Improved Wi-Fi AP position estimation using regression based approach

Youngsu Cho, Myungin Ji, Yangkoo Lee, Jooyoung Kim, Sangjoon Park
Electronics and Telecommunications Research Institute
Positioning Information Technology Research Team
Daejeon, Republic of Korea
E-mail: choys@etri.re.kr, myungin@etri.re.kr, yk_lee@etri.re.kr, kimjy@etri.re, sangjoon@etri.re.kr

***Abstract*—This paper describes the improved Wi-Fi AP position estimation method for building more accurate Wi-Fi AP position DB in complex indoor signal propagation environment. One of our project's goals is to produce the estimated position DB of Wi-Fi AP by using indoor survey for higher Wi-Fi AP position accuracy. Our contribution focuses on the Wi-Fi AP position estimation method. In previous works, there were several methods for Wi-Fi AP position estimation by using such as weighted centroid localization (WCL), the trilateration using traditional signal propagation model etc. In case of WCL, the estimated Wi-Fi AP position depends on the amount and location of collecting points (CP) so that the Wi-Fi AP position accuracy can be poor in low DoP (Delusion of Precision) collecting environment, like the corner or edge of a building. Trilateration method can be more efficient than weighted centroid in case of low DoP collecting scenario but its method using the traditional path loss model is vulnerable to complex indoor area where lots of attenuation, reflection and refraction exist. To enhance Wi-Fi AP position accuracy in case of low DoP collecting environment, the Wi-Fi AP position estimation using Gaussian process regression (GPR) is proposed. Firstly, the Wi-Fi AP received signal strength (RSS) using GPR is predicted and analyzed whether it can be used to estimation Wi-Fi AP position. Next, the comparison of Wi-Fi AP position accuracy by using WCL with by using GPR/WCL is described. From the test result, we will show that the position accuracy of the proposed method will be better than using WCL.**

Keywords: *Wi-Fi, AP, Gaussiran Process, Regression, position, estimation*

I. INTRODUCTION

Currently, the mobile devices' localization using Wi-Fi access point (AP) RSS is one of the popular technologies to obtain its location in GNSS-less environments, in especially indoor area. Considering the explosive increase of smartphones which mostly support Wi-Fi connectivity and the extension of Wi-Fi access points as the alternative communication infrastructure for dispersing the network traffic, Wi-Fi will be quickly deployed in the world-wide level, which means that they can be the important global localization resources like mobile base stations.

Specifically indoor Wi-Fi based positioning technologies can be classified as two categories, which use fingerprint DB [1] or Wi-Fi AP position DB [2]. Methods using fingerprint DB obtain relative higher position accuracy but it needs high cost for initial survey and re-collection. Besides the fingerprint DB is weak to the change of indoor environment. Meanwhile, algorithms using Wi-Fi AP position DB have less accurate position accuracy but its DB size is very compact and it is not affected by the change of indoor environment because the type of data is the set of position, not RSS. If algorithm using Wi-Fi AP position DB is considered in any system, the determination of the Wi-Fi AP position should be preceded. Most of previous researches assumed that this information was already known, for example, by site survey. However, there is lots of unsupervised Wi-Fi APs which are impossible to be known because some of them are securely managed by private users not by mobile network operators or service providers.

To obtain the Wi-Fi access points' position quickly and globally for metropolitan level Wi-Fi positioning, some collecting approaches have been developed. Firstly, most commercial Wi-Fi positioning systems (WPS) from Skyhook-wireless[3] use war-driving technology, which continuously collects the driving car's position and the surrounding wireless signal measurements(e.g. Wi-Fi RSS) in order to construct wide-area Wi-Fi APs' position database. The probe car, however, only moves along on the road so that more detailed collection closed to surrounding access points is not possible. To overcome this problem, user based crowd-sourcing technology was suggested [4]. It collects Wi-Fi measurements by user's contribution while stopping or moving. For example, user's location was obtained by clicking on the map. Especially, this approach can enhance the accuracy of Wi-Fi APs' position through indoor collecting process.

Not only various collecting approaches but also several positioning algorithms for Wi-Fi AP were developed. The first simple and quite accurate algorithm is the weighted centroid localization [5]. For each access point, all collecting positions were weighted averaged, where weighting factors were proportional to the RSS. The performance of this method was quite good in case that all collecting points were well-distributed in the vicinity of each access point. However, the accuracy was not guaranteed if this distribution was biased or fewer. The second algorithm is to calculate Wi-Fi APs' position by using trilateration [6]. It can provide light computation due to its closed-form solution but the unpredictable indoor signal propagation environment such as reflection, fading, refraction and scattering etc. makes its location performance poor.

To enhance the accuracy of Wi-Fi AP's position in indoor environment where non-uniformed collecting points and unpredictable signal propagation both exist, in this paper, we propose the improved Wi-Fi AP position estimation algorithm which combine WCL with a probabilistic regression, especially GPR. The details are explained from following sections and this paper is organized as follows. In section II, background theories of Gaussian process regression are simply shown. Then we explain how it is used to calculate Wi-Fi APs' position. To evaluate the suggested algorithm, the experiment is conducted and its setup and results are explained in section III and IV. Finally, the conclusions and future work are suggested.

II. WIFI AP POSITION ESTIMATION USING GAUSSIAN PROCESS REGRESSION

A. Gaussian process for Wi-Fi signal strength regression

The previous works related to Gaussian process [7] show that Gaussian process can be used to overcome current algorithm's limitation from existing approaches such as trilateration and WCL. Firstly Gaussian processes (GPs) are non-parametric regression models while the conventional signal propagation model requires parameters like attenuation coefficient which are usually dependent on the environment. As a result, by using GPs, signal strength from each AP at arbitrary locations can be predicted more accurately than using signal propagation model if well-averaged collecting points are totally covered in the indoor environment. In case of WCL, the distribution of training points has directly effect on the accuracy of each Wi-Fi AP position. For example, the biased collecting points which can be usually located at the corner of each floor give rise to poor Wi-Fi AP position accuracy. However, if GPs can be used, the RSS at the virtual collecting points in the grid can be predicted regardless of the location and distribution of real collecting points. That is, GPs can convert received signal strength map at random collecting points into received signal strength map at all grid points evenly and precisely spaced.

From preliminaries, a Gaussian process estimates the posterior distribution over functions from training data. If we assume that the Gaussian kernel is used as the kernel function, the posterior over function value is Gaussian with mean μ_{X^*} and variance $\sigma^2_{X^*}$:

$$p(f(X^*) \mid X^*, X, y) = N(f(X^*); \mu_{X^*}, \sigma^2_{X^*}) \quad (1)$$

where, $\mu_{X^*} = k_*^T (K + \sigma_n^2 I)^{-1} y$

$$\sigma^2_{X^*} = k(X_*, X_*) - k_*^T (K + \sigma_n^2 I)^{-1} k_* \quad (2)$$

Here, the kernel function as the one of covariance function is generally defined as follows.

$$k(X_m, X_n) = \sigma_f^2 \exp\left(-\frac{1}{2l^2} \left| X_m - X_n \right|^2\right) \quad (3)$$

Before estimating the mean and variance of RSS at the arbitrary point, hyper parameters ($\sigma_f^2, \sigma_n^2, l$) should be estimated by maximizing the marginal log-likelihood of the observation. They are maximized by using conjugate gradient decent method and are calculated by using GPML matlab toolbox.[8] Using obtained hyper parameters and equation (1)-(3), we can predict the mean and variance of the function value at an arbitrary point. The mean value represents the received signal strength and the variance does the uncertainty at this value.

Fig. 1 shows that the GPR modeling for one of access points in the test area. The mean and variance values are estimated at all the location where collecting data are not available. This information obtained from each AP is used to estimate its position which method is explained in the next session.

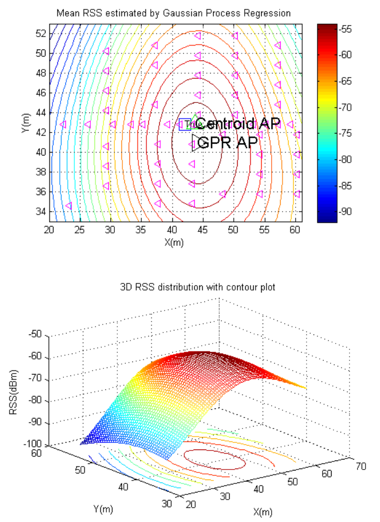


Figure 1. Mean RSS prediction using GPR for one AP

B. Wi-Fi AP position estimation using GPR

As the first step to estimate Wi-Fi AP position estimation using GPR, we need to define grid points at which the mean and variance of Wi-Fi AP RSS are predicted. The grid size should be determined by considering the accuracy of the estimated Wi-Fi AP position and here we set it to 0.5m. Note that, for the sake of generating of efficient grid points, we have one assumption that all Wi-Fi APs are installed inside a building and grid points should be defined within the test area. It may be reasonable because most of Wi-Fi APs are installed on the ceiling or wall to provide wireless communication to in-building residents in the office or at home and are also cheaply installed indoors by just to use already deployed power and network cable.

With this assumption, the mean and variance of Wi-Fi AP RSS at each grid point are predicted using GPR. Then we calculate a grid with maximum Wi-Fi AP RSS as an initial AP position. If the mean of Wi-Fi AP RSS at arbitrary grid is estimated from noisy and insufficient training data, the estimated Wi-Fi AP RSS error will increase and then local maximum problem may happen. So we need to have a check process to avoid Wi-Fi AP position to settle down at the local maximum. In this paper, we compare the Wi-Fi AP position using GPR with its position using WCL. If it is far from the pre-defined threshold (here the upper limit of WCL's position accuracy, 10m), we continue to choose the grid point with next maximum Wi-Fi AP RSS as a new Wi-Fi AP position. Finally this process is repeated until it converges to the threshold.

Even though most of Wi-Fi AP RSS prediction using GPR will be executed well, it exceptionally may fail when the correlation between the grid point(X) and Wi-Fi AP RSS(Y) is hard to model due to distorted Wi-Fi AP RSS collecting data at each grid point. In this case, as an alternative process, the final AP position estimation is determined by only using WCL.

Here, WCL calculates the weighted average of training data and is used to evaluate the performance of the proposed algorithm. The basic theory is defined as follows.

$$\hat{x}_{WCL} = \sum_{i=1}^{n}(wf_i \times x_i), \qquad \hat{y}_{WCL} = \sum_{i=1}^{n}(wf_i \times y_i) \qquad (4)$$

where, wf_i : weighting factor which is proportional to the received signal strength from each AP

III. EXPERIMETAL SETUP

The experiment was performed indoors in the 4th floor of one of ETRI's building, Daejeon, Republic of Korea. To efficiently evaluate the position accuracy of the estimated Wi-Fi AP, we deployed the 25 reference Wi-Fi APs, which are almost uniformly distributed in a test bed. In our test, a smartphone was used to obtain Wi-Fi RSS data at each collecting point, where Wi-Fi scan was performed at 0.4Hz and Wi-Fi scan duration is almost 30 seconds. During the training process, we walked and stopped at 437 known reference points in one floor and recorded the collecting data, which combine Wi-Fi RSS and the reference position. The layout and detailed specification of test bed are shown in Fig. 2 and Table. I.

TABLE I. EXPERIMENT CONDITION

Size of testbed	68m x 53 m (1 whole floor)
Number of Wi-Fi AP	25
Number of training point	437
The scan time at each trainingpoint (number of scan data at each pt.)	30s (12)
Device(scanning frequency)	Motorola RAZR(2.5s per scan)
Trainingduration(excluding preparation time before training)	437 x 30s = 13110s≈219min ≈ 3.64hour
Iteration number of process for Wi-Fi AP position estimation	10

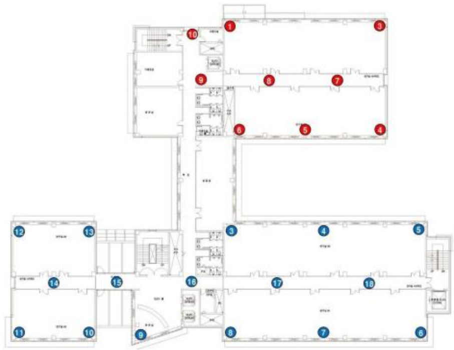


Figure 2. Testbed Configuration(The layout of test APs)

IV. EXPERIMETAL RESULTS

The performance of proposed algorithm was demonstrated by comparing it with a reference algorithm using WCL under the two conditions which have an effect on the positioning accuracy. In our test, two conditions are defined to the number of data at each collecting point and the distance interval between collecting points. The effect of these conditions on the positioning accuracy is analyzed in the following sections.

A. *The effect of the number of collecting data at each collecting point on position accuracy*

The Fig. 3 shows the averaged position accuracy of all Wi-Fi APs as a function of the number of collecting data at each collecting point. In case of both GPR/WCL and WCL, the averaged position accuracy of total Wi-Fi APs is almost even when the number of collecting data is getting more. This means that the number of collecting data has little effect on the position accuracy. So the dynamic (or in-flight) collecting approach, which obtains only 1 Wi-Fi scan dataset per collecting point, can be available while the position accuracy of both GPR/WCL and WCL still remains.

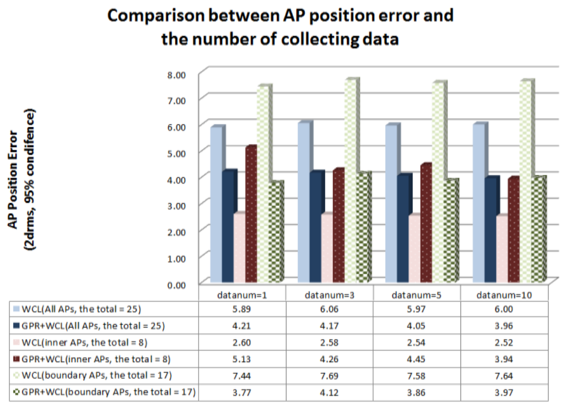


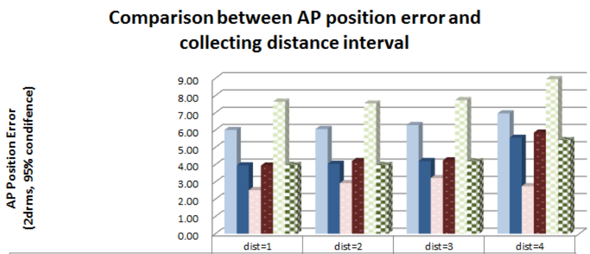
Figure 3. The position accuracy of Wi-Fi APs as a fuction of the number of collecting data at each collecting point.

For the further analysis of the effect of low DoP collecting environment on the position accuracy, the total APs are classified as two categories, inner and boundary APs. As a test results, both inner APs and boundary APs have little effect on the position accuracy when the number of collecting data increases.

B. The effect of the distance interval between collecting points on position accuracy

To evaluate the effect of the distance interval only, we set the number of collecting data to 10 to use the reliable collecting data per each reference point.

The Fig. 4 shows the averaged position accuracy of total Wi-Fi APs as a function of the distance interval between collecting points. In case of both GPR/WCL and WCL, the position accuracy is degraded more when the distance interval is getting longer. Furthermore, the best positioning accuracy using GPR/WCL is 3.96m at the 1m distance interval and totally the positioning accuracy using GPR/WCL is more accurate than WCL, irrespective of the distance interval. Furthermore, both inner and boundary APs has the similar positioning performance to total APs when the distance interval between colleting points varies.



	dist=1	dist=2	dist=3	dist=4
WCL(All APs, the total = 25)	6.00	6.06	6.28	6.96
GPR+WCL(All APs, the total = 25)	3.96	4.04	4.20	5.55
WCL(inner APs, the total = 8)	2.52	2.92	3.22	2.75
GPR+WCL(inner APs, the total = 8)	3.94	4.20	4.24	5.85
WCL(boundary APs, the total = 17)	7.64	7.53	7.73	8.94
GPR+WCL(boundary APs, the total = 17)	3.97	3.96	4.18	5.41

Figure 4. The position statistics of Wi-Fi APs as a fuction of the distance interval between collecting points

C. WiFi AP position accuracy between GPR/WCL and WCL

From the above test results, we set the best conditional value (the distance interval = 1m, the number of collecting data = 10) and calculate the Wi-Fi AP position using GPR/WCL. And then we finally compared them with Wi-Fi AP positions using WCL at all APs. Here, the position accuracy is averaged from 10 test trials.

In Fig.5, the averaged position accuracy using GPR/WCL is more accurate than using WCL in the various indoor areas, such as the corner, corridor, hall and rooms. Especially, in case of boundary APs which exist in the corner or edge of a floor, the position error when using GPR/WCL decreases from 6.00m to 3.96m (about 34% enhancement) compared with when using WCL. It shows that GPR can compensate for WCL's weakness in the low DoP collecting environments. In case of inner APs which exist in middle of a floor, the position error when using GPR/WCL is almost even or higher than when using WCL. Higher position accuracy of inner APs is due to the incomplete collecting data because the GPR is dependent on the deployment of collecting position or the reliability of collecting RSS data. However, except a few of inner APs, the position error when using GPR/WCL is accurate, under 5m.

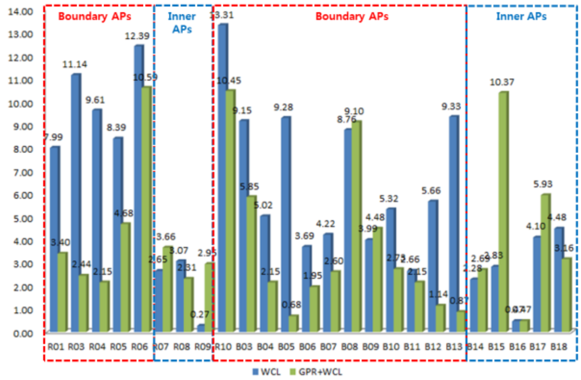


Figure 5. The position accuracy of total Wi-Fi APs in the test area. The total Wi-Fi APs are classified as inner and boundary APs.

V. CONCLUSIONS AND FUTURE WORK

We suggest a novel Wi-Fi AP position estimation algorithm using GPR/WCL. Using GPR, this algorithm is able to estimate the position of Wi-Fi AP which is hard to obtain due to the hide on the ceiling. The effect of the number of collecting data or the distance interval between collecting points on the position accuracy of Wi-Fi AP is analyzed. Furthermore, from repeated experiments in a floor, it is evaluated that its position accuracy using GPR/WCL is totally better than using WCL, in especially low DoP collecting environment, like corner. In the future, we will expand this proposed algorithm to heterogeneous mobile devices and find the effect of the RSS difference between devices on Wi-Fi AP position accuracy. [9]

ACKNOWLEDGMENT

This research was supported by the "Development of 5m accuracy LBS Platform supporting AR service" project funded by the Korea Communications Commission.

REFERENCES

[1] P. Bahl and V. N. Padmanabhan. RADAR: An In-Building RF-based User Location and Tracking System. In Proceedings of the 19th Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM, 2000.

[2] Uzair Ahmad, Choi Wan Sik, Sangjoon Park, Lim Dong-Sun, "Development of COEX WiFi APDB: Methodology, Results and Best Practices", ION Technical Meeting(ITM) 2011, Jan 2011.

[3] SkyhookWireless Inc.,http://www.skyhookwireless.com/, 2010.

[4] J. Ledlie, J. Park, D. Curtis, A. Cavalcante, L. Camara, and R. Vieira, "Mole: a Scalable, User Generated WiFi Positioning Engine", in Indoor Positioning and Indoor Navigation, 2011, pp. 1-10.

[5] J.Wang, P.Urriza, Y.Han, D.Cabric, Weighted centroid algorithm for estimating primary user location: Theoretical analysis and distributed implementation. *IEEE Transactions on Wireless* Communications, june 2011.

[6] Niculescu and B. Nath, "Ad hoc positioning system (APS)," in Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM), 2001, pp. 2926–2931.

[7] Gaussian Processes for Machine Learning. Carl Edward Rasmussen and Christopher K. I. Williams The MIT Press, 2006.

[8] http://gaussianprocess.org/gpml/

[9] Y. Cho, M. Ji, Y. Lee and S. Park, "WiFi AP position estimation using contribution from heterogeneous mobile devices", IEEE PLANS, 2012.