On the Accuracy of Wi-Fi Localization using Robot and Human Collected Signatures

Luyao Niu, Yingyue Fan, Kaveh Pahlavan, *Fellow, IEEE,*
Guanxiong Liu, *Student Member, IEEE,*, Yishuang Geng, *Student Member, IEEE*
Center for Wireless Information Network Studies, Worcester Polytechnic Institute, Worcester, MA 01609 USA
Email: {lniu, yfan, kaveh, gliu2, and ygeng}@wpi.edu

***Abstract*—As one of the most popular ways for Wi-Fi based indoor localization, the received signal strength (RSS) can provide us localization results with accuracy proportional to the quality of Wi-Fi database collection. In this paper, we collected two Wi-Fi databases using human and robot respectively. The experiments are conducted on the 3rd floor of Atwater Kent laboratory. Based on the two databases, we use an offline process to generate Wi-Fi RSS maps which can be used to estimate any arbitrary position. Furthermore, a comparison of the performance of different localization algorithms based on different databases are provided and analyzed.**

***Index Terms*—Wi-Fi, indoor localization, robot, database collection**

I. Introduction

With the guide of location aware applications, navigation in indoor environment is possible [4]. The performance of navigation highly depends on the accuracy of the localization. This dependency indicates that higher accuracy localization work is one of the best promises to higher profits of the applications [8]. However it is obvious that the easiest improvement of the accuracy of localization can be gained directly by more meticulous war driving which needs the companies to pay more for human labor. Even though the step by step RSS database collection is able to be completed, the expenses on this particular one war driving will go extremely high if such hard working is considered and admired. Finally, the catastrophe to such RSS database collection is that the indoor environment that users to localize inside are almost everywhere all over the world. Such scenario obviously violated the basic survival rule of all of the companies.

We propose to use simultaneous localization and mapping (SLAM) [6] based on robot to replace human resources in the war driving. However, robot can not be that accurate as human because of the limitations of sensor accuracy which indicates that error in robot is not avoidable. In this paper, we are contributing efforts to analyze such error. Based on the prior knowledge of the error coming from the robot sensors, the performance of different localization algorithms such as kernel method or nearest neighbor using database collected by robot can be analyzed. Meanwhile, similar performance analysis can be done based on the database collected manually. At end, we will make a conclusion on the possibility of the replacement of human with robot in RSS database collection.

The rest of the paper is organized as follows. Section II illustrates the methodologies of our experiments. It gives out a brief introduction of the error from the odometry sensor. Then we give out the process of the collection of Wi-Fi databases, one collected by human while another by robot. After the collections of the databases, a so called offline process is implemented based on the data from the databases. Finally two algorithms including K-nearest neighbor(K-NN) and kernel method are implemented on both databases to localize a mobile device with RSS reading in it. Section III showed part of our experiment results. The error we calculated from different database with different algorithm are listed in table I. Finally, a brief conclusion and possible future work are claimed in section IV.

II. Methodologies

A. Simultaneous Localization and Mapping (SLAM)

Dealing with both of localization and mapping is a dilemma until simultaneous localization and mapping (SLAM) was proposed. Basically iterations including a prediction and a observation process are implemented in SLAM solution. The prediction process is

$$x_v(k+1) = F_v(k)x_v(k) + u_v(k+1) + w_v(k+1) \qquad (1)$$

where $x_v(k) = [x_k, y_k, \theta_k]^T$ is the position state of the robot, $F_v(k)$ is the state transition matrix, $u_v(k)$ is the vector of control inputs and $w_v(k)$ is the vector of uncorrelated process noise errors with zero mean and covariance $Q_v(k)$.

In equation (1), the state vector $x(k)$ contains not only the information from the state of the robot but also the location information from all landmarks. Thus $x(k)$ can be expanded as $x(k) = [x_v^T(k)\ p_1^T\ ...\ p_n^T]^T, n \in [1, N]$ where N is the total number of landmarks. For the observation model, the process can be written as

$$z(k) = H_i x(k) + v_i(k) \qquad (2)$$

where $v_i(k)$ is a vector of uncorrelated observation errors with zero mean and variance $R_i(k)$ and H_i is the observation matrix that relates the sensor output $z_i(k)$ to the state vector $x(k)$ when observing the i^{th} landmark.

B. Sensor Evaluation

To replace human with robot in RSS collection, it is necessary to be clear how inaccurate the robot is. The coordinate

 978-1-4673-8364-6/16/$31.00 ©2016 IEEE

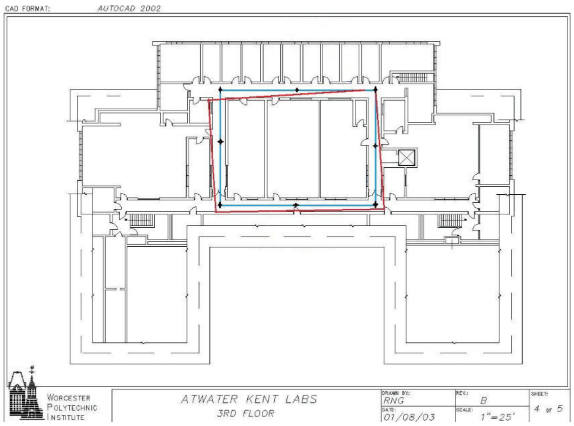
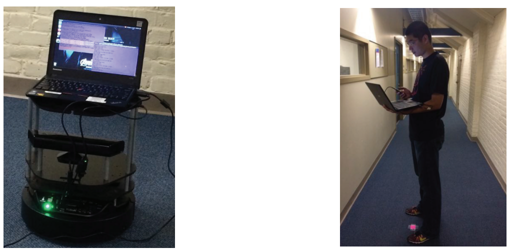


Fig. 1. 2D floor layout for Atwater Kent building, WPI.



(a) Measurement by robot (b) Measurement by human

Fig. 2. Experimental Setup: Wi-Fi signal strenght reading collection

information we got from the robot is the information from odometry sensor which accumulate the error. We place the robot on the 3rd floor of Atwater Kent (AK) building whose map is shown in Fig. 1 in Worcester Polytechnic Institute (WPI), and experimented on its error. The robot is drifting as it goes along, but the error in the whole travel is acceptable. The error introduced by the drift is included in the following discussions.

C. Database Collection

To evaluate whether robot based RSS collection provides accurate database to support localization, two databases are collected. The first one is collected by human with cellphone, and the second one is collected by robot. In case of the shadow fading brought by people walking around in the building, we collected these two databases with multiple samples to compensate the human shadowing effect. A mobile device is used to record the Wi-Fi signal data. As for robot based RSS collection the robot carried the device walking through the whole building with SLAM [9], meanwhile all of the RSS information is recorded in the mobile device associated with its coordinate information using timestamp. Since indoor environment is very complicated, there might be interference coming not only from multipath but also from RF source. Therefore, we selected out 12 MAC addresses which are the ones that can be observed at all of the positions to promise the quality of further localization. We draw the estimation path and the ground truth approximately on Fig. 1. The 8 black stars in the map are the positions that have the highest probability to be selected as RSS manual measure positions in such an environment.

D. Offline Process

Traditionally, the blind spots in the database are treated as -99 dBm which is much lower than the actual reading, or even the minimum value of the observed RSS [7]. Instead of treating them as -99dBm, our approach is to use Gaussian process [2] [1] to generate the RSS Map for each candidate AP. We build 11 modes based on the different reference points, and for each mode, we select proper access points to do Gaussian process to calculate the mean for each reference points.

Gaussian Process (GP) is a stochastic process where every variable has a Gaussian distribution and the variables have a jointly Gaussian distribution. Here these variables are the RSS for each reference location. The data is as

$$D = (x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n) \tag{3}$$

where x is the 2-Dimentional location information, y is the received signal strength indicating for one Wi-Fi access point. We assume

$$y_i = f(x_i) + \epsilon \tag{4}$$

In equation (4) y_i is related to x_i by some function f with noise in it as $\epsilon \sim N(0, \sigma_n^2)$. We assume that the covariance of the function values at different points is correlated and that the covariance of $f(x_i)$ and $f(x_j)$ is defined by x_i and x_j with a kernel function. We choose the squared exponential kernel function which is shown as following:

$$cov(f(x_i), f(x_j)) = K(x_i, x_j) = \sigma_f^2 exp(-\frac{|x_i - x_j|^2}{2l^2}) \tag{5}$$

where σ_f^2 is the signal variance, and l is the length scale for the strength of correction between positions. Take the noise into consideration, we have the covariance over the whole observations Y: $cov(Y) = K + \sigma_n^2 I$ where K is the $n \times n$ covariance matrix of X, $K_{ij} = K(x_i, x_j)$, I is $n \times n$ identity matrix, and σ_n^2 is the variance of observation error ϵ. the jointly Gaussian distribution is

$$Y \sim N(0, K + \sigma_n^2 I) \tag{6}$$

Up to now, we have three parameters $\theta = \sigma_f, \sigma_n, l$. Based on the hypothesis of these parameters, the probability density function of Y in condition of X and θ is

$$P(Y|X,\theta) = \frac{1}{\sqrt{(2\pi)^n |cov(Y)|}} exp(-\frac{1}{2}(Y)^T cov(Y)^{-1} Y) \tag{7}$$

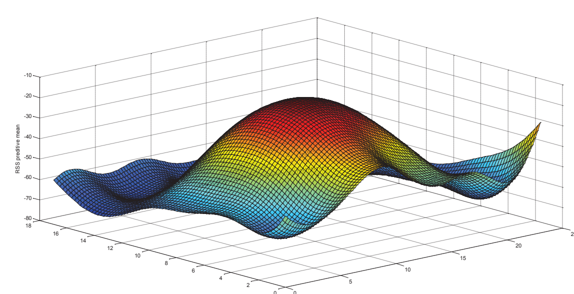


Fig. 3. relationship between RSSI and position (RSSI map)

 978-1-4673-8364-6/16/$31.00 ©2016 IEEE

θ can be estimated by maximizing the probability equation (7). With the parameters, we can generate a map constructed by the signal strength readings for each access point and the location information. With such an RSSI map, if we further hope to find out the function f_* for any arbitrary position x_*, some new datum should be interpolated into our training dataset. Therefore, the jointly Gaussian distribution can be further given as

$$\begin{pmatrix} Y \\ f_* \end{pmatrix} \sim \left(0, \begin{bmatrix} cov(Y) & K_* \\ K_*^T & K_{**} \end{bmatrix} \right) \tag{8}$$

The posterior predictive distribution is

$$p(f_*|x_*, X, Y) = p(f_*|u_*, y_*) \tag{9}$$

where $u_* = K_*^T cov(Y)^{-1} Y$, $v_* = K(x_*, x_*) - K_*^T cov(Y)^{-1} K_*$

All of the notations with star are referring to the updated value. Fig. 3 is a $23m \times 17m$ RSSI map for one access point with 48 reference points. A similar plot on the uncertainty variance versus position can also be obtained which is not shown because of page limitation. Intuitively the closer the testing point is to the training point, the lower uncertainty it has. Now in the following Wi-Fi localization algorithm implementations, all of the blind spots should be filled using Gaussian process prediction.

E. Algorithms for Wi-Fi Localization

After the databases are available for localization, some algorithms can be implemented based on the data. In our experiment, we compared Gaussian kernel method and K nearest neighbor which are two most widely used ones. [5]

1) Guassian Kernel Method: Gaussian kernel method is a statistical approach. [10]

Kernel method firstly defines a mass probability distribution function based on RSS database. The function is given in equation (10).

$$K(\mathbf{O}, \mathbf{O}_{nm}) = \frac{1}{(\sqrt{2\pi}\sigma)^K} e^{-\frac{1}{2\sigma^2}\sum_{m=1}^{M}(p_m - p_{nmk})^2} \tag{10}$$

Assume the difference between the measurement at the same position follows Gaussian random distribution, then the joint probability density function of observation and measurement can be calculated in equation (11)

$$p(\mathbf{O}|l_n) = \frac{1}{M}\sum_{k=1}^{K} K(\mathbf{O}, \mathbf{O}_{nm}) \tag{11}$$

Further, from Bayes rule, we know that $p(l_n|\mathbf{O}) = p(\mathbf{O}|l_n) \cdots p(l_n)\frac{1}{p(\mathbf{O})} = \eta \cdots p(\mathbf{O}|l_n)$

Thus the estimation of the location can be calculated as the expectation of the coordinates based on the RSS information. To eliminate the impact brought by the selection of σ, an algorithm called Gaussian Process (GP) is introduced [3].

2) K Nearest Neighbor (K-NN): Nearest neighbor is a more intuitive approach to localize the mobile device. If the power reading at an unknown position is available, the distance metric for each reference points (RPs) can be formed as

$$d_n = \sqrt{\sum_{m=1}^{M}(p_n - p_{n,m})^2}, n = 1, 2, ...N \tag{12}$$

According to the distance in the vector above, the location of RP that has the minimum distance is intuitively believed as localization result. This algorithm is obviously not precise enough, thus K nearest neighbor which can average the result of K nearest RPs is introduced to improve its performance. We propose to use weighted average whose weights of belief are calculated as $w = \arctan\frac{1}{d_n}$.

III. Experiment Results

In table I, the localization average error of kernel method and K nearest neighbor (K-NN) based on different database are listed with unit of meters. A similar table representing the variance of error can also be created which is not listed because of page limitation. The number of reference points which are uniformly distributed along the four edges increases from 4 to 58 with step size 4. Also the positions that we hope to localize are uniformly distributed along the edges as well. When we are implementing K-NN, we experimented the algorithm with K varying from 1 to 4so that we can seek a value of K that can provide us better performance. In the following parts, we will firstly separate the tables into two parts based on database, then we will analyze along both horizontal and vertical directions. The horizontal comparison is basically the performance comparison of algorithms, and the vertical observation is the analysis of the relationship between the performance and the number of RPs.

A. performance of human database

For the left part of the table I, only the first 4 rows are available. The NA terms in the table coming from the limitations of labor based war driving.

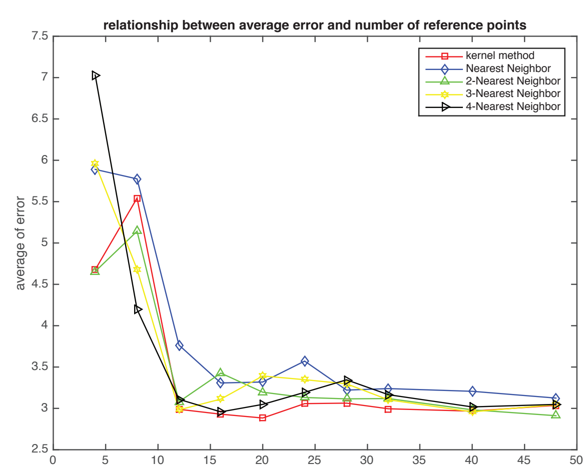


Fig. 4. relationship between average error and number of RP

978-1-4673-8364-6/16/$31.00 ©2016 IEEE

TABLE I
AVERAGE ERROR COMPARISON WITH DIFFERENT ALGORITHMS BASED ON DIFFERENT DATABASE

number of RP	database collected by human					database collected by robot				
	kernel	NN	2-NN	3-NN	4-NN	kernel	NN	2-NN	3-NN	4-NN
4 corner	2.49	4.85	2.90	4.66	6.99	4.06	6.11	3.23	4.13	5.49
4 center	5.36	5.40	5.84	7.51	9.10	4.67	5.89	4.65	5.96	7.03
8RP	4.25	5.70	3.47	5.15	6.40	5.54	5.77	5.14	4.68	4.19
12RP	2.80	3.30	3.30	3.13	3.39	2.98	3.76	3.07	2.98	3.10
16RP	NA	NA	NA	NA	NA	2.92	3.30	3.42	3.11	2.95
20RP	NA	NA	NA	NA	NA	2.88	3.31	3.19	3.39	3.04
24RP	NA	NA	NA	NA	NA	3.05	3.57	3.13	3.34	3.19
28RP	NA	NA	NA	NA	NA	3.06	3.22	3.11	3.29	3.34
32RP	NA	NA	NA	NA	NA	2.99	3.23	3.11	3.10	3.16
40RP	NA	NA	NA	NA	NA	2.96	3.20	2.98	2.95	3.01
48RP	NA	NA	NA	NA	NA	3.03	3.12	2.91	3.04	3.04

1) horizontal analysis: Focus on the first 4 rows on the left part of the table, we can observe that based on the mean of error generally the best performance can be achieved using 3-NN. 4-NN can not provide ideal performance because the number of reference points is too small. This also indicates that if there are not enough reference points, K should not be selected too large. Additionally, kernel method can provide medium performance. Generally speaking, 3-NN can provide best performance, 2-NN can perform closely to 3-NN. Kernel method is worse than 3-NN and 2-NN, but it is better than 4-NN. We can roughly conclude that 3-NN is more suitable for human database based localization because there are few number of reference points.

2) vertical analysis: If analyze along the vertical direction, we can observe that for a specific algorithm, the performance is increasing with the increment of the number of reference points. We have already known that 3-NN can perform best from the analysis before, but we can notice that lowest localization average error is 3.8 meters and the standard variance is 2.5.

B. performance of robot database

1) horizontal analysis: For nearest neighbor, we can observe that when the number of reference points is smaller than 10, 2-NN and 3-NN can provide us lower error. The reason that 4-NN have larger error is the limitation of the number of RPs which has already been illustrated before. When the number of RPs falls in the interval between 10 and 30, we can see that the cluster of NN algorithms perform closely while kernel method is better because of our offline process. When the number keeps increasing, 2-NN can provide us the lowest error while the others are close to each other. Thus we can roughly conclude that if we compare the performance of the algorithms, K-NN is more suitable if we have enough RPs because it has smallest error and best stability.

2) vertical analysis: Intuitively, the performance will increase with the increment of the number of reference points. At first, when the reference points are distributed sparsely, the error will decrease greatly when the number of reference points increases which can be seen in Fig. 4 before the number of RPs getting closer to 10. After that, the decrease of error can still be observed but it is not that obvious any more. Finally, we can see that the best performance provided by the robot collected database is 2.8 meters error and 1.5 standard variance which are much smaller than those provided by human collected database.

IV. Conclusion

In this paper, the performances of localizations based on RSS databases collected by human and robot are compared. Even though sensor error will be introduced into localization accuracy based on robot database collection, the overall performance of localization is good to indicate the replacement of human war driving with robot is reasonable.

Acknowledgements

Thanks to the support from robotic department in Worcester Polytechnic Institute (WPI). Also, great thanks to the MQP students including James Castro, Umair Rehman and Biao Zheng.

References

[1] M Aravecchia and S Messelodi. Gaussian process for rss-based localisation. In *Wireless and Mobile Computing, Networking and Communications (WiMob), 2014 IEEE 10th International Conference on*, pages 654–659. IEEE, 2014.

[2] Felix Duvallet and Ashley D Tews. Wifi position estimation in industrial environments using gaussian processes.

[3] Brian Ferris, Dieter Fox, and Neil D Lawrence. Wifi-slam using gaussian process latent variable models. In *IJCAI*, volume 7, pages 2480–2485, 2007.

[4] K Pahlavan, F Akgul, Y Ye, T Morgan, F Alizadeh-Shabdiz, M Heidari, and C Steger. Taking positioning indoors wi-fi localization and gnss. *Inside GNSS*, 5(3):40–47, 2010.

[5] Kaveh Pahlavan and Prashant Krishnamurthy. *Principles of Wireless Access and Localization*. John Wiley & Sons, 2013.

[6] Randall C Smith and Peter Cheeseman. On the representation and estimation of spatial uncertainty. *The international journal of Robotics Research*, 5(4):56–68, 1986.

[7] Jungmin So, Joo-Yub Lee, Cheal-Hwan Yoon, and Hyunjae Park. An improved location estimation method for wifi fingerprint-based indoor localization. *International Journal of Software Engineering and Its Applications*, 7(3):77–86, 2013.

[8] Xinchao Song and Yishuang Geng. Distributed community detection optimization algorithm for complex networks. *Journal of Networks*, 9(10):2758–2765, 2014.

[9] Brian Yamauchi. A frontier-based approach for autonomous exploration. In *Computational Intelligence in Robotics and Automation, 1997. CIRA'97., Proceedings., 1997 IEEE International Symposium on*, pages 146–151. IEEE, 1997.

[10] Xin Zheng, Guanqun Bao, Ruijun Fu, and Kaveh Pahlavan. The performance of simulated annealing algorithms for wi-fi localization using google indoor map. In *Vehicular Technology Conference (VTC Fall), 2012 IEEE*, pages 1–5. IEEE, 2012.

978-1-4673-8364-6/16/$31.00 ©2016 IEEE