**Dataset Description**

The complete dataset for Microsoft Indoor Location Competition 2.0 competition contains various dense indoor signatures of WiFi, geomagnetic field, iBeacons and ground truth which are collected from several hundreds of builds in cities of China [1]. For our project, the trace dataset is sample dataset including the data of two selected buildings from full set, known as site1 and site2 respectively and those data are collected by site-surveyor attached an Android smartphone to his/her body as while as walking from each point to the other along those buildings. Through each indoor traces between two positions and the recording app of sensor data are running on the Android smart phone device to collect accelerometer, gyroscope, magnetometer as well as WiFi and Bluetooth iBeacon scanning data. Those data are present as \*.txt format. In addition, floor plan metadata are also provided for each floor. An example the format of the file in Table 1.

Table 1:

|  |  |  |
| --- | --- | --- |
| **Time** | **Data Type** | **Values** |
| 1574572522291 | TYPE\_WAYPOINT | Px , Py |
| 1574572522414 | TYPE\_ACCELEROMETER | X, Y, Z, accuracy |
| 1574572522414 | TYPE\_MAGNETIC\_FIELD | X, Y, Z, accuracy |
| 1574572522414 | TYPE\_GYROSCOPE | X, Y, Z, accuracy |
| 1574572522414 | TYPE\_ROTATION\_VECTOR | X, Y, Z, accuracy |
| 1574572522414 | TYPE\_MAGNETIC\_FIELD\_UNCALIBRATED | Xb , Yb , Zb , Xa , Ya , Za accuracy |
| 1574572522414 | TYPE\_GYROSCOPE\_UNCALIBRATED | Xb , Yb , Zb , Xa , Ya , Za accuracy |
| 1574572522414 | TYPE\_ACCELEROMETER\_UNCALIBRATED | Xb , Yb , Zb , Xa , Ya , Za accuracy |
| 1574572524224 | TYPE WIFI | ssid, bbsid, RSSI, frequency,  lastseen |
| 1574572525875 | TYPE BEACON | UUID, MajorID, MinorID, Tx,  Power, RSSI, Distance, MAC,  Unix Time |

For Table 1:

* The column one is Unix Time when each data instance was recorded.
* The column two is sensor data types containing 10 types of data in total.
* The third column is data values. The values of each data vary depending on the type of data.

**Dataset Summary**

Table 2: Summarization

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Site** | **Floor** | **Height** | **Width** | **Waypoints** | … | … |
| 1 | B1 | 232 | 320 | 1034 |  |  |
| 1 | F1 | 176 | 240 | 975 |  |  |
| 1 | F2 | 177 | 240 | 1049 |  |  |
| 1 | F3 | 179 | 242 | 1012 |  |  |
| 1 | F4 | 179 | 242 | 1042 |  |  |
| 2 | B1 | 393 | 304 | 534 |  |  |
| 2 | F1 | 220 | 237 | 1006 |  |  |
| 2 | F2 | 220 | 237 | 362 |  |  |
| 2 | F3 | 220 | 237 | 278 |  |  |
| 2 | F4 | 220 | 237 | 215 |  |  |
| 2 | F5 | 220 | 237 | 298 |  |  |
| 2 | F6 | 220 | 237 | 565 |  |  |
| 2 | F7 | 220 | 237 | 273 |  |  |
| 2 | F8 | 220 | 237 | 265 |  |  |

From Table2, it shows the summary statistics of each floor in the two sites. Which including the floor plan information such as height and width, number of waypoints … etc.

… (can add more information here for other tasks)

**Data Preprocessing**

Due to the sparsity of the waypoint data which as ground truth, A technique of data augmentation is implemented thus augmenting the quantity of waypoints along the path of traces. These augmented positions’ data are as step positions and are used to present the estimated positional location of the surveyor at each timestamp.

We choose the TYPE\_ ACCELEROMETER and TYPE\_ROTATION\_VECTOR data as source to conduct the augmentation. For accelerometer magnitudes, we can use it to detect the movement of the site-surveyor [2]. When the magnitude of the accelerometer data arrived at a certain maximum or above threshold, which considered as the peak of the step. While as reduced to the minimum value or below a certain threshold, which is considered as a valley. Each peak counts as a step and the length of the stride (distance) for each step will be calculated by differencing between the two valley points of the accelerometer magnitudes. For rotation vector data, we use it to compute the direction of the site-surveyor. By using the step length, direction and actual timestamp, the directional position of the site-surveyor’s movement can be detected.

**Task: Visualization of Waypoints**

In this section, our task is to visualize the waypoints on a floor plane. Let’s take a trace of site1 – F2 - 5dda5a99c5b77e0006b1770b as an example.

Table 3: start & end points

|  |  |  |
| --- | --- | --- |
| **Timestamp** | **Px** | **Py** |
| 1574589279700 | 188.85158 | 43.87456 |
| 1574589391977 | 198.66891 | 104.87152 |

Map

Description automatically generated

Fig 1. Site 1 - F2 - 5dda5a99c5b77e0006b1770b

From Table 3 and Fig. 1, we can see the entry of waypoints data contains 3 parts: timestamp, Px coordinate and Py coordinate. The starting point at (188.85158, 43.87456) which in blue circle and the ending point at (198.66891, 104.87152) which in red circle. The discrete dots are the trace path. We use the provided function - visualize\_trajectory- and floor information to plot each trajectory of \*.txt files. It is showing the sparsity of the waypoints data.

Map

Description automatically generated

Fig. 2: Plot of Site1 - F2 - 1049 Waypoints

Chart, radar chart

Description automatically generated Chart, radar chart

Description automatically generated

(1). Original Waypoint (2). Augmented Waypoints

Fig. 3: Plots of the original waypoints and augmented waypoints for Site2 – F4

We visualized the waypoint features by Matplotlib and Seaborn function. For (1). Original Waypoint, first, we loaded plain floor image to the fig. Next, for each floor containing many text files. We processed each text file one by one, and the data points are TYPE\_WAYPOINT extracted and put them into group together. Waypoints Px, Py coordinates are added on top of the plain floor image which using the same color for marker "o" of each group of waypoints. The lines drawn are also the same color for each group as the group Markers, which represent the recorded points along the estimated trace along those points.

Fig. 2 and 3(1) are the plots of the most and least waypoint’s datapoints collected from the sample dataset respectively. For other waypoint’s plots can be seen from Appendix.

From Table 4, we can see that the waypoints of Site 1 are much more than those of Site 2. It is obvious evidence for the sparsity of waypoints. I also can observe clearly through Fig. 3(1), which are the gaps among waypoints are larger. Due to the sparsity that may cause the accuracy of collecting the data of sensors. We decided to perform a data augmentation by using the accelerometer and rotation vector’s data. As a result, by Fig. 3(2), the increase for the waypoints, shows the density of the trace path.

Table 4: Waypoints Original VS Augmented

|  |  |  |  |
| --- | --- | --- | --- |
| **Site** | **Floor** | **Waypoints** | |
| Original | Augmented |
| 1 | B1 | 1034 | 8468 |
| 1 | F1 | 975 | 8483 |
| 1 | F2 | 1049 | 9569 |
| 1 | F3 | 1012 | 9138 |
| 1 | F4 | 1042 | 8982 |
| 2 | B1 | 534 | 4363 |
| 2 | F1 | 1006 | 8759 |
| 2 | F2 | 362 | 2766 |
| 2 | F3 | 278 | 2142 |
| 2 | F4 | 215 | 1825 |
| 2 | F5 | 298 | 2270 |
| 2 | F6 | 565 | 4511 |
| 2 | F7 | 273 | 2326 |
| 2 | F8 | 265 | 2211 |

**REFERENCES:**

[1] M. Research, “Indoor location competition 2.0 (sample data and code),” 2020. [Online]. Available: https://github.com/locationcompetition/indoor-location-competition-20

[2] A. Abadleh, E. Al-Hawari, E. Alkafaween, and H. Al-Sawalqah, “Step detection algorithm for accurate distance estimation using dynamic step length,” in 2017 18th IEEE International Conference on Mobile Data Management (MDM). IEEE, 2017, pp. 324–327.