ESTIMATING PEDESTRIAN WALKING TIME ON CAMPUS BASED ON WI-FI DETECTION DATA

Z. LI [a], W. H. K. LAM [b], P. WEPULANON [c] and Z, QIN [d]
[a] *Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, China*
Email: lizhibin@seu.edu.cn
[b] *Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, China*
Email: pts_wep20@hotmail.com
[c] *Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, China*
Email: william.lam@polyu.edu.hk
[d] *School of Transportation, Southeast University, China*
Email: dongqinzhou@seu.edu.cn

ABSTRACT

Pedestrian travel time is important to the planning and design of pedestrian facilities particularly in high density populated urban areas. With the increasing use of portable electronic devices, the Wi-Fi detection data becomes a promising data source to estimate pedestrian activity patterns. The Media Access Control (MAC) address is a unique signature for each electronic device. In this study, we would make use of these Wi-Fi detection data to extract the pedestrian walking time of crossing a pedestrian tunnel that connects the Phase 8 building to the main campus of the Hong Kong Polytechnic University (PolyU). A data filtering framework is proposed to filter out noisy detections so as to extract the relevant Wi-Fi data. It follows with an efficient solution algorithm to estimate the pedestrian walking time from multiple detection records. Both the means and the variations of walking time are analyzed. The temporal characteristics of pedestrian flow patterns are discussed.

Keywords: Wi-Fi; MAC address; Filtering; Walking time

1. INTRODUCTION

Pedestrian activities are important to the design and construction of pedestrian facilities particularly in high density populated urban areas in Hong Kong (Lam et al., 2000). Pedestrian walking time is considered a good indicator of the level of service offered by a walkway (Al-Azzawi et al., 2007; Fitzpatrick et al., 2006; Yi et al., 2015). In addition, walking time is also important in the understanding of route choice behaviors of pedestrians (Hoogendoorn et al., 2004). Thus, estimating pedestrian walking time is helpful to improve the planning and design of walking facilities.

The most traditional method of measuring pedestrian walking time is to designate investigators following pedestrians and measure walking time. However, such method requires a large amount of labor work and is considered low efficient and costly. The video-based pedestrian tracking is another method of exploring walking time (Hoogendoorn et al., 2003; Wang et al., 2014; Teknomo et al., 2016). But this method works only for short sidewalk due to the limitation of vision coverage. New technologies such as facial recognition could help identify pedestrians from multiple videos and estimate walking time (Liu et al., 2015; Dewan et al., 2016). However, pairing of pedestrian faces may not be 100% accurate and video image processing is very complex and expensive. Recently, Wi-Fi detection has becoming a novel technology for pedestrian activities analysis (Kim et al., 2013; Abedi et al., 2013; Danalet et al., 2014; Wepulanon et al., 2015; Poucin et al., 2016). The MAC address detected by Wi-Fi scanner is a unique signature for each electronic device, supporting more detailed analysis of pedestrian activities.

Though with the aforementioned advantages, Wi-Fi detection-based pedestrian activity analysis contains several challenges. First, in addition to pedestrians' portable smart devices, Wi-Fi detection may include data from other electronic devices such as printer, driving video recorder, motor vehicle, etc. How to perform the data filtering to eliminate noises but keep true records is a big challenge. Second, only Wi-Fi enabled devices can be detected so that the Wi-Fi detection data may not cover all pedestrians. The ratio of number of detected devices and pedestrian count is hard to investigate and may vary in time and space. It raises challenges in the validation of Wi-Fi detection data filtering. Last but not least, Wi-Fi detection has inherent uncertainties. There is no clear pattern of when and where an electronic device is detected if entering the detection range. Signal strength indicator (RSSI) is obtained which may indicate the relative distance between the device and Wi-Fi scanner. But there are several environmental complexities affected the value of RSSI (Kim et al., 2013).

Previous studies have estimated pedestrian activities on campus using Wi-Fi connection data. But they used the university Wi-Fi logon data which only contain authenticated records from students or staffs of the university (Danalet et al., 2014; Wepulanon et al., 2015). Thus, data filtering is actually not needed. In some other studies, the Wi-Fi detection data was used to infer vehicular travel time between intersections. Wi-Fi scanners are installed at intersections to collect the MAC address data from vehicles or electronic devices carried by people. To the best of our knowledge, few studies have particularly focused on estimating pedestrian walking time using Wi-Fi detection data. The aforementioned challenges in Wi-Fi data-based pedestrian activity analysis were not well addressed.

The primary objective of the study is to propose a procedure for estimating pedestrian walking time on PolyU campus based on Wi-Fi detection data. A data filtering procedure was proposed to eliminate noisy detections but keep valid records. The data filtering results were compared with pedestrian count extracted from video records. Then pedestrian travel walking was estimated based on Wi-Fi detection data. This study provides an efficient method for acquiring pedestrian activities using existing Wi-Fi infrastructure so as to improve the usage and design of the walking facilities on the PolyU campus.

2. DATA SOURCE

In this study, two Wi-Fi scanners were installed on both sides of a pedestrian tunnel that connects the Phase 8 building to the main campus of the PolyU (denoted as Block Z and Y Figure 1). The Wi-Fi scanner can collect the information from a Wi-Fi enabled electronic device and a Wi-Fi connection record is generated. The Wi-Fi data set was derived from PolyU Wi-Fi system. The data consists of the information about timestamp, Net ID, Wi-Fi MAC address, signal strength (RSSI), vender, and the access point identification number (AP No.). Table 1 shows examples of Wi-Fi detection records.

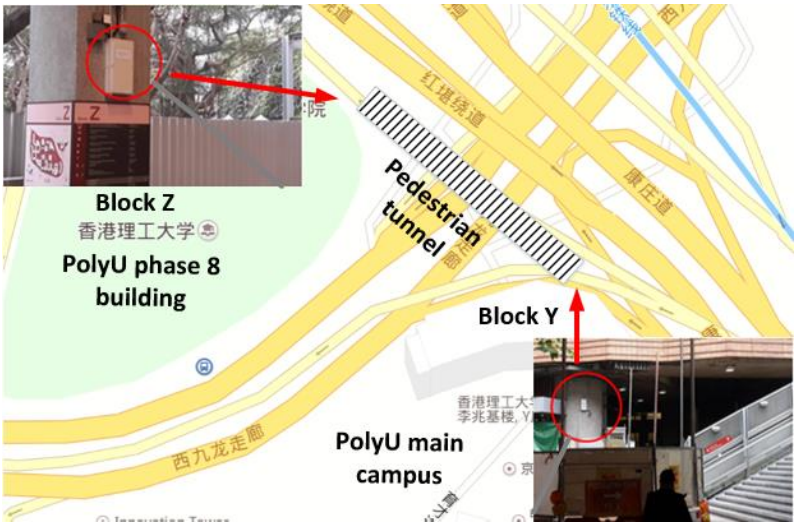


Figure 1 Illustration of Pedestrian Tunnel and Wi-Fi Scanner Location

Table 1 Examples of Wi-Fi detection records

ID_frame	MAC	TimeStamp	Location	RSSI	SSID	Vendor
342330	00:00:A1:F2:xx:xx*	2017/3/8 15:42	Z	0	NA	InPro Comm
342539	00:00:A1:F2:xx:xx	2017/3/8 15:43	Z	4	NA	InPro Comm
343042	00:00:A1:F2:xx:xx	2017/3/8 15:44	Z	3	NA	E.F. Johnson
343754	00:00:A1:F2:xx:xx	2017/3/8 15:45	Y	5	NA	E.F. Johnson
344702	00:00:A1:F2:xx:xx	2017/3/8 15:45	Y	11	NA	Airgo Networks

* The last four digital numbers are concealed.

The accurate detection range of the Wi-Fi scanner is unknown. It is possible that the detection area of both scanners may be overlapped. We conducted an experiment in which a testing smartphone was carried by an investigator walking from Block Y to Z. The detection records in a time series with RSSI information are shown in Figure 2. The time of the last detection at Block Y is after the time of the first detection at Block Z, which suggests the existence of the overlapping zone. This could bring challenges to the calculaiton of pedestrian walking time. The signal strength may indicate the distance between the device and Wi-Fi scanner. But it is also affected by a varity of environmental factors such as obstacle diversity.

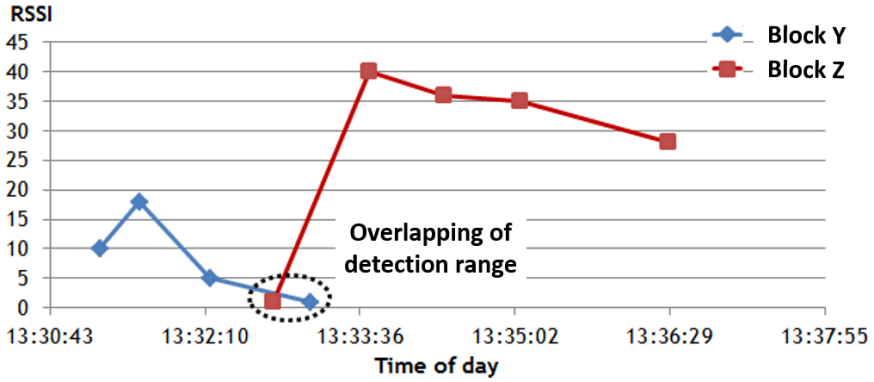


Figure 2 Detections of a Wi-Fi enabled device at the two locations

Finally, we extracted the Wi-Fi detection data from the two scanners in one week, from April 20, 2017 to April 26, 2017, for the analysis in the following sections. The raw data includes 765359 and 981225 detection records at Block Y and Z which belongs to 385464 MAC addrsses. The mean RSSI is 9.54 with the standard deviation of 10.01.

3. DATA PROCESSING RESULTS

A preliminary investigation shows that the raw Wi-Fi detection data include a lot of noisy detections which do not belong to portable electronic devices of pedestrians crossing the tunnel. Thus, a data filtering process is necessary to eliminate the noisy detections and keep those valid detections. In our study case, the pedestrian tunnel is not for a purpose of general passage. It is built to connect Block Z to the main campus. Thus, people who cross the tunnel should have some activities at Block Z. Thus, we first performed the data filtering based on the reasonableness of activities inferred from Wi-Fi data at Block Z. After that, another data filtering was performed to identify the records of tunnel crossing activities of pedestrians. Based on the processed Wi-Fi detection data, the pedestrian walking time between Block Y and Z was estimated. The overall data processing framework is shown in Figure 3.

3.1 Data Filtering Process

In Wi-Fi data filtering, two failing results could occur: 1) a valid detection which belongs to an electronic device of a pedestrian was filtered out; and 2) a noisy detection was remained. Note that once a valid detection was filtered out, there is no way to bring it back. As a consequence, in the design of each filtering step, the principle was to try the best to retain the valid detections even though

some noisy detections were not successfully excluded. In our study, multiple criteria were applied in the data filtering to distinguish reasonable and unreasonable activities of people at Block Z. The steps are introduced as follows.

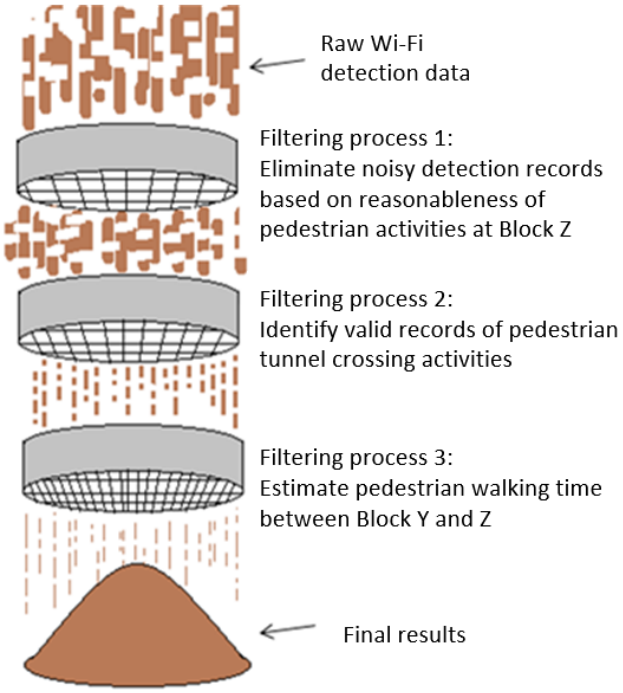


Figure 3. Data processing framework for Wi-Fi detections

Step 1: Delete data with empty MAC addresses. This step is very straightforward as invalid MAC address cannot be used to identify pedestrians.

Step 2: Delete MAC addresses with only one detection record. Those data belong to some devices that occasionally pass the Wi-Fi scanner. A pedestrian should have multiple detection records.

Step 3: Delete MAC vendors that obviously not belong to pedestrians' portable devices. We ranked the vender by detection frequency and manually identified the validity of top 20 vendors. The invalid venders include Wi-Fi bus, driving video recorder, printer, Octopus, etc., and were excluded from the dataset.

Step 4: Delete MAC addresses with only one record in its adjacent 48 hours (24hour before, 24 hour after). This is based on the assumption that if a pedestrian has an activity at the Block Z, it should have at least two detection records within 24 hours.

Step 5: Identify each walking in or walking out activity at Block Z. If the interval between two detections is longer than 600 s (10min), they were considered as belonging to two activities. Considering people may wait for elevator (maximum 3 min based on site observation) and detection range of a Wi-Fi scanner, 10 min is a reasonable threshold for not excluding valid detections.

Step 6: In each activity, delete MAC addresses with more than 8 records or duration is longer than 600 s. This is based on the assumption that if a MAC address is detected repeatedly many times in a long duration, it does not belong to a passing-by pedestrian.

Step 7: Delete MAC addresses with excessive detection records. Assuming that the reasonable number of coming in and out activities by a pedestrian at Block Z should be less than 30 per day with a detection record of 5 per activity, the maximum number was 7*30*5=1050.

Step 8: If the RSSI of all records of a MAC address is smaller than a threshold, delete the records of the MAC address. This is based on the assumption that if the MAC is valid, its records should contain data with strong signals. We used the mean value (RSSI=10) as the threshold.

The data filtering result was compared with the ground truth data. The number of pedestrians coming in and out of the BLOCK Z was manually identified from the video camera installed at the entrance. An example of comparison result between Wi-Fi device count and pedestrian count was shown in

Figure 4. The device count in raw data was also plotted. It is identified that there is a big difference between the device count in raw data and the pedestrian count. After data filtering, the device count curve has a similar variation trend as the pedestrian count curve. The percentage of device count divided by pedestrian count was 66.54%, suggesting that some pedestrian may not enable the Wi-Fi of portable devices in the detectable mode

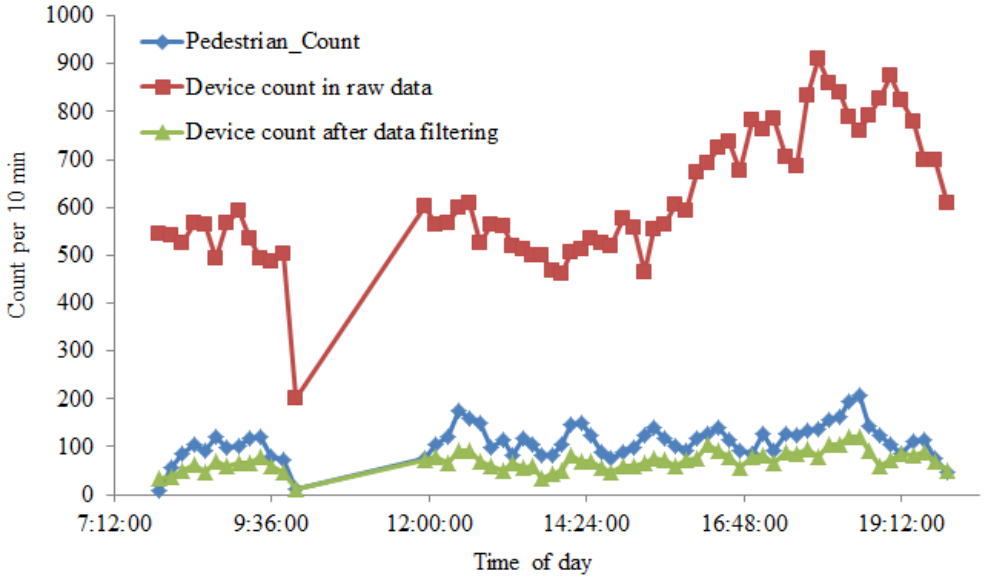


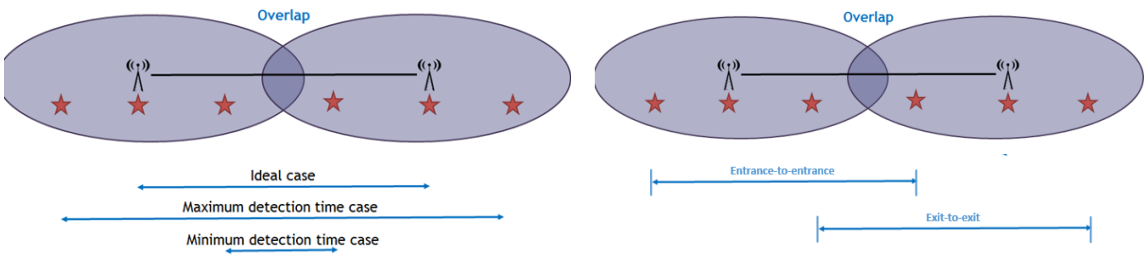
Figure 4. Comparing Wi-Fi detection data with pedestrian count (7:50 – 19:50, 21 August, 2017)

After the above data filtering process, we then paired the Wi-Fi detection data at Block Z with those at Block Y. These MAC addresses retaining at Block Z were excluded if they did not show up at Block Y. With the paired data at Block Y and Z, we then applied an algorithm to identify each tunnel crossing activity. Empirical investigation shows that the distance between two scanners is 150 m. Assuming a normal walking speed should be faster than 0.5 m/s, the longest walking time of crossing the tunnel was set to be 5 min. Thus, the MAC addresses at Block Y (or Z) that were not detected at Block Z (or Y) in its adjacent 10 min (5min before, 5min after) were discarded. Each tunnel crossing activity contains multiple Wi-Fi detection records at both sides within the 5 min period. Now we have a relatively clean dataset for pedestrian walking time analysis which was discussed in the following section.

3.2 Estimation of Pedestrian Walking Time

The illustration of Wi-Fi detections by the two scanners is shown in Figure 5 (a). The accurate walking time is the ideal case in Figure 5 (a), which is calculated as the difference of time when the pedestrian is at the two scanner locations. However, we actually don't know the location of the pedestrian from the Wi-Fi detection data. Thus, the accurate pedestrian walking time cannot be directly calculated. Methods were applied to estimate the walking time from the Wi-Fi detection. Note that as there is an overlap zone, the minimum detection time calculated from Wi-Fi detection in our case could be a negative value.

Abbott-Jard et al. (2013) have used Wi-Fi detection data to estimate vehicle travel time between two intersections. The entrance-to-entrance and exit-to-exit methods were applied, as shown in Figure 5 (b). The Entrance-to-Entrance method involves removing the reoccurring devices after the first detection. This method calculates pedestrian walking time as they enter the range of the two scanners. The exit-to-exit method involves removal of all but the last occurrence of the MAC address from the data. Note that the two methods only estimate the walking time between scanners due to the uncertainties in Wi-Fi detection.



(a) (b)

Figure 5 Walking time estimation based on Wi-Fi detection

Based on cleaned Wi-Fi data, the pedestrian walking time was calculated by the entrance-to-entrance and exit-to-exit methods for different times of day. The results are shown in Figure 6 and Table 2. The average walking time between the two Wi-Fi scanners is about 90.56 to 94.86 seconds, with a standard deviation of 40.08 to 40.79 seconds. The statistical information for each walking direction was also shown in Table 2. Figure 6 showed that the walking time in the early morning is shorter than in other time slots, probably because people tend to walk faster when there are less people in tunnel. Another reason is that due to the small sample size in the early morning, a noise that is not successfully filtered could bias the result a lot. The walking time during morning peak (8:30 -9:30) and noon time (13:30-14:30) is slightly longer than other time slots. This is probably because there are more pedestrians crossing the tunnel causing more interactions and disturbances among each other.

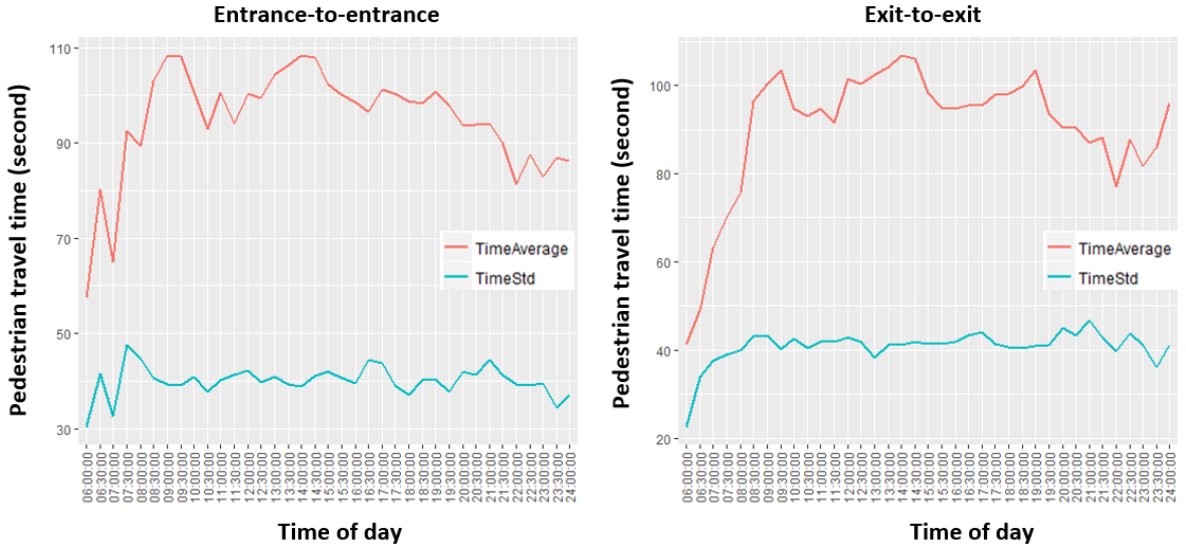


Figure 6. Estimated pedestrian walking time based on WI-Fi detection data

Table 2 Estimation of pedestrian walking time based on Wi-Fi detection

Pedestrian walking time	Entrance-to-entrance method	Exit-to-exit method
Two directions		
Average walking time (second)	99.59	96.33
Standard deviaiton of walking itme (second)	41.28	43.05
From Block Y to Z		
Average walking time (second)	106.82	102.04
Standard deviaiton of walking itme (second)	41.42	43.61
From Block Z to Y		
Average walking time (second)	92.35	90.62
Standard deviation of walking time (second)	41.15	42.49

We also calculated the flow patterns of pedestrians crossing the tunnel. The flow aggregated in every 30 min was shown in Figure 7. The results indicate that there are several peak periods with large

pedestrian flow in a day. They are the morning peak (8:30 – 10:00), the afternoon peak (17:30-19:00), and the noon peak (12:00-14:00). In the evening and early morning, pedestrian flow is very small. The results are consistent with intuition which shows the validity of the data filtering process applied in the study.

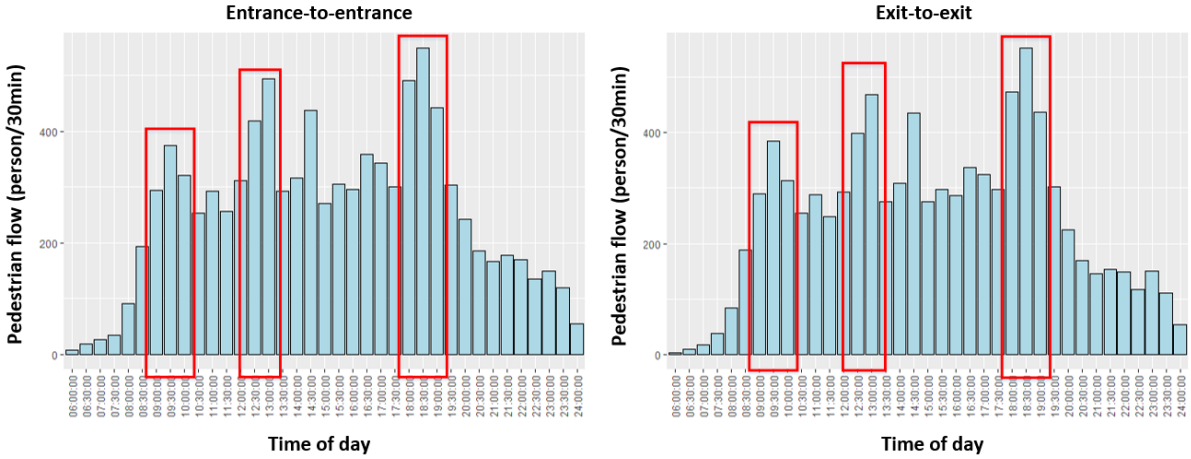


Figure 7. Pedestrian flow estimation based on Wi-Fi detection

4. CONCLUSIONS

This study proposed a framework to estimate pedestrian walking time on PolyU campus based on Wi-Fi detection data. The Wi-Fi scanner was installed on each side of a pedestrian tunnel. A data filtering procedure was followed to exclude noisy detections but keep valid records belonging to portable electronic devices of pedestrians. The data filtering result was then compared with the ground truth pedestrian count extracted from video tapes. Then the walking time was estimated based on the entrance-to-entrance and exit-to-exit methods. Pedestrian flow features were also estimated based on the cleaned data. The results showed that after the data fileting, the trend shown in the number of devices in Wi-Fi detection is consistent with pedestrian count. The average pedestrian walking time between the two Wi-Fi scanners was estimated to be 90.56 to 94.86 seconds, with a standard deviation of 40.08 to 40.79 seconds. The peak periods with large pedestrian flow were identified from the Wi-Fi detection. Findings of this study can provide useful information for understanding of the pedestrian activity patterns so as to improve the usage and design of the walking facilities on the PolyU campus.

5. ACKNOWLEDGEMENT

The work was jointly supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project Nos. PolyU 5243/13E & 152057/15E) and the Research Committee of The Hong Kong Polytechnic University (Project No. 1-ZVFJ).

6. REFERENCES

Abedi, N., Bhaskar, A., & Chung, E. (2013). Bluetooth and Wi-Fi MAC address based crowd data collection and monitoring: benefits, challenges and enhancement.

Al-Azzawi, M., & Raeside, R. (2007). Modeling pedestrian walking speeds on sidewalks. Journal of Urban Planning and Development, 133(3), 211-219.

Danalet, A., Farooq, B., & Bierlaire, M. (2014). A Bayesian approach to detect pedestrian destination-sequences from WiFi signatures. Transportation Research Part C: Emerging Technologies, 44, 146-170.

Dewan, M. A. A., Granger, E., Marcialis, G. L., Sabourin, R., & Roli, F. (2016). Adaptive appearance model tracking for still-to-video face recognition. Pattern Recognition, 49, 129-151.

Fitzpatrick, K., Brewer, M., & Turner, S. (2006). Another look at pedestrian walking speed. Transportation Research Record: Journal of the Transportation Research Board, (1982), 21-29.

Hoogendoorn, S. P., Daamen, W., & Bovy, P. H. (2003). Extracting microscopic pedestrian characteristics from video data. In Transportation Research Board Annual Meeting (pp. 1-15).

Teknomo, K., Takeyama, Y., & Inamura, H. (2016). Determination of pedestrian flow performance based on video tracking and microscopic simulations. arXiv preprint arXiv:1609.02243.

Kim, Y., Shin, H., Chon, Y., & Cha, H. (2013). Smartphone-based Wi-Fi tracking system exploiting the RSS peak to overcome the RSS variance problem. Pervasive and Mobile Computing, 9(3), 406-420.

Lam, W. H., & Cheung, C. Y. (2000). Pedestrian speed/flow relationships for walking facilities in Hong Kong. Journal of transportation engineering, 126(4), 343-349.

Liu, Q., Liang, P., Zhou, Y., & Lao, X. (2015, December). Pedestrian movement monitoring based on face recognition and RFID recognition. In Computer Science and Network Technology (ICCSNT), 2015 4th International Conference on (Vol. 1, pp. 673-676). IEEE.

Poucin, G., Farooq, B., & Patterson, Z. (2016). Pedestrian Activity Pattern Mining in WiFi-Network Connection Data. In Transportation Research Board 95th Annual Meeting (No. 16-5846).

Yi, S., Li, H., & Wang, X. (2015). Pedestrian travel time estimation in crowded scenes. In Proceedings of the IEEE International Conference on Computer Vision (pp. 3137-3145).

Wang, T., Gong, S., Zhu, X., & Wang, S. (2014, September). Person re-identification by video ranking. In European Conference on Computer Vision (pp. 688-703). Springer, Cham.

Wepulanon, P., Lam, W. H. K., & Sumalee, A. (2015). Using in-campus wi-fi data for analysis of student activity sequence.