ON THE PERFORMANCE ANALYSIS OF WIFI BASED LOCALIZATION

Baoqi Huang, Minmin Liu, Zhendong Xu and Bing Jia

College of Computer Science, Inner Mongolia University, Hohhot 010021, China
Email: jiabing@imu.edu.cn

ABSTRACT

Currently, WiFi (or IEEE 802.11) enabled infrastructures and devices have become ubiquitous, which makes it promising to provide indoor positioning services via WiFi signals. However, most existing studies on its performance are carried out based on simulations and experiments, and it is still challenging to mathematically characterize the localization error. Therefore, in this paper, the Cramer-Rao lower bound (CRLB) for the localization error by using WiFi signals is established and enables us to carry out a theoretical analysis on the fundamentals of WiFi based localization. Extensive experiments based on the well-known WiFi fingerprint-based localization are then carried out and confirm the correctness of the proposed CRLB model as well as the analysis.

Index Terms— WiFi, CRLB, Localization, Performance analysis

1. INTRODUCTION

Since location information plays a vital role in both indoor and outdoor applications, great efforts have been devoted to developing various localization techniques [1, 2]. Since GPS cannot be applied in indoor environments [3], it is promising to develop indoor positioning services based on widely available WiFi infrastructures. Thus, various indoor positioning and navigation techniques have been presented [4, 5].

Due to its simplicity and tolerance to pervasive multipath effects in indoor environments, the WiFi fingerprint-based method [6–8] has gained most attention in both academia and industries. Basically, the fingerprint-based method involves two steps. In the first step, a radio map consisting of a number of fingerprints labelled with respective reference points (i.e. predefine locations) within a service space is constructed via an offline site survey [9, 10]. In the second step, when a device sends an online location query containing its current received signal strength (RSS) from multiple APs, its location can be inferred according to the existing radio map [11, 12].

Thanks to the National Natural Science Foundation of China (Grants No. 61461037, 41761086 and 61761035), the Natural Science Foundation of Inner Mongolia Autonomous Region of China (Grant No. 2017JQ09), and the "Grassland Elite" Project of the Inner Mongolia Autonomous Region (Grant No. CYYC5016).

However, due to the complexity in the indoor propagation of WiFi signals, it is really challenging to characterize the performance of the WiFi based localization techniques. Most studies on the localization performance are based on experiments and simulations. In [13], a preliminary analytical model was developed for a localization system instance with simplified assumptions on the signal propagation and system design. In [14], a comparative study on indoor localization was reported with experiments, and the results are highly dependent on the environment and device implementation. In [15], a general but complicated probabilistic model was presented to assist in the performance analysis and further helps to design an optimal fingerprint reporting strategy. However, all the above studies cannot directly answer how the WiFi based localization error is affected by various factors.

In this paper, we deal with the performance of a general WiFi based localization problem and formulate the Cramer-Rao lower bound (CRLB) to characterize the resulting localization error. It is shown that the CRLB is inversely proportional to the number of RSS measurements from WiFi access points (APs). Moreover, the RSS measurements from different APs contribute diversely to localization according to their corresponding RSS gradients around the true location. Then, extensive experiments are conducted and a performance analysis confirms the correctness of the CRLB model.

2. FUNDAMENTALS OF WIFI BASED LOCALIZATION

Throughout this paper, we shall use the following mathematical notations: $(\cdot)^T$ denotes transpose of a matrix or a vector; $\mathrm{Tr}(\cdot)$ denotes the trace of a square matrix.

2.1. A Generic Localization Model

As was commonly assumed in the literature [15], the RSS measurements of the signals propagated from n APs to a receiver at a position $\mathbf{x}$, denoted $\mathbf{y} = [y_1, y_2, \cdots, y_n]^T$, are independent and identically distributed, namely

$$\mathbf{y} \sim N(\mathbf{m}(\mathbf{x}), \sigma^2 \mathbf{I}_n), \tag{1}$$

where $\mathbf{m}(\mathbf{x}) = [m_1(\mathbf{x}), m_2(\mathbf{x}), \cdots, m_n(\mathbf{x})]^T$ is a vector function containing the functions of mean RSS measurements at the position $\mathbf{x} = [x_1, x_2]^T$ from the n APs.

The localization problem aims to infer the unknown position $\mathbf{x}$ given the corresponding vector of RSS measurements, i.e. $\mathbf{y}$. Suppose that the RSS measurement model in (1) is known, and the localization problem can be solved through, e.g. the maximum likelihood estimator (MLE).

Define gradients $\mathbf{r}_i = \frac{\partial m_i(\mathbf{x})}{\mathbf{x}} = [r_{i1}, r_{i2}]$ and formulate $\mathbf{r} = [\mathbf{r}_1^T, \mathbf{r}_2^T, \cdots, \mathbf{r}_n^T]^T$. The likelihood function is

$$L(\mathbf{y};\mathbf{x}) = \log p(\mathbf{y}|\mathbf{x}), \tag{2}$$

and the Fisher information matrix (FIM) is

$$\mathbf{F}(\mathbf{x}) = -\operatorname{E}\left(\frac{\partial^2 L(\mathbf{y};\mathbf{x})}{\partial \mathbf{x} \partial \mathbf{x}^T}\right) = \frac{1}{\sigma^2}\mathbf{r}^T\mathbf{r}. \tag{3}$$

The inverse of the FIM, namely the Cramer-Rao lower bound (CRLB), is formulated as [16–18]

$$\begin{aligned} \mathbf{F}^{-1}(\mathbf{x}) &= \sigma^2 \left[\sum_{i=1}^{n} \begin{pmatrix} r_{i1}^2 & r_{i1}r_{i2} \\ r_{i1}r_{i2} & r_{i2}^2 \end{pmatrix}\right]^{-1} \\ &= \frac{\sigma^2 \sum_{i=1}^{n} \begin{pmatrix} r_{i2}^2 & -r_{i1}r_{i2} \\ -r_{i1}r_{i2} & r_{i1}^2 \end{pmatrix}}{\sum_{i=1}^{n} r_{i1}^2 \sum_{i=1}^{n} r_{i2}^2 - (\sum_{i=1}^{n} r_{i1}r_{i2})^2} \end{aligned}$$

The mean squared error is defined as follows

$$\begin{aligned} \operatorname{Tr}(\mathbf{F}^{-1}(\mathbf{x})) &= \frac{\sigma^2 \sum_{i=1}^{n}(r_{i1}^2 + r_{i2}^2)}{\sum_{i=1}^{n} r_{i1}^2 \sum_{i=1}^{n} r_{i2}^2 - (\sum_{i=1}^{n} r_{i1}r_{i2})^2} \\ &= \frac{2\sigma^2 \sum_{i=1}^{n} \|\mathbf{r}_i\|^2}{\sum_{i=1}^{n}\sum_{j=1}^{n}(\|\mathbf{r}_i\|\|\mathbf{r}_j\|\sin\theta_{ij})^2}, \\ &= \frac{2\sigma^2}{\sum_{i=1}^{n}\left(\frac{\|\mathbf{r}_i\|^2}{\sum_{k=1}^{n}\|\mathbf{r}_k\|^2}\sum_{j=1}^{n} h_{ij}^2\right)} \\ &= \frac{2\sigma^2}{n\sum_{i=1}^{n}\left(\frac{\|\mathbf{r}_i\|^2}{\sum_{k=1}^{n}\|\mathbf{r}_k\|^2}\bar{h}_i^2\right)} \\ &= \frac{2\sigma^2}{n\bar{h}_w^2}, \end{aligned} \tag{4}$$

where θ_{ij} denotes the angle subtended by $\mathbf{r}_i$ and $\mathbf{r}_j$

$$h_{ij} = \|\mathbf{r}_j\| \sin\theta_{ij}, \tag{5}$$

$$\bar{h}_i = \sqrt{\frac{1}{n}\sum_{j=1}^{n} h_{ij}^2}, \tag{6}$$

$$\bar{h}_w = \sqrt{\sum_{i=1}^{n}\left(\frac{\|\mathbf{r}_i\|^2}{\sum_{k=1}^{n}\|\mathbf{r}_k\|^2}\bar{h}_i^2\right)}. \tag{7}$$

In Fig. 1, the gradients $\mathbf{r}_i$ and $\mathbf{r}_j$ are illustrated by vectors, and h_{ij} denotes the distance from one vertex of the vector $\mathbf{r}_i$ to another vector $\mathbf{r}_j$. As such, $\bar{h}_i^2$ with $i = 1, \cdots, n$ actually represents the average squared distance associated with the gradient $\mathbf{r}_i$ and all the other gradients $\mathbf{r}_j$ with $j = 1, \cdots, n$. Likewise, $\bar{h}_w^2$ represents the weighted average squared distance $\bar{h}_i$ among the gradients $\mathbf{r}_i$ with $i = 1, \cdots, n$.

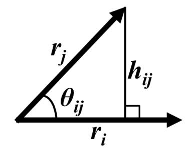


Fig. 1. The illustration of h_{ij} associated with the gradients $\mathbf{r}_i$ and $\mathbf{r}_j$.

2.2. Performance Analysis

According to (4), it can be straightforwardly concluded that: (a) the localization error scales with the magnitude of the noises in the RSS measurements, namely σ^2; (b) the localization error is inversely proportional to the number of RSS measurements, namely n; (c) the localization error inversely proportional to h_w^2, which is a complicated variable depending on all the gradients $\mathbf{r}_i$ with $i = 1, \cdots, n$.

In light of the formulation of $\bar{h}_w^2$, its value is mainly determined by two factors: the magnitudes of the gradients $\mathbf{r}_i$ with $i = 1, \cdots, n$ and the angles subtended by each pair of the corresponding vectors. Therefore, an analysis shall be presented based on these two factors.

Firstly, the magnitude of the gradient $\mathbf{r}_i$ reflects the spatial variation of (mean) RSS measurements when signals are emitted from the corresponding AP, implying that: (a) in general, the larger are the variations, the larger is the weighted average squared distance $\bar{h}_w^2$, the less is the CRLB, namely that good performance can be obtained for those positions with large gradients, and vice versa; (b) if the magnitude associated with one AP is trivial, the contribution of this AP to localization can be neglected on account of the resulting trivial weight in calculating $\bar{h}_w^2$, such that the number of AP-s used in localization as well as the size of the corresponding radio map can be significantly reduced by removing those with trivial weights; (c) specifically, in the most commonly adopted WiFi fingerprint-based approach, an offline site survey must be conducted to build a radio map (or a fingerprint database), which factually provides rough information about the gradients so as to evaluate (4); as a result, the efficiency of fingerprint-based localization systems can be improved to some extent (this will be elaborated in the next section).

Secondly, since $\sin\theta_{ij}$ is employed in (4), it can be concluded that: (a) when the gradients are collinear, the denominator of (4) will equal to zero, such that the CRLB does not exist, meaning that the localization performance is severely degraded in this case; (b) on the contrary, the CRLB can be minimized provided that two APs are used and their gradients are orthogonal to each other.

Thirdly, (5), (6) and (7) reveal that $\bar{h}_i$ is averaged over n and $\bar{h}_w$ is the weighted average value of $\bar{h}_i$ with $i = 1, 2, \cdots, n$, implying that $\bar{h}_w$ is nearly independent of n especially when n is sufficiently large. Hence, supposing that $\bar{h}_w$ keeps constant when n is above a threshold, say N_{th}, it follows from (4) that continuously increasing n above N_{th} can reduce the CRLB by only less than $100/(N_{th}+1)$ percents, indicating that having more than APs N_{th} hardly contributes to localization accuracy.

In summary, considering the difficulties in predicting indoor propagations of WiFi signals due to multipath effects, it is hard or even impossible to exactly determine the gradients of mean RSS measurements, so that optimizing the WiFi based localization performance by using (4), though is attractive, turns out to be challenging.

3. EXPERIMENTS

3.1. Setup

In the experiments, practical RSS measurements are firstly collected in our lab, which is Room 316 in the Building of College of Computer Science. Specifically, the 65 points on the 13×5 lattice with the side length of 1 m are chosen as reference positions and another 15 random positions are chosen as part of the testing positions; at each reference or testing position, a smartphone (Redmi Note) is placed on a 1.5 m high tripod to scan WiFi APs in 5 minutes. Consequently, at each reference position, 200 RSS sample vectors are random selected and their average vector is used as the fingerprint; 54 testing positions, including 39 reference positions and 15 random testing positions, are used for testing purposes, and 100 RSS sample vectors are selected as testing samples. The smartphone at each position might detect a different number of WiFi APs, ranging between 31 to 64, and thus the element in a RSS sample vector will be set to be -100 dBm if the corresponding AP cannot be detected.

To validate (4), we need to obtain the gradients at the testing positions. Hence, the gradient at one position is approximated by using its nearest five neighboring positions and the multiple linear regression function $regress$ in Matlab.

The WiFi fingerprint-based localization adopts the nearest neighbour (NN) method as well as the k nearest neighbours (kNN) with different choices of k to determine the final location estimate, which is realized in Matlab.

3.2. Performance Comparison

Firstly, the relationship between the localization error and the number of APs is considered for validation. To do so, different dimensions of the fingerprints are emulated by virtually removing certain APs, and accordingly, the elements corresponding to those removed APs will be deleted from both fingerprints and testing RSS sample vectors. Specifically, the dimension is varied from 3 to 24 with the step size of 3, and given a specific number of APs, 50 different combinations of APs are generated for localization.

The CRLB and final localization errors obtained by NN and kNN with $k = 3, 6$ and averaged over the 50 combinations are depicted in Fig. 2. As can be seen, with increasing the number of APs, both the CRLB and the practical localization errors are decreasing, but the decreasing rates become slow, which is consistent with (4).

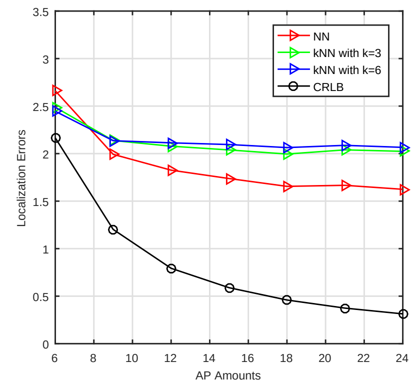
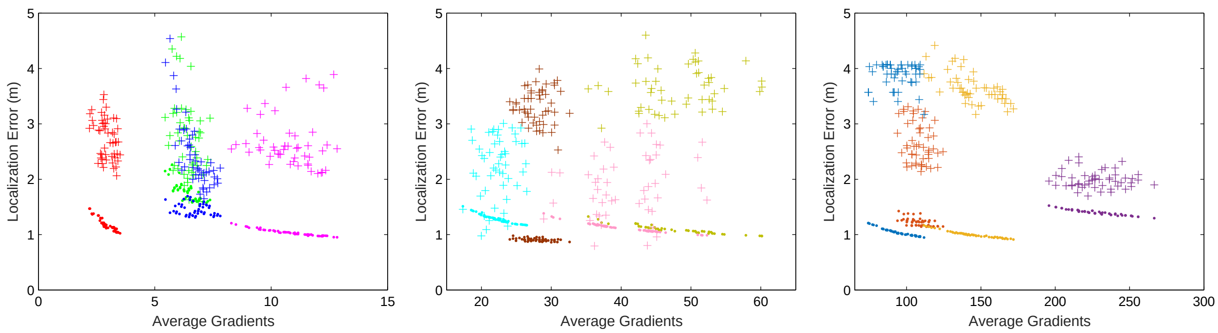


Fig. 2. The localization errors on different number of APs under the CRLB and the fingerprint-based approach.

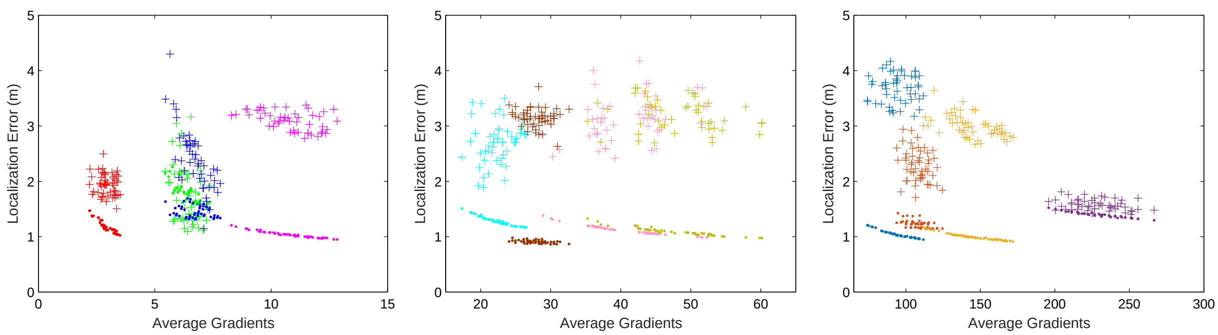
Secondly, the relationship between the localization error and the RSS gradient is considered for validation. By letting the number of APs used for localization be 9, we randomly choose 50 combinations of APs again. At each testing position, one testing RSS sample vector is randomly selected and its average RSS gradient of the corresponding 9 RSS gradients from 9 APs is calculated for use. In Fig. 3, the localization error and CRLB against its corresponding average RSS gradient is plotted with respect to NN and kNN with $k = 3, 6$. For clear of illustration, each subfigure in Fig. 3 only depicts the results of 4 testing positions. As can be seen, in most cases, the CRLB associated with any one testing position displays an obvious descending trend, whereas the localization error in most cases appears to decrease with the corresponding average RSS gradient increasing. This observation confirms that there does exist the relationship between the localization performance and RSS gradients, which can be utilized in practical localization, say to select optimal AP for localization.

4. CONCLUSIONS

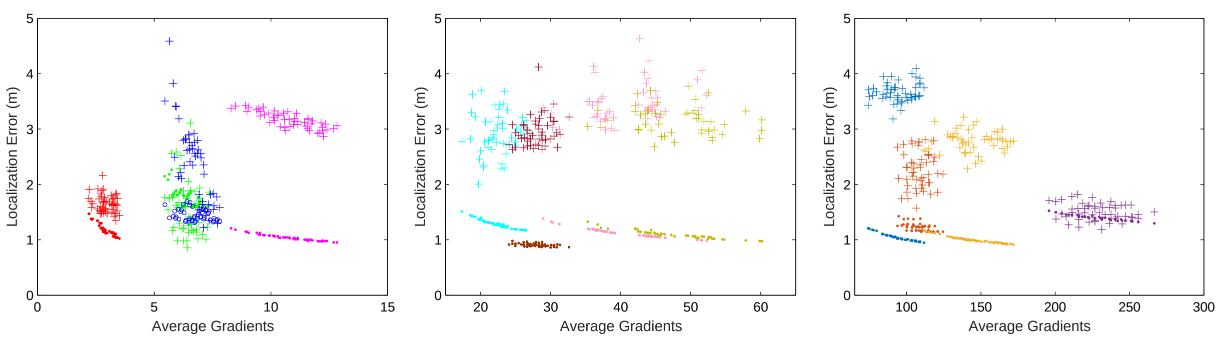
This paper presented a theoretical analysis on the performance of WiFi based localization through the CRLB which is widely adopted in the literature. The CRLB revealed that the localization performance is dependent on the number of APs and the RSS gradients. Then, the experiments were conducted based on WiFi fingerprint-based localization in a real environment and validated the theoretical analysis. This study not only provides valuable insight into the fundamentals of WiFi based localization, but also provides useful design guidelines for practical applications.



(a) The CRLB vs. the localization error using NN.



(b) The CRLB vs. the localization error using kNN with $k = 3$.



(c) The CRLB vs. the localization error using kNN with $k = 6$.

Fig. 3. The impact of RSS gradients on the localization error under. The dot denotes the CRLB and the plus denotes the localization error.

5. REFERENCES

[1] B. Huang, L. Xie, and Z. Yang, "TDOA-based source localization with distance-dependent noises," *IEEE Transactions on Wireless Communications*, vol. 14, no. 1, pp. 468–480, 2015.

[2] L. Zhao, N. Gao, B. Huang, Q. Wang, and J. Zhou, "A novel terrain-aided navigation algorithm combined with the tercom algorithm and particle filter," *IEEE Sensors Journal*, vol. 15, no. 2, pp. 1124–1131, Feb 2015.

[3] B. Buchli, F. Sutton, and J. Beutel, *GPS-equipped wireless sensor network node for high-accuracy positioning applications*. Springer, 2012, pp. 179–195.

[4] L. Zhao, Z. Liu, T. Li, B. Huang, and L. Xie, "Moving target positioning based on a distributed camera network," *Mathematical Problems in Engineering,2014,(2014-5-13)*, vol. 2014, no. 2, pp. 1–11, 2014.

[5] B. Huang, G. Qi, X. Yang, L. Zhao, and H. Zou, "Exploiting cyclic features of walking for pedestrian dead reckoning with unconstrained smartphones," in *Ubicomp '16*. ACM, 2016, pp. 374–385.

[6] P. Bahl and V. N. Padmanabhan, "Radar: An in-building rf-based user location and tracking system," in *Infocom '00*, vol. 2. IEEE, pp. 775–784.

[7] H. Zou, B. Huang, X. Lu, H. Jiang, and L. Xie, "A robust indoor positioning system based on the procrustes analysis and weighted extreme learning machine," *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1252–1266, 2016.

[8] M. Zhou, Y. Tang, W. Nie, L. Xie, and X. Yang, "Grassma: Graph-based semi-supervised manifold alignment for indoor wlan localization," *IEEE Sensors Journal*, vol. 17, no. 21, pp. 7086–7095, Nov 2017.

[9] H. Zou, B. Huang, X. Lu, H. Jiang, and L. Xie, "Standardizing location fingerprints across heterogeneous mobile devices for indoor localization," in *2016 IEEE Wireless Communications and Networking Conference*, April 2016, pp. 1–6.

[10] H. Zhao, B. Huang, and B. Jia, "Applying kriging interpolation for wifi fingerprinting based indoor positioning systems," in *WCNC '16*. IEEE, 2016, pp. 1–6.

[11] H. Zou, Y. Zhou, H. Jiang, B. Huang, L. Xie, and C. Spanos, "Adaptive localization in dynamic indoor environments by transfer kernel learning," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, March 2017, pp. 1–6.

[12] M. Zhou, Y. Tang, Z. Tian, and X. Geng, "Semi-supervised learning for indoor hybrid fingerprint database calibration with low effort," *IEEE Access*, vol. 5, pp. 4388–4400, 2017.

[13] K. Kaemarungsi and P. Krishnamurthy, "Modeling of indoor positioning systems based on location fingerprinting," *Proceedings - IEEE INFOCOM*, vol. 2, no. 2, pp. 1012–1022 vol.2, 2004.

[14] E. Elnahrawy, X. Li, and R. P. Martin, "The limits of localization using signal strength: a comparative study," in *Sensor and Ad Hoc Communications and Networks, 2004. IEEE SECON 2004. 2004 First IEEE Communications Society Conference on*, 2004, pp. 406–414.

[15] Y. Wen, X. Tian, X. Wang, and S. Lu, "Fundamental limits of rss fingerprinting based indoor localization," in *IEEE Conference on Computer Communications*, 2015, pp. 2479–2487.

[16] Y. T. Chan and K. C. Ho, "A simple and efficient estimator for hyperbolic location," *IEEE Transactions on Signal Processing*, vol. 42, no. 8, pp. 1905–1915, Aug 1994.

[17] B. Huang, T. Li, B. D. O. Anderson, and C. Yu, "Performance limits in sensor localization ," *Automatica*, vol. 49, no. 2, pp. 503–509, 2013.

[18] B. Huang, C. Yu, and B. D. O. Anderson, "Understanding error propagation in multihop sensor network localization," *IEEE Transactions on Industrial Electronics*, vol. 60, no. 12, pp. 5811–5819, Dec 2013.