Comparison of NVA and RLOWESS Algorithms in Indoor Positioning System

Hung-Huan Liu
dept. of electronic eng.
Chung Yuan Christian University
Chung Li, Taoyuan, Taiwan
hhliu@cycu.edu.tw

Wei-Chien Chen
master programm in comm. eng.
Chung Yuan Christian University
Chung Li, Taoyuan, Taiwan

Pin-Chun Hsu
dept. of electronic eng.
Chung Yuan Christian University
Chung Li, Taoyuan, Taiwan

***Abstract*—In the Wi-Fi-based indoor positioning system, by using signals of Wi-Fi access points (APs) but without the position of APs, the scene analysis method has better positioning accuracy. In the previous study, we proposed the quick radio fingerprint collection (QRFC) and neighboring vertices averaging (NVA) algorithm as a way to collect radio fingerprints. In this study, we compared NVA and RLOWESS, a well-known filtering algorithm from the point of view of positioning accuracy. A cluster AP problem which causing large positioning error when using Euclidean distance formula to estimate the position in a corridor is also discussed. From our experiment, positioning accuracy of NVA and RLOWESS are similar, which evidence the NVA is usable and can be used as a smoothing method of radio fingerprint.**

Keywords—Indoor Positioning, QRFC, NVA, RLOWESS

I. Introduction

In recent years, the development of wireless sensor network technologies drive several new applications that provided location-based services (LBS) to serve users based on their location. Among these applications, due to the penetration of Wi-Fi around the world along with affordable installation costs and acceptable positioning accuracy for some LBSs [[1],[2]], Wi-Fi-based Indoor positioning and navigation services have attracted wildly attention.

Wi-Fi-based positioning methods are currently divided into three main categories based on the positioning principles: proximity, trilateration, and scene analysis. Among them, the scene analysis is most commonly used because of the acceptable positioning accuracy. Scene analysis is divided into two phases. The first phase, also called offline phase, samples Wi-Fi signals offline. The RSSI of each AP and its basic service set identification (BSSID) is sampled at several predetermined positions in a building (called "sample points") and recorded in a database (called "radio fingerprint"). In the second phase, also called the position estimation phase, a user uses a smartphone to collect each AP's RSSI and BSSID at a specific located point of interest. Similarity measure is used to compare these measurements with the measurements stored in the radio fingerprint to determine the user's position. The nearest neighbor in signal space algorithm was proposed in[3] as a means of computing the signal space distance between the observed and recorded measurements by Euclidean distance. The nearest neighboring sample point in signal space is regarded as the user's position. The advantage of this approach is its ability to reduce multipath problems[4]; however, the disadvantage is the density of the sample points directly affects the positioning accuracy.

There is a key problem of scene analysis positioning is the high cost of the offline radio fingerprint building process, which is a time- and labor-consuming task. In the traditional approach, a coordinate system for the building is created beforehand, marking the location of every sample and then collecting the AP information at each sample point one by one. At each sample point, generally, several tens of samples are performed and the average RSSI and their variance are recorded in the fingerprint database. This procedure of the traditional signal sampling is named as the static sampling (SS) method in this study.

In our previous work, we proposed a quick radio fingerprint collection (QRFC) algorithm [5][6] that uses moving sampling (MS) and stepped moving sampling (SMS) to reduce the time required for signal sampling, and an AP RSSI shaping, called neighboring vertices averaging (NVA), to calibrate collected Wi-Fi signals. In this paper we compare the performance between NVA and robust locally weighted scatterplot smoothing (RLOWESS)[7]. A cluster AP problem which causing large positioning error when using Euclidean distance formula to estimate the position in a corridor is also discussed.

The remainder of this paper is organized as follows. In Section 2, SMS, NVA, and RLOWESS and the respective analyses of Wi-Fi signals collected using these three methods are described In Section 3, presents the experimental results and a related discussion. Conclusions are offered in Section 4.

II. Wi-Fi Signal Sampling and Smoothing Methods

The Wi-Fi signal sampling methods used in the present study were divided into two types: SS and SMS. SS is the conventional method for Wi-Fi signals collection. It requires to build a coordinate system for the indoor environment, chose target sampling points beforehand, and sample signals at each point in order. In general, several tens of samples are collected and the average RSSI is used as the final record. On the other hand, SMS samples the Wi-Fi signals along predetermined routes inside a building. For every step taken during sampling, the user stop and send Wi-Fi scan request and does not take the next step until the phone has acquired the signal data. SMS is sometimes viewed as a variant of SS where sampling occurs once for each step. After sampling on a path, assuming that the step length is equal, interpolation is used to calculate the position of each step and set the position as sample points.

Radio map constructed by SMS can be expressed by $\boldsymbol{S_{sms}}$, which is the set of all signal samples, that is

$$\boldsymbol{S_{sms}} = \{\boldsymbol{M_i}\},\ 0 < i \le N, \quad (1)$$

where i is the index of sample point and there are total N sample points in the testing area, and

$$\boldsymbol{M}_i = \cup_{l \in I_A}(i, r_{i,k}, b_k), \quad (2)$$

where I_A is the set of all possible APs, b_k is the unique MAC number of access point AP_k, and $r_{i,k}$ is the value of the RSSI value measured from the AP_k.

In similar fashion, $\boldsymbol{S}_{ss}$ is the radio map constructed by SS and can be expressed by

$$\boldsymbol{S}_{ss} = \{\boldsymbol{S}_i\},\ 0 < i \leq N. \quad (3)$$

Remind that SS performs several samples at one sample point so that has a sequence of RSSIs from one AP, the j^{th} sample has the form

$$\boldsymbol{S}_{i,j} = \cup_{l \in I_A}(i, r_{i,j,k}, b_k), \quad (4)$$

where $r_{i,j,k}$ is the value of the RSSI value measured from the AP_k at the j^{th} sample, and

$$\boldsymbol{S}_i = \cup_{l \in I_A}(i, \bar{r}_{i,k}, b_k) \quad (5)$$

the $\bar{r}_{i,k}$ is the mean RSSI, which is

$$\bar{r}_{i,k} = \frac{1}{|r_{i,k}|}\sum_{j=1}^{|r_{i,k}|} r_{i,j,k}, \quad (6)$$

where $|\boldsymbol{r}_{i,k}|$ is the number of samples performed at sample point i.

In the [5][6] study, we proposed the QRFC for use in wireless indoor positioning systems. Its main feature is that compared to the traditional SS, QRFC can save a lot of time on wireless fingerprint database construction while maintaining the same level of positioning accuracy. A heuristic NVA algorithm was proposed in this architecture as a way to RSSI shaping. The details of the algorithm from[5] are excerpted as follows:

NVA Algorithm:

Preliminary:

1. Set the sequence of raw RSSI to follow the sampling sequence and is described as $S = \{(\boldsymbol{X}_0, r_0), (\boldsymbol{X}_1, r_1), \ldots, (\boldsymbol{X}n, r_n)\}$, where $\boldsymbol{X}$ is a coordinate position and r is an RSSI value. If r_p denotes the highest RSSI value, $0 \leq p \leq n$. Thus, the right and left halves respectively form two different sets denoted respectively by S_R and S_L, defined as follows:
   $S_R = \{(\boldsymbol{X}_p, r_p), (\boldsymbol{X}_{p+1}, r_{p+1}), \ldots, (\boldsymbol{X}_n, r_n)\}$
   $S_L = \{(\boldsymbol{X}_p, r_p), (\boldsymbol{X}_{j-1}, r_{p-1}), \ldots, (\boldsymbol{X}_0, r_0)\}$
2. Set the windows size to w.

S_L:

1. All r values are copied once and named sr.
2. for each $i = p$ to n-1
3. for each $j = i + 2$ to k, where $\max_k\{\boldsymbol{X}_k - \boldsymbol{X}_i \leq w\}$
4. Draw a straight line connecting r_i and r_j.
5. If sr_a is lower than this line, sr_a is updated as a value for the line at $\boldsymbol{X}_a$, where $i < a < j$.
6. end j
7. end i
8. Correct the margin as follows:
   if ($r_n \geq sr_{n-1}$),
9. do nothing
10. else
11. Extend the line connecting sr_{n-2} and sr_{n-1} to $\boldsymbol{X}_n$. If sr_n is lower than the line,
    sr_n is updated as a value for the line at $\boldsymbol{X}_n$.

S_R:

12. Repeat the steps performed for S_L but in the opposite direction.

Output:

13. Save sr as the modified RSSI distribution.

There are three reasons to use NVA:

1. In an indoor environment, because of the serious multipath fading problem, the AP RSSI received by the Smartphone changes drastically, but according to the propagation characteristics of the wireless signal, RSSIs received in two adjacent locations should have some degree of relevance. We consider the point with stronger signal is the point at which less affected by the multipath fading, and the point where the neighbor signal is weaker is considered to be more affected by the multipath fading at the time. NVA can compensate for signals that are greatly affected by multipath fading.
2. Different similarity measures are used to find the best match between observations and the radio map. The common choice for the comparison measure is to use the Euclidean distance to assign a non-negative value to the fingerprint $\boldsymbol{M}_i$ or $\boldsymbol{S}_i$ [3]. As we know that the RSSI distribution of APs improve the accuracy of indoor positioning algorithms to a higher degree than does RSSI value. NVA algorithm suggest that the RSSI distributions of APs obtained through SMS was similar to those obtained through SS.
3. NVA features low calculation load and acceptable positioning accuracy compare with other methods.

Filtering is an effective method to reduce the noise of the signal. However, the RSSI received at a sample point by smartphone does not follow the Gaussian distribution, the linear filters and Gaussian-based filters are improper to adopt in indoor environment. Meanwhile, due to the difficulty of specifying the functional form of the whole data set, it is also hard to smooth the data with parametric regression. The RLOWESS is a nonparametric statistical approach and is used to smooth the original measured RSSI in this study for the sake of comparing and evidence the usability of the NVA algorithm.

Similar with the NVA algorithm, the raw sampling sequence is $S = \{(\boldsymbol{X}_0, r_0), (\boldsymbol{X}_1, r_1), \ldots, (\boldsymbol{X}n, r_n)\}$, for any $r_i \in S$, the result of RLOWESS can be defined as

$$r_i = \hat{r}_i + \varepsilon_i \quad (7)$$

where $\hat{r}_i$ is the smooth result of r_i and ε_i denotes a random variable for estimation error. Because the RLOWESS is a well-known algorithm, we ignore unnecessary explanation and use python statsmodels [8] to perform the RLOWESS calculation.

Our goal is to compare the positioning errors for NVA and RLOWESS. Several similarity measures are considered to find the best match between observations and the radio map.

The common choice for the measure is the Euclidean distance which assigns a non-negative value to the fingerprint $\boldsymbol{M}_i$ or $\boldsymbol{S}_i$ [3]. For instance, if $\boldsymbol{y}$ is the observation at certain position, then the distance between $\boldsymbol{y}$ and $\boldsymbol{S}_i$ is

$$dist(\boldsymbol{y}-\boldsymbol{S}_i)^2 = \sum_{k\in I_{A,i}}\left(y_k - \bar{r}_{i,k}\right)^2 \qquad (7)$$

where $I_{A,i}$ is the set of APs which both seen by $\boldsymbol{y}$ and $\boldsymbol{S}_i$. For SMS, let $\widehat{\boldsymbol{M}}_i = \bigcup_{l\in I_A}\left(i, sr_{i,k}, b_k\right)$ is the output of NVA at sample point i,

$$dist(\boldsymbol{y}-\widehat{\boldsymbol{M}}_i)^2 = \sum_{k\in I_{A,i}}\left(y_k - sr_{i,k}\right)^2 \qquad (8)$$

For the convenience of the follow-up instructions, we define three symbols here: $\widehat{\boldsymbol{R}}_i = \bigcup_{l\in I_A}\left(i, rl_{i,k}, b_k\right)$ is the output of RLOWESS at sample point i, and $\boldsymbol{S}_{\boldsymbol{nva}} = \{\widehat{\boldsymbol{M}}_i\}$, $\boldsymbol{S}_{\boldsymbol{rl}} = \{\widehat{\boldsymbol{R}}_i\}\ 0 < i \le N$, are the radio fingerprint constructed by NVA and RLOWESS, respectively.

III. Experimentation and Discussion

Figure 1 shows the floor plan of the experimental site. The building has eight floors. Each floor has staircases on the left- and right-hand sides and another stairway and elevator in the middle. The fifth floor has an open space for students to engage in recreational and academic activities. The experiment was conducted in mid-May 2019, we collected AP signals on a path from (1, 0) to (55, 0) along the hallway of this floor by SMS and SS, using the HTC M8 android phone. The HTC M8 is a pretty old smartphone released by March 2014. The reason for using this phone is that the new version of Android (since Android 8) limits the functionality of Wi-Fi Scan, making it impossible to perform wireless signal scanning on newly released Android smartphone.

In this building, over 40 AP signals were captured in each scan, totaling more than 600 APs, thereby indicating that AP deployment is dense. Four observations are made for estimating the positioning error. A man with a height of about 175cm carries out position estimations by holding a mobile phone. Blue dots in Fig. 1 denote the position of observations and the arrow indicate the directions the person is facing. The coordinates of point A, B, C, and D are (5.4, 0.8), (15.9, -0.8), (31.9, 0.9), and (49.9, 1) respectively.

Three AP Signals are chosen to explain the smoothing results, shown in Fig. 2. They are one of the strongest AP signals around the right, middle, and left side of the corridor. Stronger signals played a crucial role in the positioning algorithms, whereas weaker signals were eliminated from the algorithms. The curve labeled ss-mean denotes the raw data collected by SS and taking average; ss-nva and ss-rl are the

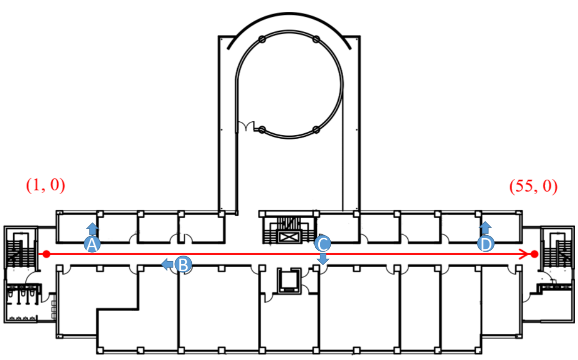

Fig. 1 Floor plan of the experimental site.

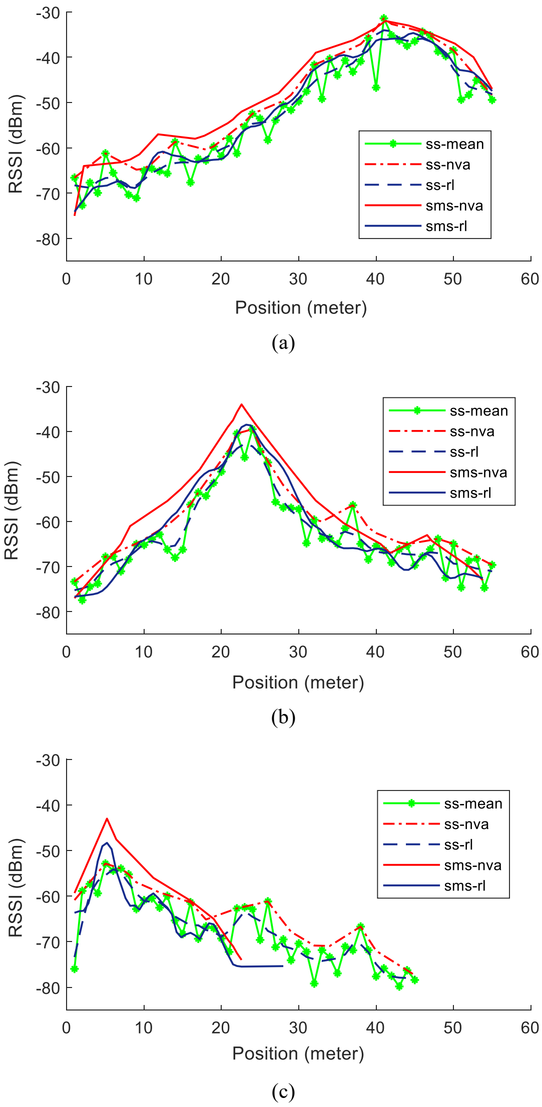

Fig. 2 RSSI distributions of signals collected from various APs (a) signals collected at the right side of the path, (b) middle, (c) left side.

curve which smoothing ss-mean using NVA and RLOWESS, respectively. The curve labeled by sms-nva and sms-rl are the curve of raw data collected by SMS and then smoothed through NVA and RLOWESS, respectively.

Due to wiring and signals coverage limitation, the APs' installation positions are usually concentrated at certain locations. For example, most APs are located at the center of corridors, with none being situated in staircases. The stronger signal detected at the middle is perhaps from campus network, with strong signals and large coverage, Fig. (a), (b). On the other hand, the stronger signals detected at the right and left sides perhaps come from AP routers set up by individual laboratories, with smaller signals and coverage. As shown in Figure 2(c), the signal collected by the SMS ends at around 25 meters, because it is obscured by the stronger signal in the middle part.

In order for equation (7) to be calculated, a predetermined value p is proposed. The p strongest APs' information detected by observation point, and the p APs have record in the fingerprint, are used for position estimation. With p, a pre-

Positioning Error for Different Scenarios

Observations	Independent test		First *n* RSSI were averaged			
	mean[a,b]	*std*	*n=1*	*5*	*10*	*30*
A	1.07	0.82	1.40	2.00	1.40	1.40
	1.52	1.95	1.00	1.00	1.00	1.00
	2.69	2.27	1.60	1.60	5.60	0.40
	1.15	1.98	0.60	0.60	0.60	0.60
	2.70	3.56	0.60	1.60	1.60	0.60
B	1.74	0.34	1.70	1.70	1.70	1.70
	0.42	0.16	0.50	0.50	0.50	0.50
	1.10	0.00	1.10	1.10	1.10	1.10
	0.39	0.39	0.10	0.10	0.10	0.10
	0.90	0.40	1.10	1.10	1.10	1.10
C	5.36	7.40	2.70	2.10	21.90	2.70
	1.34	0.31	1.50	1.50	2.10	1.50
	5.97	3.25	8.10	8.10	8.10	8.10
	2.81	4.91	1.90	1.90	21.10	1.90
	1.28	0.82	0.90	0.90	2.10	0.90
D	4.34	0.27	4.50	4.50	4.50	4.50
	3.84	0.32	3.90	3.90	3.90	3.90
	6.02	7.73	19.90	1.10	1.10	1.10
	18.83	5.86	21.90	20.90	20.90	20.90
	3.65	4.64	3.10	2.10	2.10	2.10

a. top to bottom: sms-nva, sms-rl, ss-mean, ss-nva, ss-rl

b. unit is meter

filter is executed, a subset of $\boldsymbol{S_{ss}}$, $\boldsymbol{S_{nva}}$, and $\boldsymbol{S_{rl}}$ for each observation point is selected.

In this experiment, 30 samples were performed for each observation, and the interval between each sample was about 2.5 seconds (there is no way to be faster because of the limitations of Android phones). Table 1 shows the mean and standard deviation of the positioning error obtained after the 30 pieces of data were independently evaluated. Table 1 also shows the results of the position evaluation after the first *n* pieces of RSSI were averaged first, where *n* = 1, 5, 10, 30.

The values listed in Table 1 are the projections of errors on the x-axis. In the case of observation point B, the coordinate of B is (15.9, -0.8). In SS mode, the distance from the nearest sample point is 0.1 meter. Therefore, some errors listed in Table 1 are 0.1, which is actually the best result. The positioning results of the five methods are actually similar, and the numbers may vary in different experiments. However, they are all within acceptable range and are suitable for use in pedestrian indoor navigation and guidance applications. Comparing between sms-nva and sms-rl, they are neck and neck in the positioning accuracy among the point A, B, and C. However, sms-nva has huge positioning error on point D.

At observation point D, the use of the Euclidean distance measure produces a large error that often occurs at the boundary of a single path lacking good wireless network planning. Because of wiring and signals coverage limitation, APs with strong signal are installed at around 40 meters in this floor, as shown in Fig. 2(a). If -50dBm was observed at point D, according to the Euclidean distance formula, two local minima will be generated at around 50 meters and 30 meters. Unfortunately, the values calculated near 30 meters won the battle in some cases. The individual testing result for observation D is shown in Fig. 3. Not every experiment presents the same results, and we just take a special case to illustrate using the Euclidean distance formula with no refined

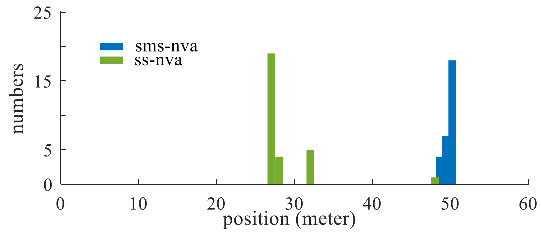

Fig. 3 The individual testing result for observation D (total 30).

radio fingerprint to estimate the position in the corridor is not a good choice.

IV. Conclusion

The indoor positioning system based on Wi-Fi signal has become the most important technology for the applications of indoor pedestrian navigation and guidance. This paper compares the use of the NVA algorithm and the RLOWEE algorithm to shape the acquired wireless signals in a corridor environment to show the evidence of the usability of the NVA. In our experiment, RLOWESS and NVA have similar performance on positioning accuracy while NVA has less computational complexity than RLOWESS. Along with SMS, the radio fingerprint can be quickly created with acceptable positioning accuracy. In addition, this study also found that in a space without good planning, Wi-Fi APs are usually concentrated installed in a certain area, which makes the positioning algorithm based on Euclidean distance have an opportunity to generate large positioning errors, especially at the borders of the corridor. This problem is also what we want to study and solve in the future.

References

[1] C. Yang, C. and H.R. Shao, “WiFi-Based Indoor Positioning,” *IEEE Commun. Mag.*, vol. 53, 2015, pp. 150–157.

[2] Y. Zhuang, Z. Syed, Y. Li, and N Ei-Sheimy, “Evaluation of Two WiFi Positining Systems Based on Autonomous Crowdsourcing of Handheld Devices for Indoor Navigation,” *IEEE Trans. Mob. Comput.*, vol 15, 2015, pp. 1982–1995.

[3] P. Bahl and V. N. Padmanabhan, "RADAR: an in-building RF-based user location and tracking system," in *Proc. IEEE INFOCOM 2000*, pp. 775-784, May 2000.

[4] H. Liu, H. Darabi, P. Banerjee and J. Liu, “Survey of Wireless Indoor Positioning Techniques and Systems,” *IEEE TSMCC*, vol. 37, 2007, pp. 1067–1080

[5] H.-H. Liu and C. Liu, “Implementation of Wi-Fi Signal Sampling on an Android Smartphone for Indoor Positioning Systems,” *Sensors*, vol. 18, no. 2, p. 3, Dec. 2017 [Online]. Available: http://dx.doi.org/10.3390/s18010003

[6] H.-H. Liu, “The Quick Radio Fingerprint Collection Method for a WiFi-Based Indoor Positioning System,” *MONET*, vol. 22, issue 1, 2017, pp. 61–71.

[7] W. S. Cleveland, “Robust locally weighted regression and smoothing scatterplots,” *Journal of the American statistical association*, vol. 74, no. 368, pp. 829–836, 1979.

[8] StatsModels, Statistics in Python, https://www.statsmodels.org/devel/generated/statsmodels.nonparametric.smoothers_lowess.lowess.html, available at: May 5 2019