Analysis of Campus Crowd Behavior Based on Location Data and Physical Environment Data: a Case Study of Southeast University Wuxi Campus

Ye Tang, Junqiang Sun, Guangjin Wang, Wenjin Hong and Li Li

March 17, 2025

Analysis of Campus Crowd Behavior Based on Location Data and Physical Environment Data: A Case Study of Southeast University Wuxi Campus

Ye Tang[1], Junqiang Sun, Guangjin Wang, Wenjin Hong, Li Li[2]

Abstract. The study on the behavior of on-campus individuals provides valuable insights for campus management, resource allocation, and planning layout. The application of multi-source data offers more objective and in-depth opportunities for exploring behavioral phenomena. Focusing on the Wuxi campus of Southeast University, this research utilized Wi-Fi probe positioning technology combined with a physical environment sensor system to comprehensively collect 28.87 million positioning data points and 340,000 environmental data points over a period of 14 days. After cleaning redundant, missing, abnormal, drifting, and ping-pong data, both types of data underwent visual analysis, and their correlations were studied. Additionally, trajectory feature extraction was conducted using a convolutional autoencoder neural network. The study revealed the temporal distribution of pedestrian flow, the spatial distribution of stopover behavior, and the spatiotemporal characteristics of pedestrian trajectories. This provides a reliable basis for guiding crowd behavior by improving specific campus areas and the physical environment.

Keywords: Wi-Fi probes, Physical environment sensors, Data visualization, Characterization, Data correlation

1 Introduction

High school campus spaces belong to the observational scale in urban spaces and feature comprehensive and concentrated urban facilities. Faced with the challenge of limited resources in urban spaces, various universities in China opt to expand their scale through new campus development, imposing higher demands on management, resource allocation, and planning. The characteristics of crowd behavior have become

[1] Ye Tang (🖂)
Southeast University, Si-Pai-Lou 2#, Nanjing, Jiangsu Province, China
e-mail: 220230214@seu.edu.cn

[2] Li Li (🖂)
Southeast University, Si-Pai-Lou 2#, Nanjing, Jiangsu Province, China
e-mail: 101012053@seu.edu.cn

a focal point in urban design, reflecting both the spatiotemporal features of urban areas and the behavioral needs of users. Analysing crowd behavior on university campuses helps obtain concise and diverse research samples, assisting designers and administrators in understanding the environment and providing site-specific design guidance. Currently, research on crowd behaviour characteristics is mature, with diverse data collection methods such as Ultra-Wideband (UWB), base station positioning, Wi-Fi positioning, and GPS positioning. However, due to the concentrated nature and moderate scale of campuses, a single data source may not comprehensively capture crowd behavior.

Taking Southeast University's Wuxi campus as an example, this paper combines the visualization and feature analysis of on-campus crowd behavior using positioning and physical environment data. The research provides essential references for campus management, resource allocation, and planning, offering valuable insights for the preliminary research of expansion plans for other universities.

2 Data Collection Tools and Methods

The study covers the detailed collection, cleaning and visualization of pedestrian positioning data and environment perception data (**Fig. 1**). Obtaining sufficient and valid data is a critical step in this study before proceeding with data processing. This chapter aims to provide a comprehensive overview of the methods used to acquire positioning data and physical environment data in the related study, with an in-depth discussion of the instrumentation employed, the data collection procedures, and the preprocessing steps used to obtain the dataset, with the aim of establishing a solid foundation for subsequent data analysis.

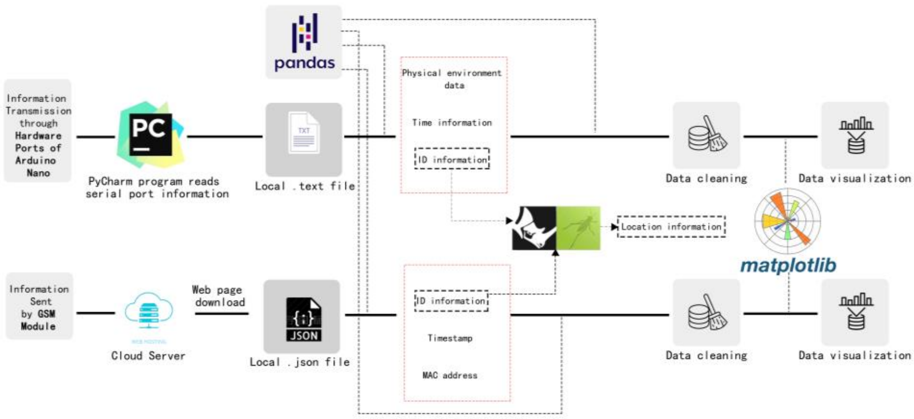

Fig. 1 Flowchart for data collection, cleansing, and visualization

2.1 Existing Equipment and Sensor Construction

Traditional crowd tracking methods suffer from significant manpower and resource input, as well as subjective influences. In 2011, Cyriac et al. introduced Wi-Fi probe technology, utilizing MAC addresses for anonymous tracking of Wi-Fi devices. Scholars like Hongyu Wan [1] extended the analysis of positioning data to architectural design using low-precision Wi-Fi probe devices. Yizhou Wu et al. [2] created layered user mobility models with Wi-Fi access point data. In this study, a self-developed Wi-Fi probe device integrates GSM, Wi-Fi, and power modules on a circuit board.

Traditional methods for obtaining building environmental data include satellite remote sensing, building information model simulations, and GIS for regional meteorological information. Yizhou Wu et al. [3] analyzed the impact of the physical environment on human behavior by simulating sound, light, heat, and landscape. Wang W et al. [4] compared datasets from EPW data, urban climate stations, and local microclimate stations. Purchasing meteorological station products has been a common method in past research, but these products are expensive, especially when monitoring multiple points.

For studies of this nature, a more suitable setup for physical environment sensors was implemented (**Fig. 2**): cost-effective and compact DHT11 temperature sensors, light-sensitive sensors, and three-cup ABS material anemometers were selected. A local area network communication system was constructed using LoRa modules to cover the entire campus, avoiding the use of expensive 4G modules. Sensor integration involved connecting the main control board Arduino Nano, LoRa module, and various sensors using a breadboard and dedicated jumper wires to enhance LoRa signal stability. Waterproofing measures were applied to the encapsulated sensors, ensuring normal operation under various weather conditions.

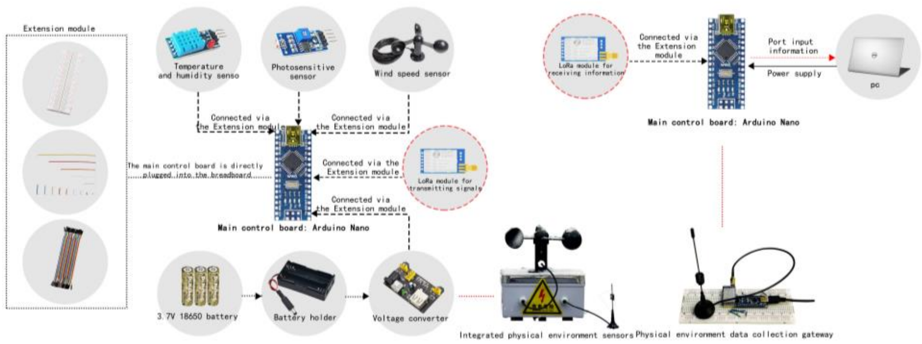

Fig. 2 Physical data sensors construction using embedded system technology

2.2 Data Collection

This study focuses on the Southeast University Wuxi Campus, which is organized in a circular layout with various functional zones surrounding the central landscape area. To comprehensively capture representative trajectories of human behaviour, the campus is divided into eight regions: Student Dormitory Area, Academic Experiment Zone, Professional Teaching Zone, Life Service Area, Library, Central Landscape Area, Public Teaching Zone, Sports and Fitness Zone, and Faculty Dormitory Area. Wi-Fi probe monitoring points are strategically placed at key intersections and public spaces within each region, following principles such as positioning them at high-frequency usage locations, ensuring unique paths between monitoring points, keeping them away from buildings to minimize indoor device impact, maintaining distance between monitoring points to reduce ping-pong data, and establishing points at campus entrances and exits.

After a week of preliminary data collection, 19 monitoring points were finalized, including critical traffic nodes and public space nodes like 202, 204, 205, 208, 213, 210, 215, 214, 211, 299 (**Fig. 3**). The information from these 20 points is directly uploaded to a cloud server via a 4G module. Over the course of a 14-day winter data collection period, a total of 28.87 million positioning data points were collected.

Physical environment monitoring points are placed in the main public spaces within the Wi-Fi probe coverage area. Sensors are positioned at heights corresponding to human standing behavior to ensure that monitoring accurately reflects people's perceptual conditions. Information from the 10 monitoring points is transmitted via LoRa modules to two gateways located on the 5th floor of the campus library and the 1st floor of the Two Rivers Building. Over the 14-day winter data collection period, a total of 3.4 million pieces of physical environment information were collected.

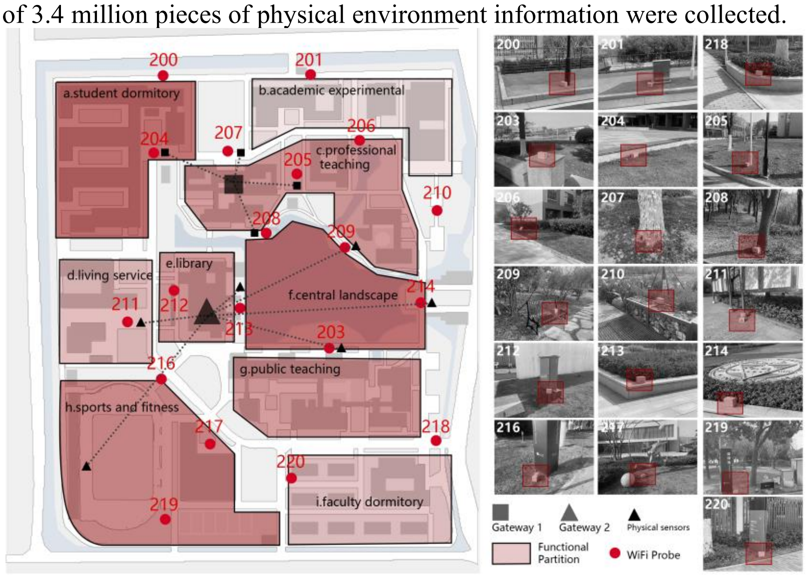

Fig. 3 Layout of monitoring points for collection devices

2.3 Data Cleansing

Data cleansing aims to detect and clean the dirty data and noise in the original data to ensure that the cleaned dataset reflects the nature of the data more realistically and improves the data visualization.

The data pre-processing stage requires judgment based on the specific project context. After analysis, five types of data requiring pre-processing were identified: ① Anomalous data: data lacking certain critical information; ② Time restrictions: removing data between 23:00 at night and 6:00 in the morning to reduce computation; ③ Drift data: data representing positions outside the possible reach within a certain time due to smart device or equipment malfunctions; ④ Redundant data: MAC addresses appearing continuously or intermittently within a monitoring range over a period; ⑤ Ping-pong data: devices repeatedly detected between two probe monitoring points. Simultaneously, a cleaning framework was established to sequentially clean these five types of data, improving cleaning efficiency. Ultimately, 400,000 valid data points were obtained, achieving a cleaning rate of 96%. This provides a high-quality data foundation for subsequent analysis.

The characteristics of physical environment data are simpler, and its data cleaning framework mainly includes two parts: data import and data preprocessing. The information collected by the gateway is written into a local .TXT file through the serial port of the computer, which is converted into a .CSV file or a .JSON file for subsequent cleaning. Information transmission through the Lora module will lead to a part of the missing information, the two parts of the extracted keywords need to be merged into a complete information. Some outliers need to be filtered out by setting thresholds for special values.

3 Spatial Distribution Based on Stopover Behavior and Physical Environment

Physical environmental data exhibit variations in microenvironments within the same region, primarily influenced by campus layout. Therefore, conducting a correlation analysis between the dwell rate in campus public spaces and the physical environment is a crucial aspect of an in-depth exploration of campus layout. By determining the minimum dwell time of samples on the same device within a specified period in positioning data, it can be assessed whether a sample has lingered at a device point. In this study, the dwell rate (J) is calculated as the ratio of the number of unique MAC addresses (W) to the count of recorded MAC addresses (S), indicating preferences in people's dwellings at different monitoring points in space (**Fig. 4**).

In addition, visualizing preferences in individuals' walking paths is achieved by restoring Wi-Fi probe data. However, due to the limited number of Wi-Fi probes, initial data can only be transformed into "point-to-point" line graphs, making it challenging to accurately represent actual trajectory situations. To optimize trajectory maps, the

Dijkstra algorithm is employed to find the shortest paths between different monitoring points (**Fig. 5**). The trajectory of a specific MAC throughout the day is presented on a two-dimensional plane, where point size indicates the duration of stay, and line thickness represents the frequency of walking (**Fig. 6**).

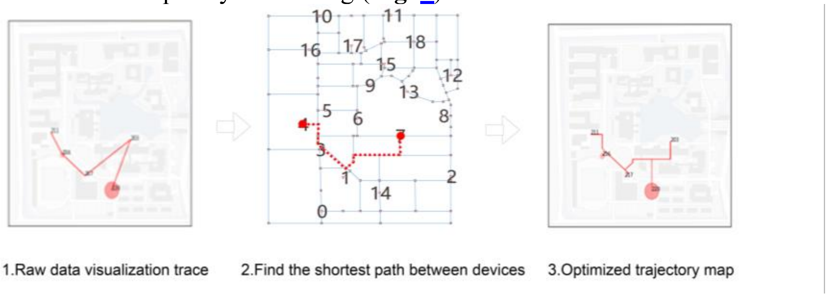

Fig. 4 Schematic diagram of optimized trajectory using Dijkstra algorithm

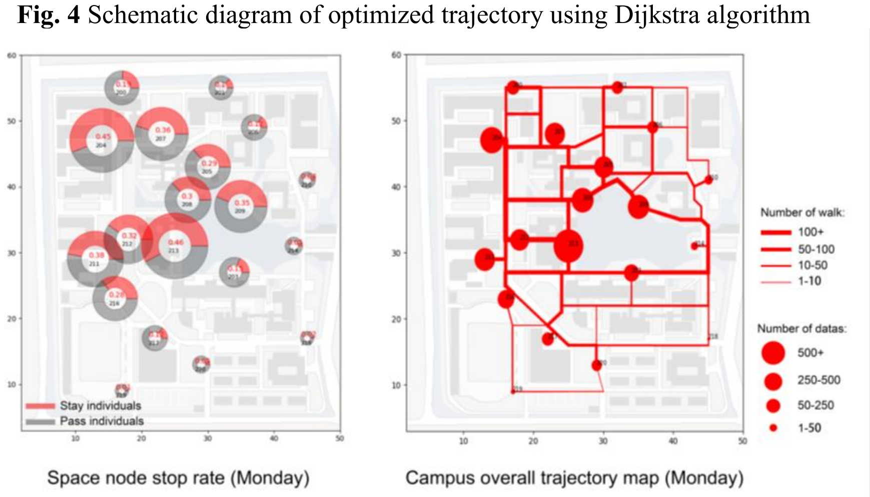

Fig. 5 On-campus Personnel Stopover Rate Plan
Fig. 6 Layout of monitoring points for collection devices

3.1 Spatial Visualization of Physical Data

By distributing the physical environment data spatially in the form of heat maps, it is possible to visualize the differences in the physical environment of different public spaces on campus (**Fig. 7**). The following characteristics were observed when observing the thermograms: ①The monitoring points around the centre lake, such as 213, 214, and 215, exhibited higher wind speeds. It is worth noting that point 204 has an unusually high wind speed despite being in an area surrounded by buildings. This area is a windy plaza that is usually underutilized despite the presence of public facilities. ②215, 205, 202, and 204 have relatively high temperatures. Among them, 202, 204, and 205 are in the dormitory area and workstation complex, which have higher temperatures relative to other public spaces. ③209, 214, 215 and 213 at the lake shore have significantly higher humidity than the other monitoring points, while point 299 located in the playground has the lowest humidity and is the driest public space. ④ 213 on the east side of the lake and 215 on the south side of the lake have the highest illuminance. They share the common characteristics of being open environments, away from buildings, and shaded by fewer trees.

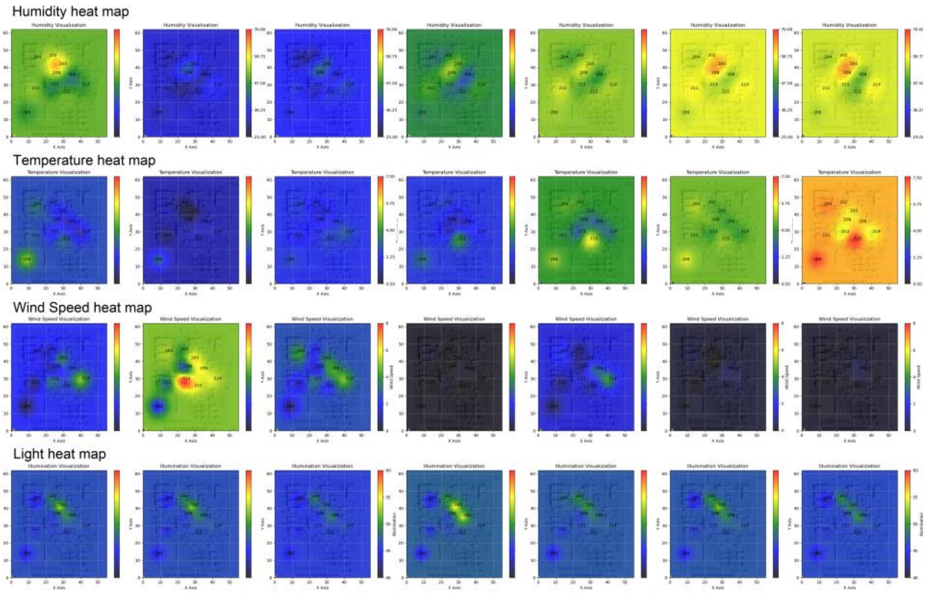

Fig. 7 Four visual views of the physical environment for each day of the week

3.2 Fitting of Stopover Rates to Environment Data Indicators

Upon visualizing the environmental data, particular attention was directed towards the dwell rates at ten key public space nodes. Calculations were performed at three time points during the day—morning, noon, and afternoon. The results from both aspects were integrated and simulated, yielding line graphs illustrating the data fluctuations (**Fig. 8**).

From the indicator line chart analysis, it is evident that the stay rates at all monitoring points are negatively correlated with wind speed and unrelated to illumination. The relationship between stay rates and temperature-humidity varies across different monitoring points, for examples, the stay rates at 208, 214, 213, and 215 are more closely related to humidity, while those at 209 and 205 are more closely related to temperature. In conclusion: ① In the campus under study, people's choice to stay in a particular public space is most affected by the wind environment. ②Due to fewer obstructions and wider roads on campus, the wind environment can easily cause discomfort to individuals during winter. ③Temperature and humidity are secondary factors influencing people's choices for staying activities, with individuals more likely to linger in leisure-oriented public spaces during the winter when humidity and temperature are higher. ④The stay rates at some public spaces serving as transportation hubs show no apparent correlation with temperature and humidity.

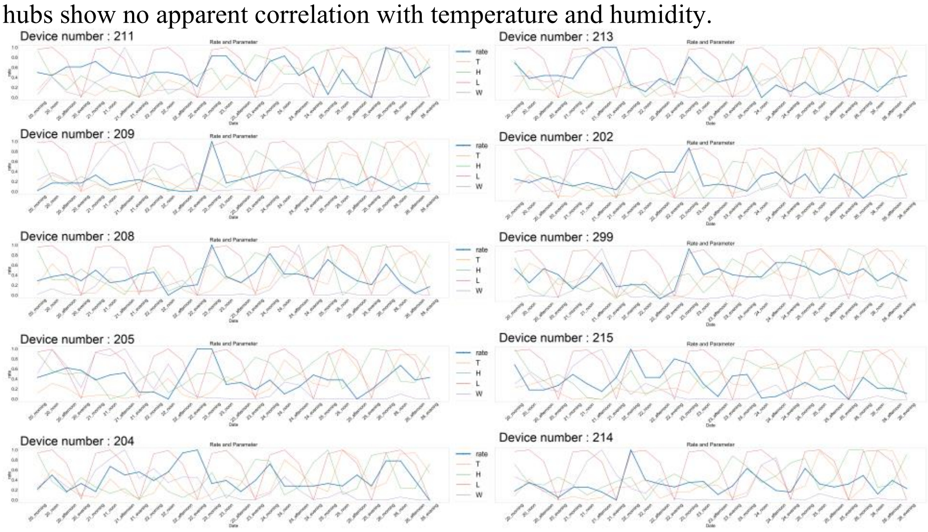

Fig. 8 Fitting line chart of stop rate and physical data indicators

3.3 Behavioural Trajectory Visualization and Clustering

Analyzing the overall trajectories across the campus for seven days reveals crucial public space nodes, such as the library, central landscape area, and dormitory areas. However, as two-dimensional trajectories cannot capture time information, this representation introduces time as the third dimension. If a MAC experiences a stationary behavior, the trajectory grows along the z-axis. This approach allows the observation of the trajectory's starting and ending points, facilitating the inference and categorization of activities.

Based on the cleaned data, this study obtained over 9,000 trajectory images within a week (**Fig.** 9). The diverse coordinate patterns of each trajectory make it challenging to intuitively extract patterns from numerous images. To address this issue, a Convolutional Auto-Encoder (CAE) was employed for trajectory clustering. In this study, a set of RGB images with pixel dimensions of 88×88 extracted over the week was used as the training base. During the clustering analysis, feature extraction was performed using the pre-trained ResNet-50 model. After a series of convolutions and pooling, the feature dimensions were compressed, and the K-means++ clustering algorithm was applied. The clustering results were then compressed into three-dimensional visualizations using t-SNE. Both two-dimensional and three-dimensional trajectory clustering yielded four distinct results (**Fig.** **10**).

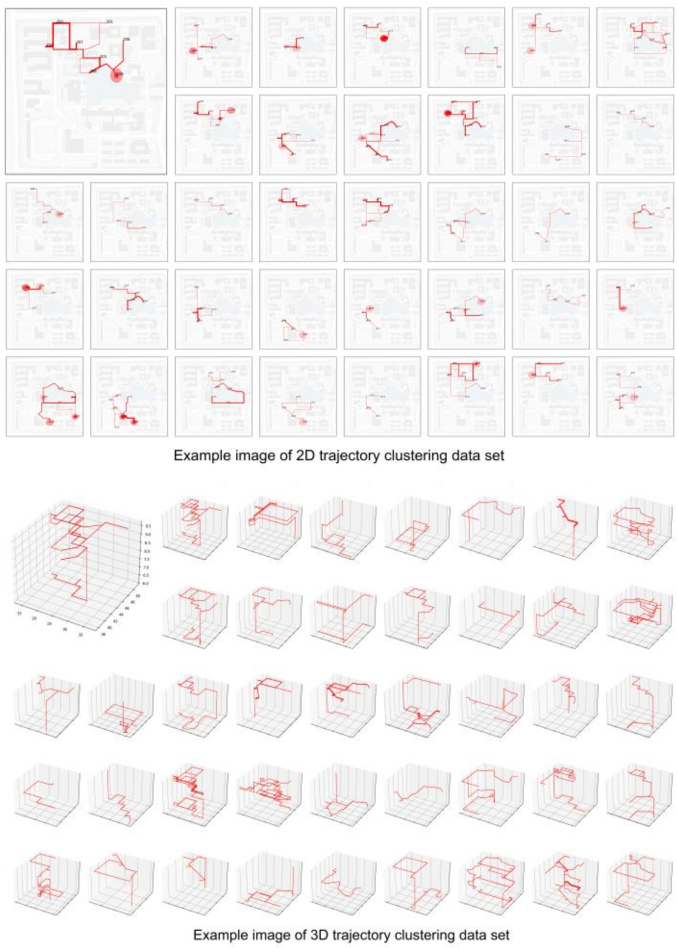

Fig. 9 Parts of 2D-3D trajectory map used as a training set for clustering

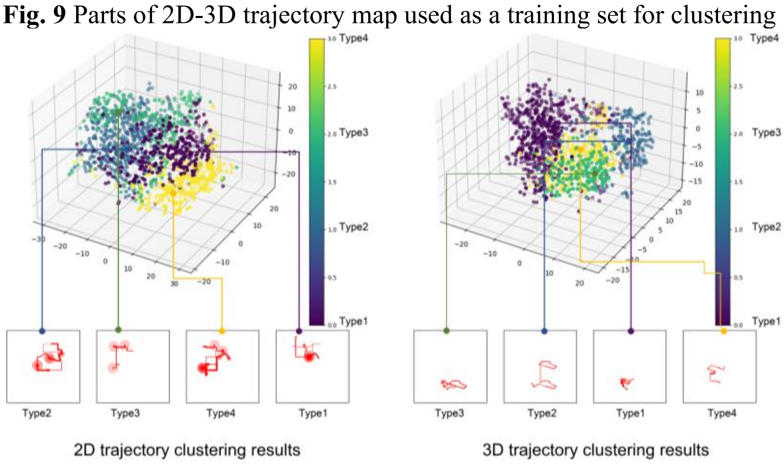

Fig. 10 Clustering results for 2D-3D trajectories

From the two-dimensional trajectory clustering results, campus trajectories can be categorized into four types: ① stationary activities around the central landscape area, ② activity trajectories centered around dormitories, cafeterias, and workstations with weak mobility, ③ more active trajectories centered around dormitories, cafeterias, and the library, ④ activity trajectories centered around professional classrooms. In the three-trajectory clustering results, people's trajectory ranges were generally singular and limited, mostly consisting of closed-loop trajectories around specific points, with a few non-closed trajectories differing in their starting points.

6 Summary and discussion

This study combines Wi-Fi probe technology and physical environmental sensors to comprehensively analyze pedestrian flow and environmental conditions at the Southeast University Wuxi campus. The selection of effective data collection techniques forms the foundation of the research, offering insights for studying urban spaces at various scales. From preliminary investigations to the analysis of current campus conditions, this data-driven research process holds practical significance for assessing and enhancing urban environments and validating design proposals. Despite achieving initial outcomes, the study faces limitations. Challenges include the insufficient sample size affecting the comprehensive representation of crowd behavior in trajectory clustering. The sensors based on LoRa modules are influenced by buildings and trees, resulting in signal losses. The data collection period spans only two weeks in winter, reflecting the impact of seasonal variations on human activity. Future research endeavors could optimize data cleaning processes, increase sample sizes, enhance sensor deployment strategies, and extend data collection durations. This would contribute to a deeper understanding of dynamic urban spatial features, ultimately improving the practicality and applicability of the research.

References

1. Wan, H., Wang, S., Du, J., Li, L., & Zhang, J. (2023). Trace analysis using Wi-Fi probe positioning and virtual reality for commercial building complex design. Automation in Construction
2. Cohen, A., & Cohen, B. (2023). Integrating Large-Scale Additive Manufacturing and Bioplastic Compounds for Architectural Acoustic Performance. In Proceedings of the 28th CAADRIA Conference
3. Wu, Y. Z., Wu, S. Q., Qiu, X. L., Wang, S., Yao, S. Y., Li, W. T., You, H. Y., Zhang, J. R., Xia, S. Y., & Guo, Y. F. (2022). Integrated Evaluation Method of the Health-Related Physical Environment in Urbanizing Areas: A Case Study from a University Campus in China. Frontiers in Public Health, 10, Article 801023.
4. Wang, W., Li, S. G., Guo, S. Y., Ma, M., Feng, S. H., & Bao, L. (2021). Benchmarking urban local weather with long-term monitoring compared with weather datasets from climate station and Energy Plus weather (EPW) data. Energy Reports, 7, 6501-6514.