



A METHODOLOGICAL FRAMEWORK FOR EXTRACTING PEDESTRIAN MOVEMENTS FROM WI-FI SENSING DATA IN A CONGESTED CENTRAL URBAN AREA

Takumi Komatsubara¹, Kazuyuki Takada¹, Yuma Morisaki², Makoto Fujii²

¹Tokyo Denki University, Japan

²Kanazawa University, Japan

Abstract

Effective design and management of pedestrian spaces require reliable measurement of pedestrian volume, dwell time, and movement routes. Video observation and questionnaire surveys have traditionally been used for this purpose; however, continuous implementation has become increasingly difficult because of labor requirements and concerns regarding privacy and personal data protection. Mobile phone population statistics derived from smartphones and other devices are also used for urban analysis, but their spatial resolution is often too coarse to capture pedestrian dwell time and circulation behavior within central urban areas. This study therefore develops a survey method using Wi-Fi packet sensing to estimate pedestrian dwell time and movement patterns in a central urban area. The study area is the historic tourist district of Kawagoe City, located in the Tokyo metropolitan suburbs. This district attracts many visitors, and pedestrians and vehicles are densely mixed on weekends, causing chronic congestion. Raw Wi-Fi packet sensing data therefore include not only devices carried by pedestrians but also devices carried by occupants of cars and buses. Without appropriate filtering, these records can bias estimates of pedestrian circulation behavior and dwell time. To address this issue, this study proposes a stepwise filtering procedure that combines three criteria: (1) sensor-location attributes, (2) travel speeds consistent with pedestrian movement, and (3) simultaneous co-movement of multiple devices along the same route during low-speed movement. The results demonstrate that pedestrian-derived records can be extracted even in pedestrian-vehicle mixed traffic environments. The analysis confirmed that interzonal movements tended to concentrate around major tourist attractions. In addition, dwell-time distributions revealed location-specific differences in dwell characteristics. These findings indicate that Wi-Fi packet sensing combined with appropriate filtering can provide effective data for analyzing pedestrian dwell time and movement patterns in tourist districts where pedestrian and vehicular traffic coexist.

Keywords: passive Wi-Fi sensing, Wi-Fi probe requests, pedestrian movement, dwell time, origin-destination

1 Introduction

Planning and managing safe and attractive pedestrian spaces require an understanding of pedestrian volume, dwell time, and route choice. In tourist districts in particular, these factors are closely related to urban design and visitor experience. It is therefore necessary to identify congestion levels and spatial concentrations of visitors in order to develop effective pedestrian circulation strategies. Pedestrian movement has traditionally been measured using video observation and questionnaire surveys.

These methods can provide detailed information on actual pedestrian behavior, but they require substantial labor and are difficult to implement continuously over long periods. Mobile phone population statistics and other smartphone-based aggregate data have also been applied to urban analysis. However, these data are often spatially aggregated, which limits their applicability for evaluating pedestrian behavior at the street, facility, or point scale. Against this background, Wi-Fi packet sensing has attracted attention as a practical method for observing pedestrian movement. However, many historic tourist districts in Japan contain narrow streets where pedestrians, private vehicles, and buses move in close proximity. In such pedestrian-vehicle mixed traffic environments, where pedestrians and vehicles are not physically separated, probe request records from in-vehicle devices are included in the observed data. If these records are not appropriately removed or suppressed, estimates of dwell time and movement patterns may be substantially biased.

This study therefore proposes a stepwise method for extracting pedestrian-derived trip chains from Wi-Fi packet sensing data observed in pedestrian-vehicle mixed traffic environments. The method uses sensor-location attributes, estimated inter-sensor travel speed, and simultaneous co-movement patterns of multiple devices as classification criteria. By reducing the inclusion of records from in-vehicle devices, the method aims to clarify pedestrian movement patterns and dwell characteristics in the historic tourist district of Kawagoe City.

2 Previous research

This section reviews previous studies on pedestrian behavior analysis using Wi-Fi packet sensing and clarifies the contribution of this study. Many studies have analyzed pedestrian behavior and tourist movement patterns using Wi-Fi packet sensors. Terabe et al. [1] reviewed behavioral surveys using Wi-Fi packet sensors and showed that such sensors have been used to estimate visitor volume, dwell time, and origin-destination (OD) patterns. Kimura et al. [2] analyzed pedestrian flows and dwell behavior in street spaces using Wi-Fi packet sensor data, reconstructed movement routes, and classified dwell-time distributions. Gao and Schmöcker [3] modeled route choice and dwell time in a tourist area, while Hou et al. [4] proposed a method using Wi-Fi probe data to distinguish walking for transport from walking accompanied by strolling or dwelling. These studies indicate that Wi-Fi data can be applied not only to OD estimation and dwell-time measurement but also to the analysis of behavioral characteristics such as route choice and walking purpose. The uncertainty of observed Wi-Fi data must also be considered. Musa and Eriksson [5] estimated travel paths using Wi-Fi monitor data and evaluated the accuracy and limitations of the method through comparison with GPS data. Chilipirea et al. [6] examined preprocessing methods for crowd analysis using Wi-Fi data and emphasized the need for data cleaning that accounts for false detections, missed detections, duplicate detections, and changes in device identifiers. Schauer et al. [7] showed that the use of RSSI and detection timestamps improves the accuracy of pedestrian-flow estimation based on Wi-Fi and Bluetooth data. These findings suggest that preprocessing and accuracy-improvement procedures must be designed according to the characteristics of the sensing data.

Previous studies have accumulated knowledge on the use of Wi-Fi packet sensing for pedestrian behavior analysis, route-choice estimation, dwell-time estimation, data screening, and accuracy improvement. However, a systematic method has not yet been sufficiently established for extracting pedestrian-derived movement records while suppressing records from in-vehicle devices in pedestrian-vehicle mixed traffic environments, where pedestrians and vehicles move in close proximity. This study addresses this gap by combining sensor-location attributes, estimated travel speed, and simultaneous co-movement patterns of multiple devices to extract pedestrian-derived records from Wi-Fi packet sensing data.

3 Study area and field survey

The study area is a historic tourist district in Kawagoe City, located in the Tokyo metropolitan suburbs. This survey focused on the Kurazukuri Zone (Old Storehouse Zone), including Ichibangai Street, Toki no Kane (Time Bell Tower), Kashiya Yokocho (Penny Candy Lane), the approach to Kawagoe Hikawa Shrine, and the surrounding streets. These tourist attractions and adjacent streets attract many visitors and form the main pedestrian circulation area of the district. At the same time, pedestrian and vehicular traffic tend to concentrate on narrow streets around major tourist attractions and commercial facilities, producing a pedestrian-vehicle mixed traffic environment. The field survey was conducted on Saturday, November 1, 2025, from 10:30 a.m. to 4:30 p.m. (JST). During the survey period, temporary vehicle traffic restrictions were implemented in part of the Ichibangai area. To capture diverse route choices and continuous movement routes of visitors, 34 Wi-Fi packet sensors were installed throughout the study area. Each sensor received probe request frames transmitted by Wi-Fi-enabled devices. The recorded variables were detection time, received signal strength indicator (RSSI), and MAC address used as a device identifier. For privacy protection, the content of communications was not accessed.

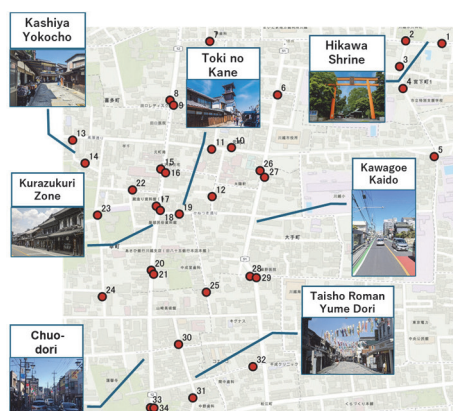


Figure 1 Installation locations of Wi-Fi packet sensors

4 Data screening

Outdoor Wi-Fi packet sensing data are affected by reception conditions and device specifications. In particular, RSSI can fluctuate over short time intervals because of sensor installation conditions, surrounding obstacles, building reflections, and human-body shielding. In addition, probe request transmission intervals vary depending on the device operating system and sleep state; consequently, a device may not be detected continuously even when it remains near the same location. Recent smartphones also randomize MAC addresses, which makes long-term tracking of the same individual difficult. Given these characteristics, a screening procedure was applied to organize the observations into a dataset suitable for analysis. First, records with RSSI values between -98 dBm and -30 dBm were retained, and extremely weak signals and outliers were excluded. In addition, each sensor was assigned to either a pedestrian space or a pedestrian-vehicle mixed space according to the characteristics of the observation location. Next, consecutive detections of the same MAC address by the same sensor were aggregated into a single observation episode. If the interval between detections exceeded 10 min, the detections were treated as separate episodes rather than as a continuous stay at the same location.

This rule prevents a device that left a location and later revisited it from being counted as a single continuous stay. For each observation episode, duration, packet count, and maximum RSSI were calculated. Episodes lasting 5 min or longer were classified as dwell episodes, whereas episodes shorter than 5 min were classified as pass-through episodes. After single-sensor processing, consecutive observations of the same MAC address by different sensors were connected to reconstruct inter-sensor movement records. For each movement record, travel speed was estimated by dividing the shortest-path distance between the corresponding sensors by travel time. Records with reversed time order or unrealistic time differences were excluded from subsequent analysis.

5 Pedestrian data extraction by filtering

This section describes the filtering procedure used to extract pedestrian-derived records from the screened data. The classification is based on three criteria: travel speed, sensor-location attributes, and simultaneous movement of multiple devices.

5.1 Determination based on travel speed

Non-pedestrian movement was identified using the inter-sensor travel speed estimated in the preceding section. In this study, movements with travel speeds exceeding 2.2 m/s (7.92 km/h) were excluded as likely non-pedestrian movements. Through this criterion, 20,396 of the 28,965 observed records were extracted as pedestrian-derived candidate records.

5.2 Classification based on sensor-location attributes

For the candidate records extracted in section 5.1, sensor-location attributes were used to determine whether the observed space was a pedestrian space or a pedestrian-vehicle mixed space. Records observed in pedestrian spaces were treated as pedestrian-derived records, whereas records observed in pedestrian-vehicle mixed spaces were subjected to the additional classification described in section 5.3. As a result, 12,040 of the 20,396 candidate records were classified as observations in pedestrian spaces, and 8,356 were classified as observations in pedestrian-vehicle mixed spaces.

5.3 Classification based on simultaneous movement of multiple devices

In pedestrian-vehicle mixed traffic environments, vehicle speeds can decrease during congestion. Under such conditions, distinguishing private vehicles, buses, and pedestrians by travel speed alone is difficult. The analysis therefore focused on the characteristic that multiple devices may exist inside the same vehicle. For the 8,356 records classified as observations in pedestrian-vehicle mixed spaces, a co-movement filter was applied to identify simultaneous movement of multiple devices. Specifically, records in which multiple devices moved along the same route within a short time interval were detected and excluded as records likely to originate from the same vehicle. After this procedure, 8,309 records were retained as pedestrian-derived records.

5.4 Filtering results

Through the three-stage filtering procedure, 20,349 observations were extracted as pedestrian-derived records out of 28,965 observed records. These comprised 12,040 records observed in pedestrian spaces and 8,309 records retained after applying the co-movement filter in pedestrian-vehicle mixed spaces. At the device-address level, 1,344 of the 1,392 observed addresses were retained as pedestrian-derived addresses.

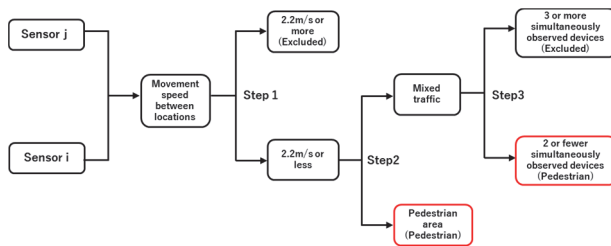


Figure 2 Data Filtering process for multiple sensors

6 Analysis of movement patterns and dwell times

6.1 Analysis of movement patterns

Using the pedestrian-derived dataset extracted in section 5, pedestrian circulation patterns within the study area were analyzed. The study area was divided into 16 zones based on the distribution of tourist facilities and the street structure, and sensor observations were aggregated at the zone level. Figure 3 shows the interzonal transition probabilities. Blue circles denote dummy nodes representing inflow to and outflow from the study area. Line thickness indicates movement volume based on the number of MAC addresses. Only transitions with transition probabilities of 5% or higher are displayed. As shown in the figure, interzonal movements tended to concentrate in specific zones. In particular, many movements were directed toward major tourist attractions such as the Kurazukuri Zone, Kashiya Yokochi, and Kawagoe Hikawa Shrine. In contrast, movements in multiple directions were observed around Honmachi, suggesting that this area may function as a connection point within the circulation network of the study area.

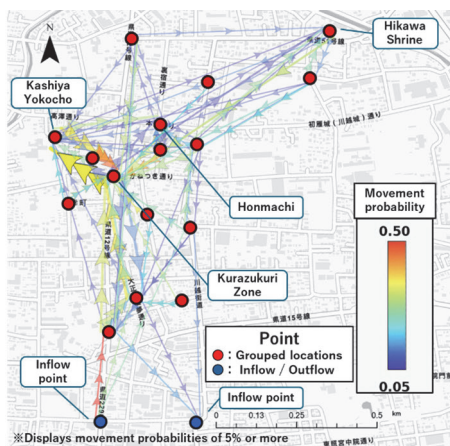


Figure 3 Transition probability

6.2 Analysis of dwell time

Dwelling conditions by visitor location were analyzed using the dwell times observed at each location. In this study, estimated dwell times at each location were aggregated into 5-min intervals, and cumulative percentages were calculated to characterize dwell-time distributions by location. This made it possible to compare whether each location tends to be passed

through in a short time or tends to be used for stays of a certain duration or longer. Comparison of the distribution shapes in figure 4 revealed differences in dwell characteristics among locations. At Kashiya Yokocho, the Kurazukuri Zone, and Kawagoe Hikawa Shrine, relatively long dwell times were observed for a consistent share of records, suggesting that these locations have a strong destination-type character. In contrast, the alley west of the Kurazukuri Zone and Kawagoe Kaido showed many short-duration observations, indicating a passing-type character. In addition, Chuo-dori Street and Taisho Roman Yume-dori Street showed both short-duration passing and stays of a certain duration, suggesting that these locations combine passing and dwelling functions.

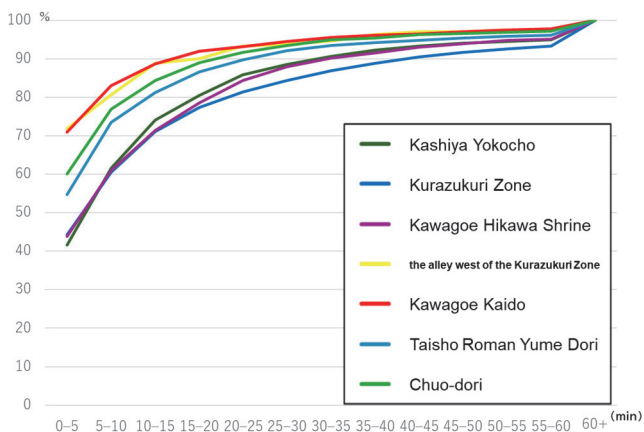


Figure 4 Cumulative distribution of dwell time

7 Conclusion

This study proposed a method for extracting pedestrian-derived observation records from Wi-Fi packet sensing data in the historic tourist district of Kawagoe City. To address the challenges of pedestrian-vehicle mixed traffic environments, three classification criteria were applied: sensor-location attributes, estimated travel speed, and a co-movement filter. The proposed procedure reduced the inclusion of records derived from vehicles and other non-pedestrian sources and produced a dataset suitable for analyzing pedestrian circulation and dwell behavior. The analysis using the constructed dataset confirmed that interzonal movements tended to concentrate toward major tourist attractions. In addition, comparison of cumulative dwell-time distributions showed that visited locations can be characterized as destination-type locations, pass-through spaces, or mixed-function locations.

This study has several limitations that should be addressed in future work. First, the analysis is based on observations from only one day in one study area. Therefore, observations on different days and in other areas are needed to verify the applicability of the proposed method. Second, Wi-Fi packet sensing data are affected by device communication specifications and sensor installation environments, which may influence the estimation of dwell time and movement speed. Future work should compare the results with external reference data, such as video observation, to evaluate estimation accuracy. Third, because MAC address randomization makes long-term tracking of individuals difficult, the proposed method is suitable for capturing short-term consecutive movements and stays. Future work should examine how changes in MAC addresses affect the continuity of reconstructed trip chains and improve the accuracy of short-term movement analysis.

References

- [1] Terabe, S., Ichii, K., Yaginuma, H., Ono, M., Tanaka, K., Kang, N.: Tourists' behavioral survey using Wi-Fi packet sensors: comprehensive review of existing research and comparison of origin-destination matrices, *Journal of the Japan Society of Civil Engineers, Ser. D3 (Infrastructure Planning and Management)*, 75 (2019) 5, pp. I_669-I_679, DOI: 10.2208/jscejipm.75.I_669
- [2] Kimura, Y., Higashikawa, A., Susaki, J.: Analysis of pedestrian flows in street spaces using Wi-Fi probe data: with a focus on the Higashiyama area, Kyoto City, *Japanese Journal of JSCE*, 79 (2023) 22, Article ID: 22-22008, DOI: 10.2208/jscej.22-22008
- [3] Gao, Y., Schmöcker, J.D.: Estimation of walking patterns in a touristic area with Wi-Fi packet sensors, *Transportation Research Part C: Emerging Technologies*, 128 (2021), 103219, DOI: 10.1016/j.trc.2021.103219
- [4] Hou, J., Zhang, E., Long, Y.: Measuring pedestrian flows in public spaces: Inferring walking for transport and recreation using Wi-Fi probes, *Building and Environment*, 230 (2023), 109999, DOI: 10.1016/j.buildenv.2023.109999
- [5] Musa, A.B.M., Eriksson, J.: Tracking unmodified smartphones using Wi-Fi monitors, 10th ACM Conference on Embedded Network Sensor Systems (SenSys '12), pp. 281-294, Toronto, Canada, 6-9 November 2012., DOI: 10.1145/2426656.2426685
- [6] Chilipirea, C., Petre, A.C., Dobre, C., van Steen, M.: Presumably simple: monitoring crowds using Wi-Fi, 2016 17th IEEE International Conference on Mobile Data Management (MDM), pp. 220-225, Porto, Portugal, 13-16 June 2016., DOI: 10.1109/MDM.2016.42
- [7] Schauer, L., Werner, M., Marcus, P.: Estimating crowd densities and pedestrian flows using Wi-Fi and Bluetooth, 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MobiQuitous 2014), pp. 171-177, London, United Kingdom, 2-5 December 2014., DOI: 10.4108/icst.mobiquitous.2014.257870

