CSIT 6910D Independent Project Report

Indoor BLE Positioning Based on Fingerprinting Algorithm

Student Name: Wang Zhongyi

Student ID: 20618841

E-mail: zwangem@connect.ust.hk

Supervisor: Prof. Shueng-Han Gary CHAN

Abstract

This project is a continuous work of indoor localization from last semester. The goal is to achieve speed accurate velocity estimation, and we divided the localization into three parts: data collection, location estimation and velocity computation. Specifically, I mainly focus on the second part-location estimation. When drawing thoughts on this part, we would prefer an online estimation mode rather than purely offline estimation. Considering this, two feasible ways are proposed and can be applied to different settings flexibly. The first method is similar to nearest neighbors, the second is related to Bayesian Inference and Bayes model.

Contents

1 Preliminary

In order to achieve velocity estimation, one possible approach is to first acquire the estimated location of the target and then calculate its current speed in a minor span of time Δt.

1.1 Framework structure

The framework of the project consists of three parts: RSSI data collection, localization and velocity computation. Since RSSI data information is deployed last semester, the main work lies in location estimation part. According to the algorithms, changes can be made to collect data to obtain better performance. By studying the RSSI data, some insights about the distribution, frequency and some other useful statistics at different calibration points can be made and stored beforehand, when a new coming signal of the target arrives, these statistics then are utilized to estimate the location.

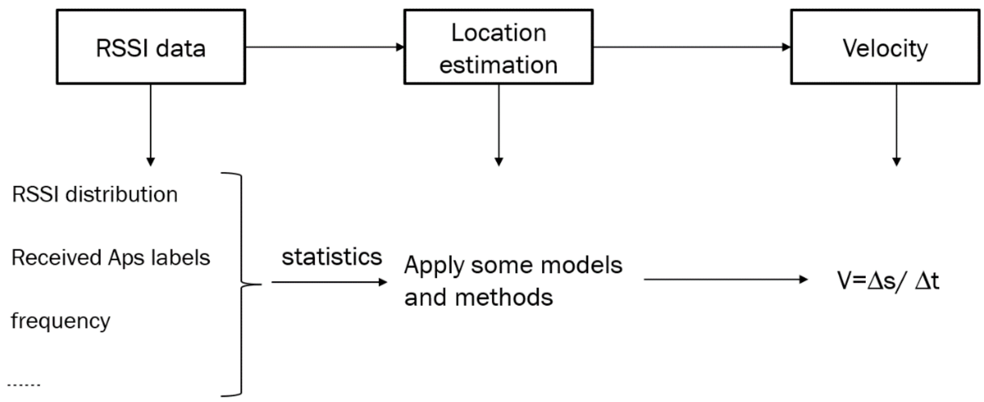

Figure 1.1 Framework of localization

Figure 1 shows the framework structure, two statistical models nearest neighbors and Bayes model are designed in the second part to estimate the location.

1.2 Setting Analysis

With the purpose to achieve good performance, studying the realistic setting is crucial. Figure 1.2, 1.3 and 1.4 show the actual BLE beacon deployment in different stations.

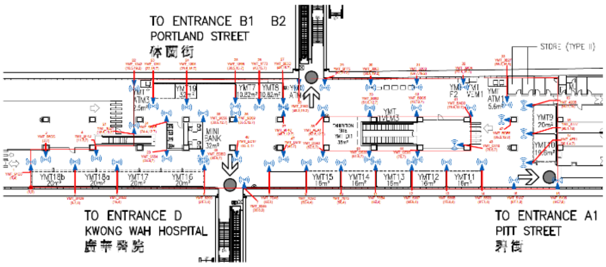

Figure 1.2 Floorplan of KLB station

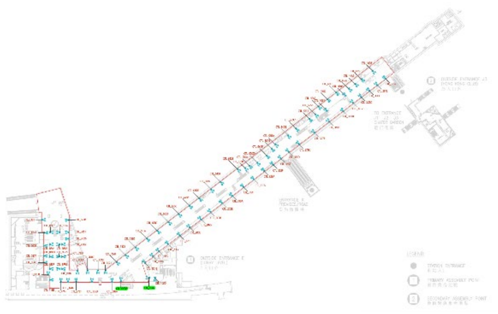

Figure 1.3 Floorplan of CTL station

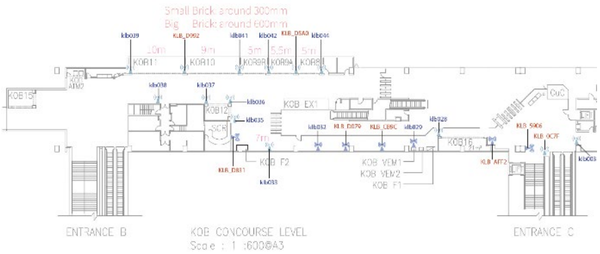

Figure 1.4 Floorplan of KLB station

Obviously, the paths along these station maps all have the shape comprises long thin segments. Based on this observation, the path can be simplified as kind of a rectangular region where beacons are deployed and the target cargo travels along. Figure 1.5 shows the simple version of the setting.

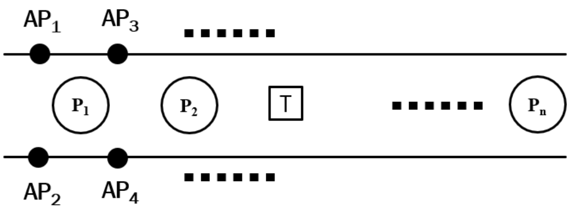

Figure 1.5 Simple version of setting

T is the target, AP_i denotes the beacons deployed along the path, P_j refers to the calibration points of which the coordinate is known and is used in calculating the target's estimated position.

2 Weighted Nearest Neighbors Method

In my implementation. There are four steps to derive the position using weighted nearest neighbor. First step is to collect the RSSI data at each P_j, mean values of different RSSI values at P_j of each AP are able to calculate. Step 2 involves normalization, fix one AP RSSI value as the standard value, and difference of other AP_i's RSSI value and the fixed value is the taken as AP_i's RSSI, this is the reference information stored to compare with the new arriving target signal. Step 3 is the online comparison phase, which applies distance measurement to calculate the distance between the target RSSI information and data stored in step 2, 2-nearest neighbors are therefore computed. Step 4 introduces a weight to derive a more accuracy weight estimation, the idea is that the target is not likely to locate at the exact one of the P_js.

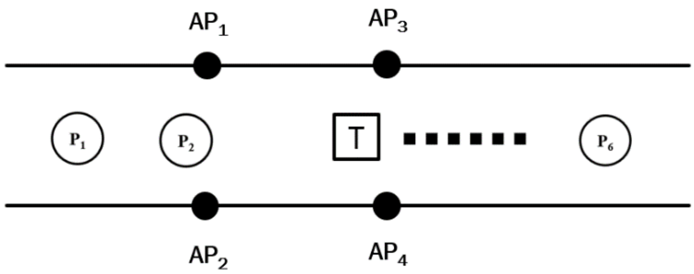

Figure 2.1 Weighted nearest neighbor approach

Figure 2.1 illustrates the idea of weighted nearest neighbor approach. Suppose there are 4 AP_i beacons along the path and 6 calibration coordinates $P_j(x_j, y_j)$.

In step 1, RSSI information is collected, mean is calculated and stored, as shown in table 2.1.

P_j	1	2	3	4	5	6
(x, y)	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
AP_1	-65	-55	-59	-63	-72	-80
AP_2	-66	-56	-58	-64	-74	-81
AP_3	-88	-82	-70	-60	-67	-75
AP_4	-89	-81	-69	-59	-66	-76

Table 2.1 Collected mean RSSI values at each fixed position

In step 2 normalization, set AP_1 as the standard point, and take the absolute subtraction $|AP_i - AP_1|$, then set all AP_1 values to 0. Result shown as table 2.2.

P_j	1	2	3	4	5	6
(x, y)	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
AP_1	0	0	0	0	0	0
AP_2	1	1	1	1	2	1
AP_3	23	27	11	3	5	5
AP_4	24	26	10	4	6	6

Table 2.2 Normalized RSSI value

In step 3, online estimation phase, let the input signal from the target be (-64, -65, -58, -57) for AP_1, AP_2, AP_3, and AP_4 sequentially. Perform the normalization same as in step 2, we transform the target signal to (0, 1, 6, 7). Now compute each Euclidean distance between the normalized target signal and the stored ones. The two smallest distance $\sqrt{3}$ and $\sqrt{2}$ are aligned with P_5 and P_6.

Finally, assign weights on the two selected coordinates, the weight for P_5 and P_6 are $w_5 = \frac{\sqrt{2}}{\sqrt{2}+\sqrt{3}}$ and $w_6 = \frac{\sqrt{3}}{\sqrt{2}+\sqrt{3}}$, respectively. The larger the distance, the smaller the weight is assigned. Consequently, the weighted estimation of the coordinate is $w_5 * (x_5,\ y_5) + w_6 * (x_6,\ y_6) = (1,\ 5.55)$.

3 Bayesian Inference

Another method I think about is to use Bayes rule to estimate the probable position given the signal value. The problem is how to define all the parameters in the Bayes rule formula (3-1).

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \tag{3-1}$$

Similar to the method in the weighted nearest neighbors, RSSI data are collected during the offline period, this time, an observation during a short span of time is defined as (3-2).

$$O_k = \{(\sigma_1, AP_1), (\sigma_2, AP_2), (\sigma_3, AP_3), \ldots, (\sigma_j, AP_i)\} \tag{3-2}$$

σ_j is the received RSSI value. Still, statistics such as mean and frequency of all the APs at each calibration point can be calculated and stored. Formula (3-3) and (3-4) are measurements that are well-designed.

$$P(S_k|O') = \frac{P(O'|S_k)P(S_k)}{\sum_{i=1}^{K} P(O'|S_k)P(S_k)} \tag{3-3}$$

$$P(O|Sk) = \prod_i^N P(f_i|S_k) \prod_j^M P\left(\sigma \middle| a_j, S_k\right) \tag{3-4}$$

Formula (3-3) is the online estimation phase, on the left-hand side, the highest probability of a location S_k given one single online observation is the desired estimated position, O' is the signal from the target O', O' has the similar form as (3-2). On the right-hand side, the statistics of the target signal such as mean and frequency can be used to query the stored information and then calculate the probability.

Formula (3-4) is the statistics stored during offline phase. The probability $P(f_i|S_k)$ is the relative frequency probability storing different RSSI value frequency for each AP at position S_k. In histogram view, it means at each calibration location S_k, there are histograms of RSSI values(as frequency) for every received beacon.

For example, suppose after miscellaneous(different time of day, etc) measurements at position S_1, AP_1, AP_2, AP_3 are received, so that the RSSI values are adequate to describe the distribution. We will get the histogram of AP_1 similar as follow:

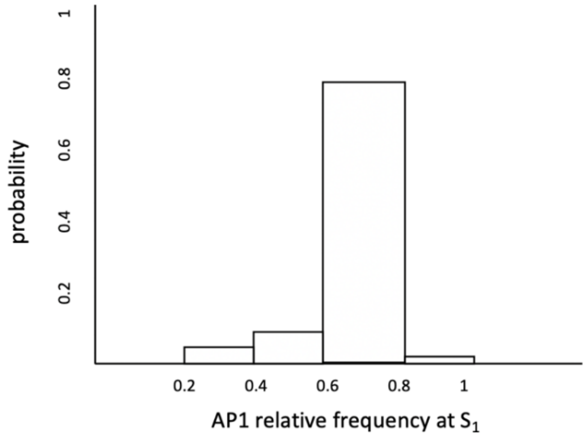

Figure 3.1 Offline AP_1's relative frequency-probability at S_1

Figure 3.1 means considering all the observations, the relative ratio of AP_1's signals. Sometimes, AP values may miss due to the environment and distance. At each location S_k, there are similar histograms of all the APs, from AP_1 to AP_n deployed.

The second probability $P(\sigma|a_j, S_k)$, the RSSI value distribution of a particular AP a_j at location S_k. Also, in histogram view, the distribution of different range of RSSI value is clear to obtain. The histogram for RSSI values of AP_1 will be something as follow:

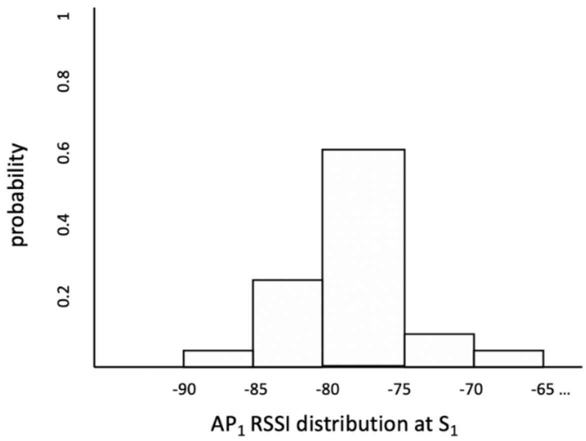

Figure 3.2 Offline AP_1's RSSI distribution-probability at location S_1

Similar histogram can be drawn for other APs received at this location. The probability $P(\sigma|a_j, S_k)$ is well-defined. Now the statistical data of those two probabilities are collected and stored in the database, and a current observation O' is received at target location for which the coordinate is unknown and to be estimated. O' has the similar form to (2-1). Then the histogram of O' can also be drawn.

Formula (3-4) and its parameters are given clearly, now we use the new arriving observation from the target signal to get the probability in the offline phase.

For the online estimation formula (3-3), let us say that at this target location S_t, AP_1, AP_2, AP_3 are detected and have the following histograms, one is the frequency:

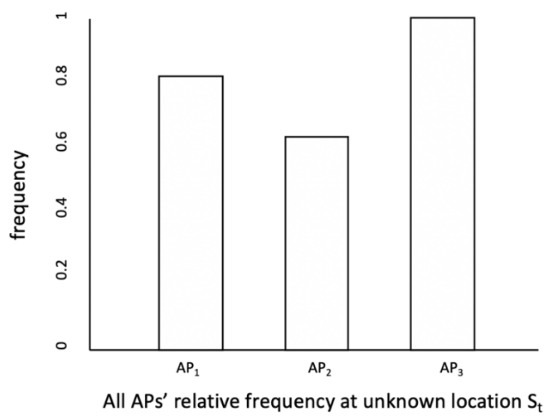

Figure 3.3 online relative frequency at location S_t

The other is the RSSI distribution:

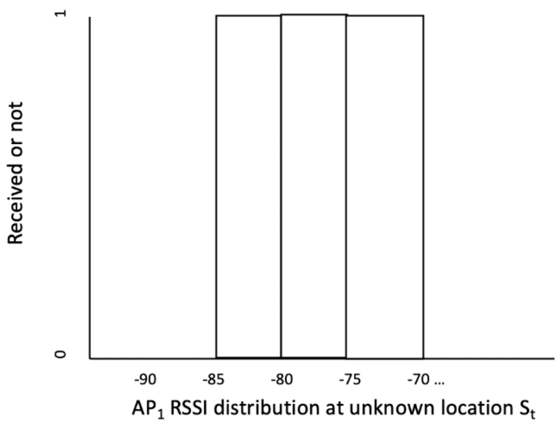

Figure 3.4 online RSSI distribution at location S_t

In the online phase, frequencies of all the received APs are known, and once the frequency is counted, the probability $P(O'|Sk)$ as in formula (3-4), consists of two probabilities, $P(f_i|S_k)$ and $P(\sigma|a_j, S_k)$. $P(f_i|S_k)$ can be acquired using the frequency in figure 3.3 and frequency-probability in figure 3.1. $P(\sigma|a_j, S_k)$ can be obtained by combine figure 3.4 and figure 3.2.

4 Improvements and Prospects

This project offers me a wonderful chance to study fingerprinting and related topics into more depth. I am able to apply knowledge and techniques learned in my lectures in practice, during the process, both my understanding and engineering skills are honed drastically.

Concerning the two methods, adaboost algorithms can further be implemented. Since the two methods can be regarded as weak classifiers. Additionally, one advantage is that weights for weak classifiers are computed in advance, and will not consume time during online phase, but serves as a strong tool to boost efficiency and effectiveness.

Personally speaking, I can improve myself in several ways. First is that I can be more active in communicating with other team members, I do not have to wait until figuring out some breakthroughs or important thoughts. Moreover, I could also go to the actual setting to have a deeper insight towards the project. Also, as a pragmatic project, experiment results are also very important, I can always doing some experiments even I may not have a very precise and concrete plan.

Appendix A

Minutes of the 1st Project Meeting

Date:	13 March 2020(Friday)
Time:	15:00
Place:	Zoom Meeting
Attending:	Prof. Gary Chan, Wang Zhongyi

1. Approval of minutes

    This is the first formal group meeting. Approved.

2. Discussion items

    - Specific topic
    - Preliminary thoughts on the project
    - Clarified main work tasks and directions

3. Meeting adjournment and next meeting

    The meeting is held on time. The next meeting will be held in about two or three weeks later by email announcement.

Minutes of the 2nd Project Meeting

Date: 30 March 2020(Friday)

Time: 15:00

Place: Zoom Meeting

Attending: Prof. Gary Chan, Wang Zhongyi

1. Approval of minutes

   Approved.

2. Discussion items

   - Current progress
   - Thinking and thoughts
   - Discussion about the deployment

3. Meeting adjournment and next meeting

   Postponed to 15:30. The next meeting will be announced by email.

Minutes of the 3rd Project Meeting

Date:	23 April 2020(Thursday)
Time:	15:00
Place:	Zoom Meeting
Attending:	Prof. Gary Chan, Wang Zhongyi, Zhong Shuhan, Chen Geng, Zeng Jiayu

1. Approval of minutes

   Approved.

2. Discussion items

   - Progress report
   - Questions and advice

3. Meeting adjournment and next meeting

   Postponed to 16:00. The next meeting will be around final.

Minutes of the 4th Project Meeting

Date:	21 May 2020(Thursday)
Time:	15:00-16:00
Place:	Zoom Meeting
Attending:	Prof. Gary Chan, Wang Zhongyi, Zhong Shuhan, Chen Geng, Zeng Jiayu

1. Approval of minutes

   Approved.

2. Discussion items

   - Final report assignments
   - Advice and conclusions on final progress

3. Meeting adjournment and next meeting

   On time. This is the last meeting.