DOI:
10.1109/LSP.2016.2519607

An improved K-Nearest-Neighbor indoor localization method based on Spearman distance

Yaqin Xie, Yan Wang, *Member, IEEE*, Arumugam Nallanathan, *Senior Member, IEEE*, and Lina Wang

***Abstract*—Indoor localization based on existing Wi-Fi Received Signal Strength Indicator (RSSI) is attractive since it can reuse the existing Wi-Fi infrastructure. However, it suffers from dramatic performance degradation due to multipath signal attenuation and environmental changes. To improve the localization accuracy under the above-mentioned circumstances, an improved Spearman-distance-based K-Nearest-Neighbor (KNN) scheme is proposed. Simulation results demonstrate that our improved method outperforms the original KNN method under the indoor environment with severe multipath fading and temporal dynamics.**

I. Introduction

The proliferation of wireless communication and mobile computing has driven the demand of location-based services (LBSs). For outdoor open environment, the Global Navigation Satellite Systems (GNSS), such as Global Positioning System (GPS), GLONASS, BeiDou Navigation System and Galileo Positioning system can provide high location accuracy. However, the GNSS signals from satellites cannot be always hearable in many indoor areas, which limits their applications. In recent decade, Wi-Fi infrastructures are widely deployed in many indoor environments such as airports, supermarkets and shopping malls etc. Furthermore, most of the current off-the-shelf smart equipments, such as smartphones, laptops and Ipad are integrated with Wi-Fi modules which makes it possible for location based on Wi-Fi signal strengths in indoor environment.

However, performing received signal strength (RSS) based indoor localization is particularly challenging due to the following three reasons: Firstly, it is difficult to obtain an accurate received signal strength indicator (RSSI) values. According to the measurement in [1], the variance of RSSIs collected from an immobile receiver in one minute is up to 5dB. Secondly, RSSI is easily varied by the multipath and NLOS effect which is unavoidable in an indoor environment where a ceiling, a floor, furniture, walls and the movement of people are present. Thirdly, manufacturing variations among different Wi-Fi devices might also affect the RSSI measurement accuracy. Therefore, different Wi-Fi devices might yield different RSSI values at the same location, especially for those equipments made by different manufacturers. However, as RSSI fingerprints can be easily obtained from most off-the-shelf wireless network infrastructures, it is also an attractive approach for location determinations.

The current RSS-based localization methods can be divided into ranged-based localization methods and RSS fingerprint location techniques. The former converts the RSSI values to distances according to the propagation loss model[2–3] before performing localization by lateration methods. The latter can usually be divided into two phases: training and locating. Although the procedure of training is time-consuming, labor-intensive, and vulnerable to environmental dynamics, it is inevitable for fingerprinting-based approaches. With the development of Google maps, indoor google maps can provide more and more indoor localization. Until July 2015, over 10,000 locations around the world are available.

In fact, the absolute RSSI values are unstable and quite different when measured by different Wi-Fi terminals. To mitigate these effects, the relative RSSI values, i.e. their rankings, are used to replace the absolute ones for location determinations[4–9]. This means that, even if the absolute RSSI values of a set of Access Points (APs) in the covering area might be quite different when measured by different Wi-Fi measurement devices or over different time, their ranking is more likely to remain the same or, at least, more similar. This is based on the assumption that the RSSI values monotonically decrease when the distance between the source and APs increases[10]. To evaluate the similarity among different rankings of the same set of APs, the Spearman rank correlation coefficient[11] is utilized which is a nonparametric measure of statistical dependence between two variables. This Spearman's coefficient assesses how well the relationship between two variables can be described using a monotonic function and is appropriate for both continuous and discrete variables, including ordinal variables. Based on this reason, we propose to utilize the Spearman rank correlation coefficient to evaluate the similarity among different rankings of the same set of APs. For the purpose of comparison, RSS-based lateration algorithm proposed in [12] and the K-nearest neighbor (KNN) method[13] are also provided in this letter.

To evaluate the effectiveness of our spearman-distance-based method, we use a partition attenuation factor propagation model as in [12] to simulate the real indoor environments.

The rest of the letter is organized as follows. In Section II, we first provide some related background before introducing our proposed method. Simulation results and discussion are presented in Section III. Finally, we conclude our work in

Yaqin Xie and Lina Wang are with the Jiangsu Key Laboratory of Meteorological Observation and Signal Processing, Nanjing University of Information Science & Technology, Nanjing 210044, China(e-mail: xyq@nuist.edu.cn; wangllna@163.com).

Yan Wang is with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China(e-mail: yanwang@seu.edu.cn).

Arumugam Nallanathan is with the Department of Informatics at King's College London, UK(e-mail: arumugam.nallanathan@kcl.ac.uk).

Section IV.

II. Algorithmic description

A. Background

There are two phases in RSSI fingerprint location approaches: the off line training phase and the runtime localization procedure. During the off line training phase, various RSSI values are collected at predefined points within the coverage area. The collected RSS characteristics are location dependent and stored in the correlation databases (CDBs). Obviously, the more parameters per signal observed, the more unique the fingerprint and thus the better the location accuracy.

An RSSI fingerprint can also be classified as either a target or reference fingerprint. Undoubtedly, a target RSSI fingerprint is associated with the object node that is to be localized, that is, it contains signal parameters measured by the object node or by the associated APs. The reference RSSI fingerprints are values collected during the training phase and stored in the CDB. The target fingerprint $\mathbf{T}$ used in the remainder of this letter is written by a $N_t \times 2$ matrix:

$$\mathbf{T} = \begin{bmatrix} ID_1 & RSSI_1 \\ \vdots & \vdots \\ ID_{N_t} & RSSI_{N_t} \end{bmatrix}, \tag{1}$$

where N_t is the number of APs within the range of the object to be localized. ID_i and $RSSI_i$ are the identity and the measured RSSI from the i^{th} AP, respectively.

The reference fingerprint $\mathbf{R}$ at pixel (i,j) can be expressed as

$$\mathbf{R_{i,j}} = \begin{bmatrix} ID_{i,j,1} & RSSI_{i,j,1} \\ \vdots & \vdots \\ ID_{i,j,N_{i,j}} & RSSI_{i,j,N_{i,j}} \end{bmatrix}, \tag{2}$$

where $N_{i,j}$ is the number of APs whose predicted RSSI values are above the minimum threshold at pixel (i,j). The rows of $\mathbf{R_{i,j}}$ are classified in descending order of RSSI, that is, $RSSI_{i,j,k} > RSSI_{i,j,k'}$, if $k <= k'$.

B. Spearman Rank Correlation Coefficient

The Spearman rank correlation coefficient[14] is used to calculate the correlation between the target fingerprint $\mathbf{T}$ and the reference fingerprint $\mathbf{R_{i,j}}$. However, the target fingerprints might not have the same number of APs nor the same APs. Therefore, some modification is needed before computing the Spearman correlation coefficient.

On this basis, two $N_t \times 2$ matrices, V_T and V_R, are generated, which are initialized to be

$$\mathbf{V_T} = \mathbf{V_R} = \begin{bmatrix} ID_1 & N_t \\ \vdots & \vdots \\ ID_{N_t} & N_t \end{bmatrix} \tag{3}$$

The position of APs in the RSSI ranking of $\mathbf{T}$ must be inserted in the second column of the correspondent row in $\mathbf{V_T}$, that is to say

$$\mathbf{V_T}(n_k, 2) = k, \tag{4}$$

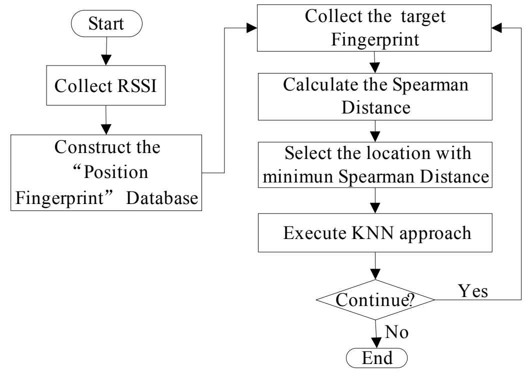

Fig. 1. Flow chart of localization.

where $\mathbf{V_T}(n_k, 1) = \mathbf{T}(k, 1)$, $n_k \in [1, N_t]$ and $k = 1, 2, \ldots, N_t$.

Similarly, $\mathbf{V_R}$ can be renewed to be

$$\mathbf{V_R}(n_k, 2) = k, \tag{5}$$

where $\mathbf{V_R}(n_k, 1) = \mathbf{R}_{i,j}(k, 1)$, $n_k \in [1, N_t]$ and $k = 1, 2, \ldots, N_{i,j}$.

The Spearman rank correlation coefficient between the target fingerprint and the reference fingerprint at pixel (i,j) can be expressed as

$$\rho_{i,j} = \frac{\sum_{n=1}^{N_t} \left[\left(V_T(n,2) - \bar{R}_T \right) \left(V_R(n,2) - \bar{R}_R \right) \right]}{\sqrt{\sum_{n=1}^{N_t} \left[\left(V_T(n,2) - \bar{R}_T \right)^2 \left(V_R(n,2) - \bar{R}_R \right)^2 \right]}}, \tag{6}$$

where

$$\bar{R}_T = \frac{1}{N_t} \sum_{n=1}^{N_t} V_T(n,2),$$

$$\bar{R}_R = \frac{1}{N_t} \sum_{n=1}^{N_t} V_R(n,2),$$

Subsequently, the Spearman distance $d_{i,j}$ can be given as

$$d_{i,j} = 1 - \rho_{i,j}. \tag{7}$$

C. Spearman-distance-based methods

By exploiting the Spearman rank correlation of RSSI measurements from different APs, we proposed an improved Spearman-distance-based KNN[15] location method. In our proposed approach, the localization procedure includes the following three steps: Firstly, the offline RSSI fingerprint database is built; Secondly, collecting the position fingerprint of the object; Thirdly, calculating the Spearman distance according to Equation (7) and then select all the locations with minimum Spearman Distance; Finally, execute the original KNN approach according to all the locations with minimum Spearman Distance and obtain the final location estimation. The detailed flow chart is shown as in Fig. 1.

III. Simulations and Discussions

A. Simulation Methodology

In our simulation, RSSI values are generated from the predefined 400 locations which are shown in Fig. 2 by the small blue circles. This area is a 10m by 10m square field with all these predefined points evenly spaced in this field. The position of the object to be localized is randomly generated inside this coverage area. That is to say, the x and y label of the objects to be localized are Nx $\times$ rand and Ny $\times$ rand, respectively, where Nx and Ny represent the maximize value in axis x and y of the simulated area which are both equal to 10m in our simulation and "rand" is the Matlab function. Four APs at the four corners of the square area are deployed as shown in Fig. 2. In order to simulate the indoor environments, the whole coverage area is divided into nine districts which represent nine different rooms, which are separated by concrete walls. The wireless signals from devices in different rooms are decayed by various number of walls before reaching the APs. To simulate the indoor signal propagation, a partition attenuation factor (PAF)[2] is added into the shadow fading propagation model. Therefore, the shadow fading propagation model can be written as[6]

$$P(d) = P(d_0) - 10\gamma\log_{10}\left(\frac{d}{d_0}\right) - W \times PAF + X_\sigma, \quad (8)$$

where $P(d)$ is the total path loss measured in decibel, $P(d_0)$ is the path loss at the reference distance d_0, γ is the path loss exponent, PAF is used to denote a specific obstruction such as walls in indoors. We use it here to simulate the penetration loss when signals pass through the wall. W is the number of walls between the object node and the APs. X_σ is a normal random variable. For a better understanding, consider the wireless device located in Room4 of Fig. 2, W is 1, 3, 3, 1, respectively, for AP_1, AP_2, AP_3 and AP_4.

Moveover, in order to simulate the indoor environment accurately, it is important to take the correlation into account. The spatial correlation of the shadow fading effect is obtained as follows:

Firstly, a $m \times m$ covariance matrix $\mathbf{K}$ is generated which satisfies $K_{ij}(d_{ij}) = \sigma^2 \exp\left(-\frac{d_{ij}}{D_c}\right)$, where D_c represents the decorrelation distance which can range from several meters to many tens of meters, d_{ij} is the distance between the i^{th} position and the j^{th} one. Secondly, for the above obtained covariance matrix $\mathbf{K}$, execute cholesky factorization, we get $\mathbf{K} = \mathbf{L}\mathbf{L}^\mathbf{T}$. Subsequently, generate a non-correlated normal random variables $\omega = [\omega_\mathbf{1}, \ldots, \omega_\mathbf{m}]^\mathbf{T}$, then X_σ in Equation (8) can be expressed as $X_\sigma = \mathbf{L}\omega$. That is to say, the correlation of the shadow fading at location i and j is $E[X_\sigma(i) X_\sigma(j)] = K_{ij}(d_{ij}) = \sigma^2 \exp\left(\frac{-d_{ij}}{D_c}\right)$.

In this subsequent simulation, Non-line-of-sight (NLOS) environment is assumed, and then the related parameter sets are summarized as shown in Table I.

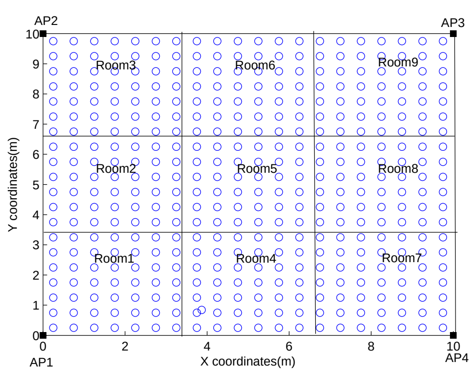

Fig. 2. Simulation layout.

TABLE I. Default Parameter Sets.

Parameters	d_0	$P(d_0)$	γ	**PAF**	D_c
Values	1**m**	-37.3dBm	3.3	5**dB**	10**m**

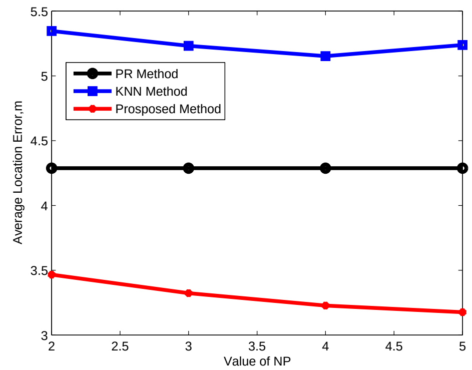

Fig. 3. Comparison of average location errors for different values of NP.

B. Impact of the number of the closest neighbor points

The effect of varying number of NP which means the number of the closest neighbor points in the location space can be examined by choosing different values from 2 to 5. The average location errors (ALE) of the PR method, the KNN method and the Proposed method are shown in Fig. 3. It is obvious that the value of ALE for the PR method is a constant because it is nothing to do with the value of NP. As shown in Fig. 3, the curve corresponds to the KNN method shows a steady decline in ALE before rising and reaches the lowest point when NP equals to 4. For the proposed method, however, there is a slow and steady decrease in ALE, ranging from 3.5m to just below 3.2 m when NP increases from 2 to 5. Thus, for fair comparison, we keep the value of NP at 4 in the following subsections.

C. Impact of shadow fading

To see the impact of shadow fading factor, we set σ range from 4dB to 8dB and repeat the same measurements for 1000 times. Figure 4 illustrates the ALE when the shadow fading factor varies within the above given scope. For the purpose

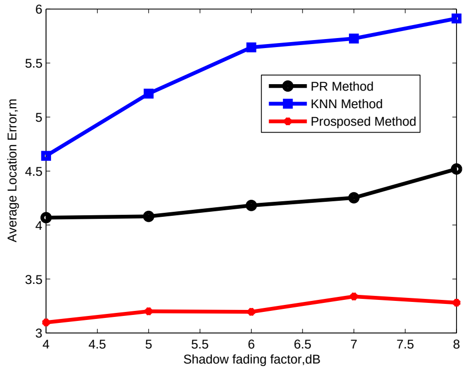

Fig. 4. Comparison of average location errors for different location methods.

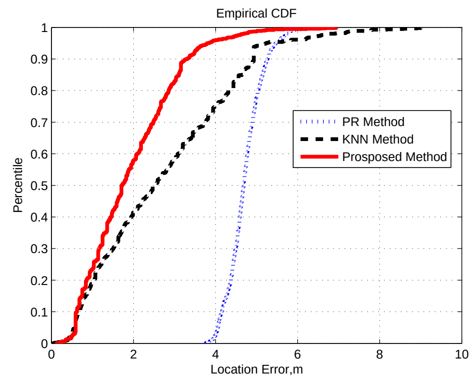

Fig. 5. Location error CDFs with $\sigma = 5$dB.

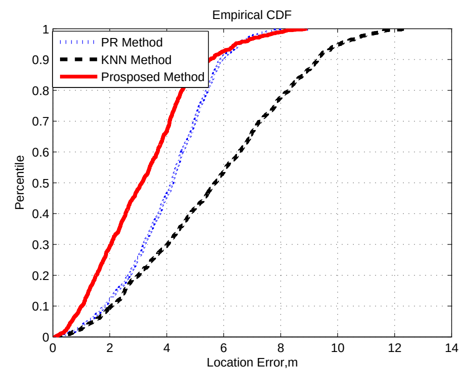

Fig. 6. Location error CDFs with $\sigma = 7$dB.

of comparison, the original KNN method and the polynomial-regression-based method (PR Method) [2] are also simulated. Obviously, the positioning error significantly increases for the original KNN method with the increase of shadow fading factor. Such impressive results are natural since that a big shadow fading forms more unreliable position fingerprint and thus the influence of location error is increased to a certain extent. The value of shadow fading, by contrast, doesn't have a significant influence on the another two methods. When σ is small, the original KNN method and the PR method show similar localization performance which are apparently inferior to the proposed method. The effectiveness of our proposed method improves significantly as the value of σ increases which demonstrates the superiority of our method in aweful indoor environments comparing with the another two approaches.

D. Cumulative Error Distribution

Fig. 5–6 illustrate the cumulative distribution function(CDF) of localization errors in the simulated indoor environment when the shadow fading factor equals to 5dB and 7dB, respectively. As can be seen from Fig. 5, the proposed scheme achieves a localization error under 2.7m for 80% of the testing samples, which is significantly smaller than those of the original KNN method(4.5m) and the PR method(4.9m). And when the shadow fading factor increases from 5dB to 7dB, the proposed scheme achieves a localization error under 4.6m for 80% of the testing samples, which is significantly smaller than those of the original KNN method(8.2m) and the PR method(5.8m).

On this basis, we can conclude that our approach exhibits a preferable property in the indoor environment where the value of RSSI is unstable and varied with time since the proposed method takes the relative ranking of RSSIs into consideration which is beneficial for improving the precision of location fingerprint.

IV. Conclusions

A novel Spearman-distance-based indoor location system is presented in this letter, which is based on the fingerprint of RSSI values obtained in advance from the APs. We collect and proceed the RSSI values as ”fingerprints” to form the radio map in the training procedure. The spearman rank correlation coefficient is then calculated after obtaining the unknown position fingerprint. Then we get the spearman distance based on the spearman rank correlation coefficient and then combine it with the original KNN approach. Experimental results show that the proposed combination method achieves better performance than the another two existing methods.

V. Acknowledgement

This work was supported in part by the Open Research Fund of National Mobile Communications Research Laboratory, Southeast University (No. 2012D20), the National Science and Technology Major Project of China under Grant 2016ZX03001022-002, the National Natural Science Foundation of China under Grant Number NSFC-61401216. It is also funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions. Partial work was finished when the first author visited King's College London, UK and supported by China Scholarship Council (Beijing, China) (CSC No: 201408320039).

REFERENCES

[1] K. Wu, J. Xiao, Y. Yi and D. Chen, "CSI-Based Indoor Localization," *IEEE Transactions on Parallel and Distributed System*, vol. 24, no. 7, pp. 1300-1309, July 2013.

[2] P. Bahl and V.N. Padmanabhan, "Radar: An In-building RF-Based User Location and Tracking System," *Proceedings of the IEEE INFOCOM*, 2000.

[3] Y. Xu, J. Zhou and P. Zhang, "RSS-based source localization when path-loss model parameters are unknown," *IEEE Communications Letters*, vol. 18, no. 6, pp. 1055-1058, April 2014.

[4] J. Machaj, P. Brida, and R. Piche, "Rank based fingerprinting algorithm for indoor positioning," *in 2011 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, IEEE, 2011, pp. 1-6.

[5] K. Yedavalli, B. Krishnamachari, "Sequence-based localization in wireless sensor networks," *IEEE Trans. Mob. Comput*, Vol. 7, Issue 1, pp. 81C94, January 2008.

[6] Deora, S. and Krishnamachari, B., "Harnessing Non-Uniform Transmit Power Levels for Improved Sequence Based Localization," *2014 IEEE International Conference on Distributed Computing in Sensor Systems (DCOSS)*, Marina Del Rey, CA, May 2014.

[7] Liu Z., Chen J., "A new sequence-based iterative localization in wireless sensor networks," *In: Proceedings of the 2nd International Conference on Information Engineering and Computer Science*, Wuhan, China, 2009.

[8] Xu X., Deng Z., "A novel sequence based localization approach for wireless sensor networks," *In: Proceedings of the 5th International Conference on Wireless Communications, Networking and Mobile Computing*, Beijing, China, 2009.

[9] Yu Y., Jiang C., Zhao X. etc., "Sequence-based localization algorithm with improved correlation metric and dynamic centroid," *Science China Information Sciences*, Vol. 54, Issue 11, pp. 2349-2358, Nov. 2011.

[10] J. Krumm, G. Cermak, and E. Horvitz, "Rightspot: a novel sense of location for a smart personal object," *in Proceedings of Ubicomp 2003*, Seattle, USA, pp. 36–43, October 2003.

[11] The MathWorks, Inc., "Pairwise distance between pairs of objects," http://www.mathworks.de/access/helpdesk/help/toolbox/stats/pdist.html, May 2010.

[12] J. Yang, Y. Chen, "Indoor Localization Using Improved RSS-Based Lateration Methods," *Proceedings of the 28th IEEE Conference on Global Telecommunications*, *GLOBECOM'09*, Piscataway, NJ, USA, Novermber 30–December 4, 2009.

[13] F. Yu, M. Jiang, J. Liang, et al., "5G WiFi Signal-Based Indoor Localization System Using Cluster k-Nearest Neighbor Algorithm," *International Journal of Distributed Sensor Networks*, vol. 2014, Article ID 247525, 12 pages, 2014.

[14] S. A. Zekavat and R. M. Buehrer, Handbook of Position Location: Theory, Practice and Advances, John Wiley and Sons, 2011.

[15] P. Torteeka and X. Chundi, "Indoor positioning based on Wi-Fi fingerprint technique using fuzzy K-nearest neighbor," *in Proceedings of the 11th International Bhurban Conference on Applied Sciences and Technology(IBCAST14)*, Islamabad, January 2014.