intensive computational resources (Liu et al., 2015) and the accuracy could be affected by various factors (e.g. camera angles and lighting environments). In the literature, indirect methods have been developed to estimate pedestrian travel time as a function of pedestrian flows or densities, and the capacity of pedestrian facility (Lam and Cheung, 1998).

The possibility to derive individual pedestrian travel time has been enabled with the advancement of sensing technologies. In particular, smart devices (e.g. smartphones, tablets) have been considered as potential sensing devices for human tracking due to the increasing popularity and the embedded sensing technologies. Positioning techniques have been developed to derive mobility information from smartphones including Global System for Mobile Communications (GSM) and Global Positioning System (GPS). The technologies are suitable for analyzing large scale mobility in outdoor environments (Wolf et al., 2001; Sila-Nowicka et al., 2016; Alexander et al., 2015; Jiang et al., 2017).

Nonetheless, accessing such data sources requires the direct permission from cellular network providers or smartphone users.

An alternative approach for deriving human mobility is monitoring Bluetooth and Wi-Fi enabled devices. The wireless technologies have been suggested for indoor mobility analysis in a smaller scale of study areas such as hospital and campus. Moreover, mobility data can be observed in non-intrusive ways without the requirement of direct participation from the device users. Basically, Bluetooth/Wi-Fi monitors can be installed at study locations in order to capture the network traces generated from Bluetooth/Wi-Fi enabled devices. Each device can be recognized by a Media Access Control (MAC) address which is assigned as its unique identification number. Therefore, the mobility of individual devices can be analyzed in both spatial and temporal dimensions with multiple monitors installed at different locations of interest. MAC address data has been considered as a complementary data source for tracking both vehicle (Haghani et al., 2010; Bhaskar et al., 2015) and pedestrian mobility (Abedi et al., 2015). A study found that the possibility of acquiring Wi-Fi MAC addresses from general travelers tends to be higher than the Bluetooth one (Abedi et al., 2015).

In this paper, a pedestrian facility monitoring system is proposed. Pedestrian flow and density profiles can be obtained by installing video cameras at strategic locations. In the same way, Bluetooth and Wi-Fi monitors can be installed to collect MAC address data which are incorporated to tackle the limitations of pedestrian travel time estimation using video data. Moreover, using MAC address data could be more feasible when there is a need to observe the time spent to traverse through some facilities such as crossing bridges, crossing tunnels, and skywalks. Consequently, pedestrian facility usage can be monitored over times in the long run.

Despite the advantages of MAC address data, particular characteristics in the data introduce extra difficulties in pedestrian travel time estimation. First, the spatial and temporal resolution of passive MAC address data is coarse and uncertain among different devices. Second, MAC address data usually provides partial data sets of the total pedestrians. Third, the environments of pedestrian facilities could provide significant noise in MAC address data. Data mining and filtering techniques for handling the challenges in MAC address data are proposed in this paper. The practicality of using MAC address data to estimate pedestrian travel time is evaluated in terms of estimation accuracy.

The rest of this paper is organized as follows. Section 2 describes a case study in this paper. Section 3 proposes the necessary data processing steps to extract pedestrian travel time from MAC address data. The numerical results are provided and discussed in Section 4. Finally, the conclusion and the further studies are summarized in Section 5.

2. DATA COLLECTION

In this study, a pedestrian crossing tunnel linking the main campus of Hong Kong Polytechnic University (PolyU) and the building of Faculty of Construction and Environment (Block Z) is used as a case study for developing the proposed pedestrian facility monitoring system. The tunnel is constructed under Chatham North Road, an eight-lane road where a heavy volume of vehicles can travel at high speeds in both directions. Figure 1 shows a simple topology of the tunnel.

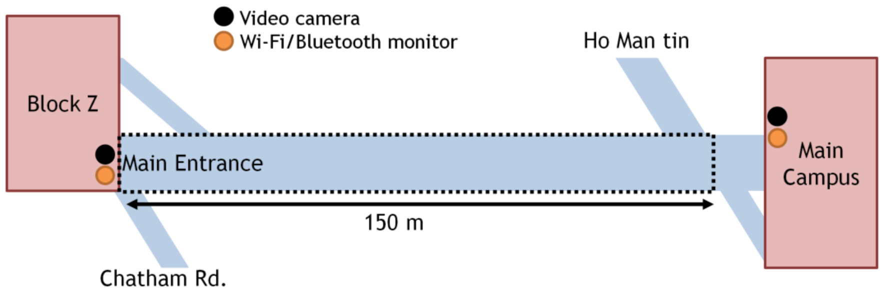


Figure 1. A pedestrian crossing tunnel under Chatham North Rd.

In order to monitor the tunnel usage, a video camera as well as a Bluetooth and Wi-Fi monitor are installed at each end of the tunnel. Due to privacy concerns, video cameras are installed to capture the top-view angle of pedestrians. Also, a median filter is applied to the video files so as to reduce the resolution of video images. Since the installation is for research purposes, the equipment should be installed in PolyU campus areas. Hence, this study assumes the crossing tunnel areas as demonstrated in Figure 1. The study area is about 150 m long.

The video streams can be transmitted to a computer server. Practical image processing techniques can be adopted to extract pedestrian flow profiles at the two locations. In the same way, Bluetooth and Wi-Fi detection data can be transmitted to the server for pedestrian travel time estimation. The following sections will focus on the methodology of using MAC address data for pedestrian travel time estimation as image processing techniques can be directly adopted for flow profiles extraction.

3. PEDESTRIAN TRAVEL TIME ESTIMATION USING MAC ADDRESS DATA

This section introduces the attributes of MAC address data acquired by Bluetooth and Wi-Fi monitors. The primary characteristics in MAC address data are described together with the necessary data processing steps for pedestrian travel time estimation.

3.1 Background and raw data

Owing to the underlying mechanisms of Bluetooth and Wi-Fi communication, the Bluetooth/Wi-Fi enabled devices such as smartphones usually broadcasts wireless communication messages. The broadcasted messages can be detected by Bluetooth/Wi-Fi monitors. The monitors continuously listen to the communication messages within the operating area and record a set of information including (i) MAC address, (ii) detection timestamp, and (iii) Received Signal Strength Indicator (RSSI) which can be used to estimate the distance between the detected device and the monitor. It is noteworthy that Wi-Fi monitors are able to discriminate the messages broadcasted by networking devices e.g. access points. Therefore, the device type (access points or mobile devices) is another data attribute provided by Wi-Fi monitors. In this study, only Wi-Fi MAC address data from mobile devices are used for pedestrian travel time estimation since networking devices are assumed to be stationary. Table 1 and Table 1 shows examples of MAC address data derived from a Wi-Fi monitor.

Table 1. Examples of MAC address data from a Wi-Fi monitor

MAC address	Timestamp	RSSI	Type	Monitor ID
C4:0B:CB:34:DE:18	16/6/17 9:30:00	10	AP	B
F4:4E:FD:38:DD:5B	16/6/17 9:30:00	12	AP	B
38:A4:ED:6D:33:A8	16/6/17 9:30:01	40	MD	B
00:18:13:CD:54:0F	16/6/17 9:30:01	14	MD	B

Despite the availability of MAC address data, estimating pedestrian travel time is the most challenging task. First, the operating range of Bluetooth/Wi-Fi monitors may cover a large area. Significant noise from surrounding environments can be included in MAC address data i.e. the mobile devices from passing vehicle on Chatham North Road in this study. Second, the spatial and temporal resolution of MAC address data is coarse. A mobile device may broadcast a message in every few seconds or several minutes based on a wide range of factors. MAC address data of the devices which are present at the operational area for a considerable time could be detected as multiple records. Identifying the precise location of a mobile device in the operating area of a monitor is a burdensome task based on the coarse MAC address data. In the literature, only the first or the last MAC address data detected by each monitors is considered for travel time estimation (Haghani et al., 2010; Abedi et al., 2015; Bhaskar et al., 2015). This study proposes a number of data processing steps for pedestrian travel time estimation as can be shown in Figure 2.

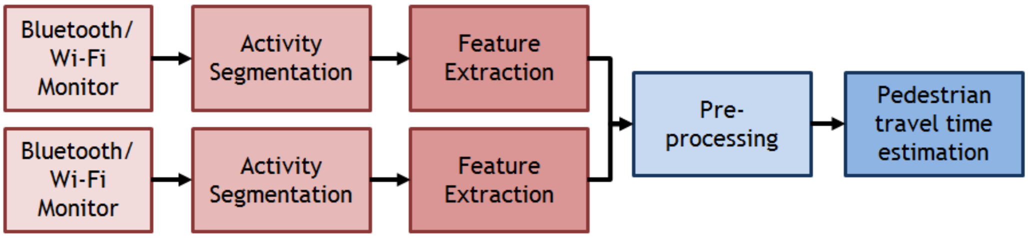


Figure 2. A conceptual diagram of the data processing steps

3.2 Activity segmentation

A pedestrian may conduct multiple activities at a particular location in a day. In this case, pedestrians could cross the pedestrian tunnel for multiple times. As a result, several detection records of the same mobile device can be detected from different trips. Figure 3 shows examples of MAC address data of a known smartphone detected by a Wi-Fi monitor. The smartphone user crossed the tunnel twice within 5 hours.

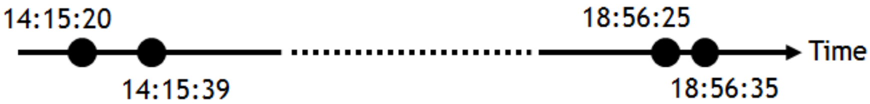


Figure 3. MAC address data of a known device over time of a day

It can be seen that the MAC address data of each activity can be identified by considering the time duration between consecutive detection. A threshold can be determined from the statistical analysis of MAC address detection frequencies. The distribution of detection frequency is shown in Figure 4.

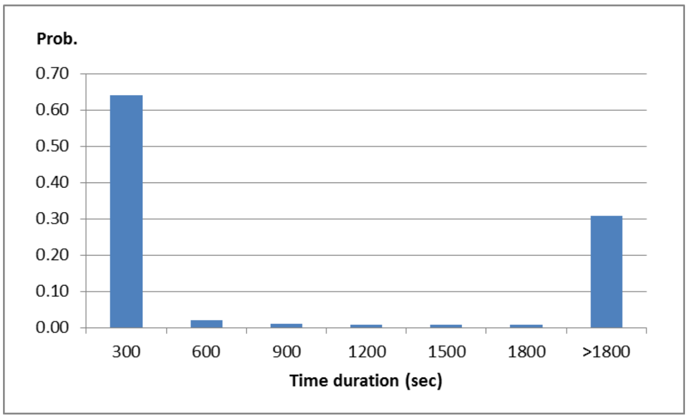


Figure 4. Probability distribution of detection frequencies

Most of the MAC addresses can be detected within 5 minutes. The duration is expected as the common detection frequency when a device is placed at the monitoring area. Another group of MAC address data can be detected more than half an hour, since a pedestrian could conduct several visits at the same place. A 10-min threshold is applied based on an assumption that the minimum time duration to revisit the same location is 10 minutes.

3.3 Feature extraction

To facilitate the further processing steps, a new data structure can be constructed to summarize the records detected from a mobile device during an activity (i.e. a crossing). Table 2 summarizes the basic features which could be extracted from a monitor during an activity.

Table 2. Cumulative features of each MAC address during each activity

Features	mac	monitor_id	t_first	t_last	t_duration	data_count	avg_rssi	max_rssi
Description	MAC address	Monitor ID	First detection time	Last detection time	Detected duration	Number of data points	Average RSSI	Maximum RSSI

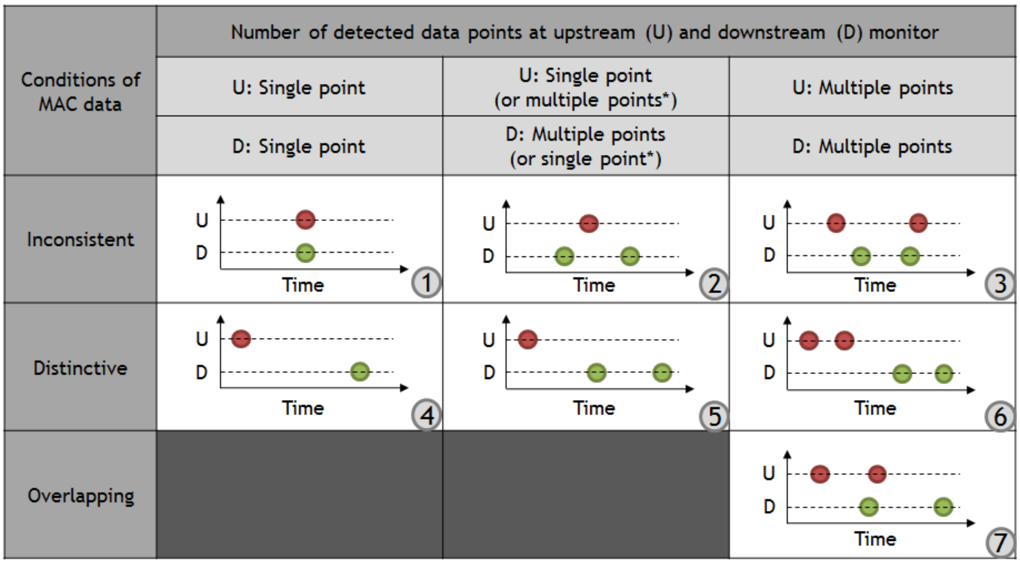
3.4 Preprocessing

The mobile devices from other sources, apart from crossing pedestrians, can be detected due to a wide operational range of Bluetooth/Wi-Fi monitors. The possible data sources consist of the mobile devices from (i) the passing pedestrians who did not make a crossing, (ii) the vehicles travelling on nearby roads, and (iii) the devices in PolyU campus where students and staffs may conduct other activities. Data preprocessing aims to retain the MAC address data which correspond to the pedestrian travel time across the tunnel.

First, the MAC addresses which were observed by only one monitor were removed since the data cannot represent the mobility across the tunnel. The valid MAC address data should be detected by both monitors within a time-window. A searching time-window can be determined by observing the upper bound of pedestrian travel time required to pass the operating area of both monitors. The time-window can be calculated as follows:

$$time\_window = [t\_first_A, ub\_duration_A + ub\_traveltime_{AB} + ub\_duration_B] \qquad (1)$$

where $ub\_traveltime_{AB}$ is the upper bound of travel time from location A to B, $ub\_duration_A$, and $ub\_duration_B$ are the upper bound of activity time duration at location A and B respectively. Basically, the activity at the end of pedestrian facility should be walking. However, the activity time duration at each location can be varied based on the type of facilities. In our case, the end of pedestrian tunnel at Block Z is also an elevator waiting area. As a result, the activity time duration at this location is usually longer, when a pedestrian travelled from the main campus to Block Z and waited for the elevators. In this paper, a 10-min time-window is determined from manual observations.



*An opposite case when a monitor detected multiple points and another one detected a single point

Figure 5. Patterns of MAC address data detected by 2 monitors

Since MAC address data cannot provide the fine-grained mobility tracking, the detected data points from a mobile device may not be sufficient for pedestrian travel time estimation. In addition, it is possible to detect the mobile devices on vehicles as the road is between the two monitors. The second preprocessing step is to identify the MAC addresses which can be employed for the estimation. Figure 5 demonstrates the possible MAC address detection patterns from the two monitors, supposed that a device is firstly detected by the upstream monitor, and then by the downstream one.

It can be observed that the patterns no.1-3 are unlikely to represent a pedestrian crossing the tunnel as the mobile device is detected by two monitors at the same time interval. The MAC address data correspond to such inconsistent patterns should be discarded.

3.5 Pedestrian travel time estimation

To this end, the remaining MAC addresses can be considered as the potential devices carried by the pedestrians crossing the tunnel. The most challenging task is to estimate the travel time based on available data points, and determine which data point is more reliable. It can be assumed that the more data points detected by a monitor can provide the more accurate entry/exit time at the monitor side. Therefore, the detection patterns no. 6 and 7 should be more reliable, followed by the patterns no.5 and 4.

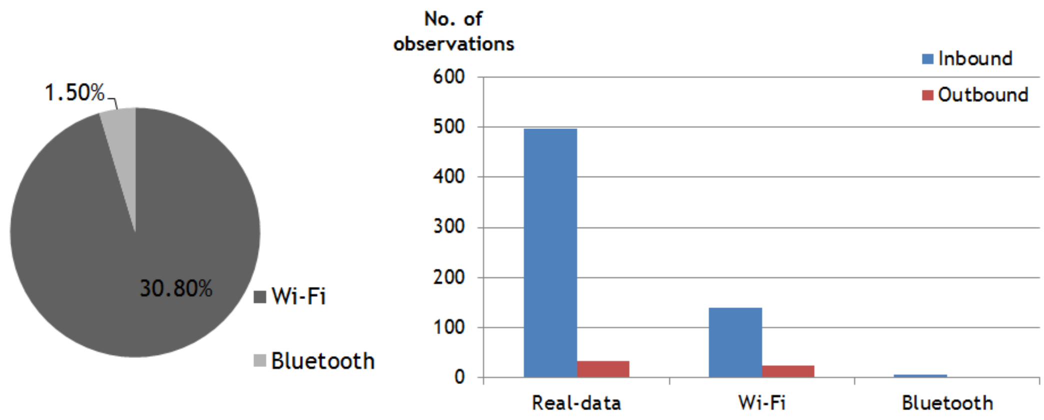
In this study, the data point detected with the strongest RSSI is assumed to be the most reliable entry/exit time at a monitor location. Hence, pedestrian travel time can be estimated as follows:

$$traveltime_{AB} = t\_\max rssi_B - t\_\max rssi_A \tag{2}$$

where $t\_\max rssi_A$ and $t\_\max rssi_B$ are the timestamp of MAC address data with the maximum RSSI value of the monitor A and B respectively. Finally, a travel time filtering technique can be adopted to remove outlier travel time data (i.e. Tam and Lam, 2011).

4. NUMERICAL RESULTS

To evaluate the system performance, pedestrian travel time was manually observed from video data, using a data set on 16th June 2017 during a morning peak time periods: 8-10 a.m. Figure 6a presents the penetration rate of Bluetooth and Wi-Fi devices compared with the actual data from manual observations, while Figure 6b shows the number of MAC addresses of the pedestrians travelling inbound (from the main campus to Block Z) and outbound.



(a) (b)

Figure 6. (a) MAC address penetration rate,
(b) the number of MAC addresses from each travel direction

The number of pedestrians travelling to Block Z is significantly greater than the outbound direction. This is expected since the travel demand to Block Z is usually high in the morning peak time period. For both travel directions, the number of detected Wi-Fi devices is around 31% of total crossing pedestrians, whereas the number of Bluetooth devices can be detected for only 1.5%. The results indicate that Bluetooth monitors can enhance the penetration rate of overall MAC address data. However, the stand-alone system may not be able to provide sufficient samples for travel time estimation. Figure 7 demonstrates the sample size of inbound pedestrians provided by Wi-Fi monitors. The samples are plotted by each 2-min time interval.

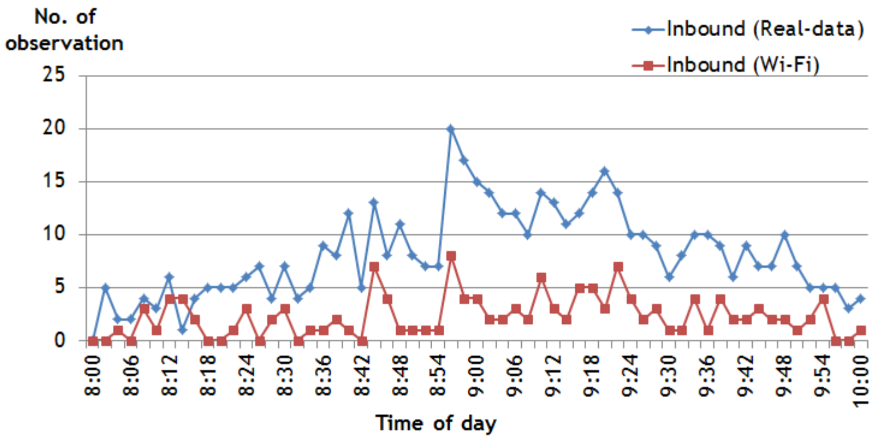


Figure 7. Number of unique Wi-Fi devices

During the 2-hour observation, there are 9 time intervals without the detection of Wi-Fi devices. Besides, the number of Wi-Fi devices is greater than observed pedestrian at a time interval during 8:12-8:14 a.m. Figure 8 shows the distribution of the estimated pedestrian travel time over the 2-hour time period, compared with actual observation data.

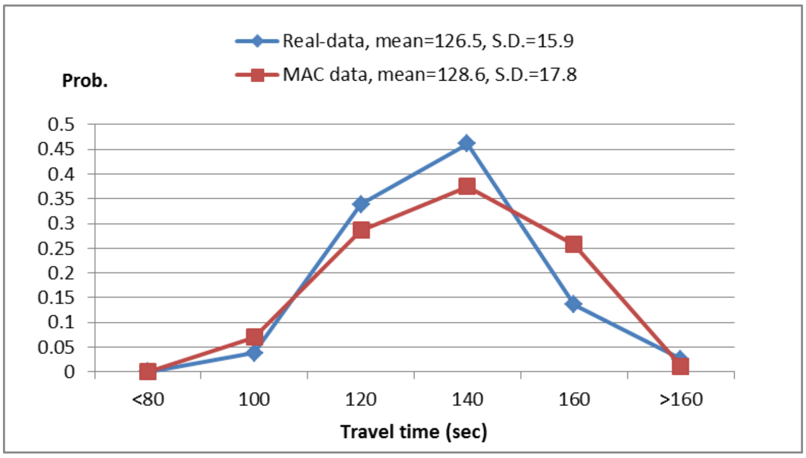


Figure 8. Distribution of pedestrian travel time over a 2-hour time period

The travel times estimated from MAC address data provide the higher mean and standard deviation then the actual data about 2 seconds. However, it can be observed from the distribution that the travel times during 100-140 seconds are tend to be underestimated while the travel time during 140-160 seconds are overestimated. The errors can cause by a wide detection range of MAC address monitors. This study assume that the maximum RSSI represents the most reliable MAC address data of each detected location. In fact, the devices can be detected at any location in the detection range. Threfore, the MAC address data with the strongest RSSI may not be captured when a pedestrian enters or exits the tunnel. The accuracy can be improved by developing the more promising methods to estimate the entry and exit time at each location.

5. CONCLUSION

This paper introduces a facility usage monitoring system using multiple sources of data. With the availability of MAC address data, pedestrian travel time can be estimated from individuals. The necessary data processing steps for developing facility usage monitoring system based on MAC address data are proposed to handle the particular data characteristics. The system performance is evaluated using a case study on a pedestrian crossing tunnel in PolyU campus. The results show that pedestrian travel time can be estimated from MAC address data with the acceptable accuracy. The proposed framework can be applied to other pedestrian facilities e.g. pedestrian crossing bridges and skywalks. In the future research, the algorithms for travel time estimation can be developed based on the patterns of MAC address data so as to enhance the valid travel time samples and improve travel time estimation accuracy.

Acknowledgement

The work described in this paper was supported by research grants from the Research Grants Council of the Hong Kong Special Administrative Region (Project No. PolyU 152057/15E and PolyU 5181/13E) and the Research Committee of The Hong Kong Polytechnic University (Project No. 1-ZVFJ)

6. REFERENCES

Liu, K., Ma, B., Zhang, W. and Huang, R. (2015) A spatio-temporal appearance representation for viceo-based pedestrian re-identification. *In Proceedings of the IEEE International Conference on Computer Vision*. pp.3810-3818.

Lam, W. H., and Cheung, C. Y. (1998) Pedestrian travel time functions for the Hong Kong underground stations–calibration and validation. *HKIE transactions*, 5(3), pp.39-45.

Wolf, J., Guensler, R. and Bachman, W. (2001) Elimination of the travel diary: Experiment to derive trip purpose from global positioning system travel data. *Transportation Research Record: Journal of the Transportation Research Board*, (1768), pp.125-134.

Alexander, L., Jiang, S., Murga, M. and González, M.C. (2015) Origin–destination trips by purpose and time of day inferred from mobile phone data. *Transportation Research Part C: Emerging Technologies*, 58, pp.240-250.

Siła-Nowicka, K., Vandrol, J., Oshan, T., Long, J.A., Demšar, U. and Fotheringham, A.S. (2016) Analysis of human mobility patterns from GPS trajectories and contextual information. *International Journal of Geographical Information Science*, 30(5), pp.881-906.

Jiang, S., Ferreira, J. and González, M.C. (2017) Activity-based human mobility patterns inferred from mobile phone data: A case study of Singapore. *IEEE Transactions on Big Data*, 3(2), pp.208-219.

Haghani, A., Hamedi, M., Sadabadi, K., Young, S., and Tarnoff, P. (2010) Data collection of freeway travel time ground truth with bluetooth sensors. *Transportation Research Record: Journal of the Transportation Research Board*, (2160), pp.60-68.

Abedi, N., Bhaskar, A., Chung, E., and Miska, M. (2015) Assessment of antenna characteristic effects on pedestrian and cyclists travel-time estimation based on Bluetooth and WiFi MAC addresses. *Transportation Research Part C: Emerging Technologies*, 60, pp.124-141.

Bhaskar, A., Qu, M., Nantes, A., Miska, M., and Chung, E. (2015) Is bus overrepresented in Bluetooth MAC scanner data? Is MAC-ID really unique?. *International Journal of Intelligent Transportation Systems Research*, 13(2), pp.119-130.

Tam, M. L., and Lam, W. H. (2011) Application of automatic vehicle identification technology for real-time journey time estimation. *Information Fusion*, 12(1), pp.11-19.