A Novel Approach for RSS-based Fingerprinting Location System in GSM

Zhongliang Deng, Enwen Hu*, Lu Yin
Beijing University of Posts and Telecommunications, Road No.10, Xitucheng, Haidian District
*Corresponding author, e-mail: owen.hu@bupt.edu.cn

Abstract
Received signal strength (RSS)-based fingerprinting location systems have attracted much attention in recent years. No additional hardware is required for its implementation to existing networks. However, by analyzing the realistic measurements from GSM, we find that not only the RSS but also the neighboring base stations heard at some reference point will change over time. But the relative order of neighboring base stations (NBS) sorted according to their RSS is more stable. In this paper, we present a confidence coefficient of base station based on a gridding approach to get a relative accurate NBS during the offline phase, and proposed a novel algorithm with Longest Common Subsequence Matching Degree (LCSMD) of base station vector (BSV), which uses the relative order of NBS sorted according to their RSS. To evaluate our proposed method, we collect the realistic data from GSM network. Results show that the method based on LCSMD improves the positioning accuracy compared to the original method.

Keywords: GSM, received signal strength, LCSMD, confidence coefficient

1. Introduction
Location schemes have gained considerable attention in the last several years, as position estimations are required in many practical applications. Currently, the Global Positioning System (GPS) [1] is the most effective positioning technology in the outdoor open environments [2]. Some methods use the time or angle of signal arrival to estimate the mobile user’s position [3-6]. However, many of these method requires additional hardware, and line-of-sight propagation between transmitter and receiver [7]. While, location fingerprinting method can remedy to these problems well. A fingerprinting method is used for some of these reasons: firstly, no additional hardware is required for it implementation to existing networks. Secondly, compared to the methods above, it performs better in areas with significant multipath propagation [8, 9]. It uses a database to determine user locations. Generally, it has two phases including offline phase and online phase. During the offline phase, database contains the records (RSS, BS, GPS location) at reference points. In the online stage, when network side receives a positioning request coming from user equipment (UE), the position of UE can be acquired by matching available UE measurements with the corresponding data of every separated grid.

Recently, several fingerprint matching algorithms have been used for user location. In [10], a fingerprint positioning method is proposed, which selects k fingerprints with the smallest Euclidean distance as potential indicators of the current location. The location of the device is estimated as an average of the latitude and longitude coordinates of the best k matches. A matching algorithm based on maximum likelihood in [11] is proposed to suit the situation where the measurement values are relatively independent, and measurement errors satisfy the Gaussian distribution.

The high sensitivity of the RSS to the environmental changes is taken into account in the previous work [12]. But they didn’t discuss the changes of NBS, which will influence the accuracy of the positioning system. To remedy this issue, a confidence coefficient of BS has been proposed in this paper. The serving BS is stored into database, as well as up to six other BSs which has the highest confidence coefficient. LCSMD is proposed to improve the accuracy and reduce the computational complexity for the positioning system.

In this paper, we proposed a novel approach based on LCSMD of BS vector. Our approach works in three steps: first, we collect the real-time RSS samples at reference points
which used to construct the offline database. Second, considering the stochastic RSS and BS, we use a gridding approach to get the analysis of statistical characteristics of them. Third, we apply the LCSMD of BS vector to get the coarse range of user’s position, and then use the K-nearest-neighbor-based method to find the most likely location.

This paper is organized as follows. The localization method is presented in Section II. In section III we present the performance evaluation of our proposed method and compare with other algorithms. The paper is concluded in section IV.

2. LCSMD Approach
In this section, we describe our location system based on LCSMD for GSM network. We start by an overview of the system followed by the details of offline and online positioning phases. LCSMD-based location system contains two phases: an offline fingerprint database construction phase and online positioning phase. During the offline phase, we collect the RSS, the serve base station and the other 6 neighboring base stations to construct the offline fingerprinting database. During the online phase, we calculate the LCSMD of the base station vector between the point to be located and offline fingerprinting database. According to the LCSMD, a coarse range of the point to be located is achieved from the offline fingerprinting database. Following by K-Nearest Neighbor (KNN) algorithm, the position of the point to be located is finally estimated.

2.1. Collecting Database
During the offline phase, the time samples of RSS readings and base station vectors are collected at known locations, referred to as the Reference Points (RP). The raw set of fingerprinting time samples collected in the real GSM network at RP \( i \) is denoted as \( F_i = \{coordinate^i_{vector}, RSS^i_{vector}, BS^i_{vector}\} \), with the representing the coordinate of RP, and denoting the received signal strength (RSS) vector, and the hearable base station vector at RP is defined as . So we can define the offline database as:

\[
F_{DB} = \begin{pmatrix}
coordinate^1_{vector} & RSS^1_{vector} & BS^1_{vector} \\
coordinate^2_{vector} & RSS^2_{vector} & BS^2_{vector} \\
\vdots & \vdots & \vdots \\
coordinate^n_{vector} & RSS^n_{vector} & BS^n_{vector}
\end{pmatrix}
\]

(1)

Where \( n \) denotes the total number of samples collected.

2.2. Database Preprocessing
As we all know, radio propagation environment is easily influenced by occasional obstructions and the complicate multipath, so the RSS and the neighboring base stations we collected are always different from others collected in the same location at different time. This requires us to stand at each location for a certain time to collect enough samples, then averaging to keep those samples be relatively accurate. But this will increase the fingerprint database construction overhead significantly, as the war-driving car should stop at each position for a certain time. In order to construct a relatively accurate database and avoid this overhead, we use a gridding approach presented in [13]. But it only considered the sensitivity of the RSS, and do not take the changes of neighboring BS vector into account. By observing the measured data in the GSM network, we find that the neighboring base stations received at referent points will change over time. In this paper, we propose a method based on the confidence coefficient of BS to remedy to the issue about stochastic neighboring base stations.

All the coordinate of RPs collected in the offline phase can be written as:

\[
coordinate = \begin{pmatrix}
lon_1 & lat_1 \\
lon_2 & lat_2 \\
\vdots & \vdots \\
lon_n & lat_n
\end{pmatrix}
\]

(2)
Where \( \text{lon}_i, \text{lat}_i \) representing the longitude and latitude of RP \( i \) respectively. Converting to \( x-y \) coordinate system, the coordinate of RPs can be given by:

\[
\text{coordinate} = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_n & y_n \end{bmatrix}
\] (3)

Then we set the grid size as \( w \) (meter), the range of \( i \)th grid can be expressed as follows:

\[
\begin{cases}
    x, y | x_{\min} + (i - 1) \times w \leq x \leq x_{\min} + i \times w \\
    y_{\min} + (i - 1) \times w \leq y \leq y_{\min} + i \times w
\end{cases}
\] (4)

Where \( x_{\min}, y_{\min} \) respectively denote the min of \( x \) and \( y \) in the database. The center of \( i \)th grid can be given by

\[
\text{coordinate}_{G,i} = \left[ x_{\min} + \left( i - \frac{1}{2} \times w \right), y_{\min} + \left( i - \frac{1}{2} \times w \right) \right]
\] (5)

It denotes the position of \( i \)th grid after gridding process in database. Next, we need to get the RSS and BS vectors to represent the all the samples collected in the \( i \)th grid. Let \( m \) denotes the number of the samples collected in the \( i \)th grid, and all the samples are denoted \( G_i = (F_1, F_2, \ldots, F_m) \). Let \( m \) represents the number of base stations heard in the \( i \)th grid, and which can be given by:

\[
BS_i = (BS_1, BS_2, \ldots, BS_p)
\] (6)

Then we can calculate the confidence coefficient of each base station as follows:

\[
\eta_j = \frac{1}{w} \frac{\text{Num}_j}{\sum_{q=1}^{p} \frac{1}{w} \text{Num}_q}
\] (6)

\[
W_j = \frac{\text{RSS}_j^i}{\sum_{q=1}^{p} \text{RSS}_q^i}
\] (7)

Where \( \eta_j \) denotes the confidence coefficient of base station \( j \), \( \text{Num}_j \) is the number of occurrences of base station \( j \) in the \( i \)th grid, and \( w_j \) is the weight coefficient of RSS from base station \( j \). According to \( \eta \), the base station vector of the \( i \)th grid can be given by:

\[
BS_{G,i} = (BS_1^i, BS_2^i, \ldots, BS_p^i)
\] (9)

The RSS vector can be written as:

\[
\text{RSS}_{G,i} = (\text{RSS}_1^i, \text{RSS}_2^i, \ldots, \text{RSS}_p^i)
\] (10)

\[
\text{RSS}_t^i = \frac{\sum_{j=1}^{\text{Num}_t} \text{RSS}_{ij}}{\text{Num}_t}(1 \leq t \leq 7)
\] (11)

Where \( \text{RSS}_t^i \) and \( \text{RSS}_{ij} \) denote the RSS from base station and in Eq. (6) at \( j \)th time respectively. So the fingerprinting data of the \( i \)th grid can be given by:

\[
F_{G,i} = \{\text{coordinate}_{G,i}, BS_{G,i}, \text{RSS}_{G,i}\}
\] (8)

After the gridding process, the offline database can be written as:
F_{G,DB} = \begin{pmatrix}
\text{coordinate}_{G,1} & BS_{G,1} & RSS_{G,1} \\
\text{coordinate}_{G,2} & BS_{G,2} & RSS_{G,2} \\
\vdots & \vdots & \vdots \\
\text{coordinate}_{G,c} & BS_{G,c} & RSS_{G,c}
\end{pmatrix}
\tag{9}

Where \( c \) denotes the number of grid.

2.3. Coarse Location

During the online phase, the user received data vectors including BS vector and RSS vector, which can be written as \( F_u = (BS_u, RSS_u) \). We divided this phase into two steps including coarse location and fine location. In this step, we use the BS vector user received to get a coarse range of the user’s position and propose a method based on Longest Common Subsequence Match Degree (LCSMD) of BSV. This method not only consider the number of the same BSs between \( BS_u \) user received and \( BS_G \) in offline base, but also take the order of RSS into account.

Finding longest common subsequence (LCS) involves comparison of two or more sequences and find the Longest Subsequence which is common to all sequences. For example, a sequence \( A_1 = \{a, c, g, h, t, e, f\} \) and another one is \( A_2 = \{a, c, e, g, t, f\} \), then \( LCS(A_1, A_2) = \{a, c, g, t, f\} \), we can use solution expressed in [14] as follows:

\[
LCS(i,j) = \begin{cases} 
0 & \text{if } i \text{ or } j = 0, \\
LCS(i-1,j-1) + 1 & \text{if } A_1(i) = A_2(j), \\
\max(LCS(i-1,j), LCS(i,j-1)) & \text{if } A_1(i) \neq A_2(j).
\end{cases}
\] \tag{10}

Firstly, BSs are sorted according to their RSS. Then, we can calculate the LCS between \( BS_u \) user received and the BS vectors in the offline database. Let \( L_i \) denotes the length of LCS between \( BS_u \) and the \( i^{th} \) BS vector \( BS_{G,i} \) in the offline database. The LCSMD can be written as:

\[
\mu_i = \frac{L_i}{N_{total}} \times 100\%
\] \tag{11}

Where \( \mu_i \) is the LCSMD between \( BS_u \) and the \( i^{th} \) BS vector in the offline database, and \( N_{total} \) is the number of base stations that user can hear. Therefore, we can get a subsample library according to the value of LCSMD. The subsample library can be given by:

\[
DB_{sub} = \{F_{G,k} | \mu_k > \mu_T\}
\] \tag{12}

Where \( F_{G,k} \) is the fingerprint data of the \( k^{th} \) grid, \( \mu_k \) is the LCSMD between \( BS_{G,k} \) and \( BS_u \), and \( \mu_T \) is the threshold of LCSMD. If the \( DB_{sub} \) is NULL, the fingerprint data in database has the same serving BS with \( BS_u \) will be chosen as the new coarse range of user location.

2.4. Fine location

In the first step, we have got a coarse range of user’s position as shown in Eq. (16). We only use the user’s BS vector in the first step. Next, we can get the fine location by using user’s RSS vector. In this step, we define a distance function as follows:

\[
D_j(RSS_u, RSS_{sub}) = \sqrt{\sum_{i=1}^{Num_j} (RSS_{u,i} - RSS_{sub,i,j})^2}
\] \tag{13}

where \( D_j(RSS_u, RSS_{sub}) \) denotes the distance between the user’s RSS vector \( RSS_u \) and the \( j^{th} \) RSS vector in the subsample library as shown in Eq. (16), and \( Num_j \) is the number of the same base stations of them. So we can use the method of K-nearest neighbor (KNN) to get the user’s final position as follows:
3. Experiments

In order to evaluate the performance of our proposed method, we collect the realistic data in cellular network environment and compare it with the CellSense-Hybrid system [15] and the Gaussian Process method [16].

3.1. Test Scenario

We collected data in the campus of Beijing University of Posts and Telecommunications which represents an urban area. Data was collected using a Redmi mobile phone which has a GPS receiver (used as ground truth for location) and running the Android 4.0 operating system. We implemented the scanning program using the Android APK. The program records the (timestamp, cell-ID, signal strength, GPS location) for the base station the mobile is connected to as well as the other 6 neighboring cell towers information as dedicated by GSM specifications. The scanning rate was set to one per second. Figure 1 shows the distribution of the base station in the test scenario.

![Figure 1. Effects of selecting different switching under dynamic condition](image)

3.2. Experimental Results

The realistic data was applied to evaluate the performance of our proposed method, compared with original-KNN and ML method. In the offline phase, we use the gridding approach to get the statistics of RSS and BS. The performance of the LCSMD-KNN positioning system and CellSense-Hybrid system with different size of grid is shown in Figure 2.

From Figure 2, we can know that the performance of our proposed method is better than CellSense-Hybrid, when the size of grid is lower than 100 m.

In the step of coarse location, we propose $\mu_T$, a threshold of LCSMD. An appropriate value of $\mu_T$ should be chosen in the LCSMD-KNN system. When $\mu_T$ is chosen from 50%, 60%, 70%, 80%, and 90% to 100%, we observe the mean location errors of them as shown in figure 3, where the size of grid is fixed at 40 m.
A Novel Approach for RSS-based Fingerprinting Location System in GSM (Zhongliang Deng)

From the Figure 3, we can know that the performance seems better when the threshold of LCSMD is chosen from 60% to 70%. Therefore, we choose 65% as the threshold of LCSMD.

Figure 4 is their numerical curves of positioning error for about 2000 test points. The performance of LCSMD-KNN is better than others at most test points as shown in Figure 4.

LCSMD-KNN improves 25 m positioning error at 67% and 97 m at 90% positioning probability, compared with the Gaussian Process method, and provides more than 10 m enhancement, compared with CellSense system, as shown in Table 1.

| method              | Number of data | Number of grid | Error at 67% | Error at 90% |
|---------------------|----------------|----------------|--------------|--------------|
| Gaussian Process    | 10271          | 1549           | 86 m         | 196 m        |
| Cellsense-Hybrid    | 10271          | 1549           | 75 m         | 117 m        |
| LCSMD-KNN           | 10271          | 1549           | 61 m         | 99 m         |
The cumulative distribution function of location error for the different algorithms with the same dataset is shown in Figure 5. It's clearly seen that, for this experiment, our proposed method appeared to give better results than the CellSense-Hybrid system and the Gaussian Process method.

5. Conclusion and Future Works

We proposed a fingerprinting positioning method based on LCSMD of BS for GSM mobile phone. We described the details of the method, and test the performance of it compared to the CellSense-Hybrid system and Gaussian Process positioning method. Our results show that LCSMD-KNN system is better than CellSense-Hybrid with about 10 m and Gaussian Process with 25 m at 67% and 97 m at 90% positioning probability. We also test the effect of the grid size and LCSMD on the location error.

Currently, we are working on improving our system with different method in fine location including probabilistic algorithm and some pattern matching method.

Acknowledgments

This work is supported by the National High Technology Research and Development Program of China under No. 2015AA124101 and National Natural Science Foundation of China (Grant No. 61372110).

References

[1] P Enge, P Misra. Special issue on GPS: the Global Positioning System. Proceedings of the IEEE. 1999: 3-172.
[2] ECL Chan, G Bacieu. Wireless Tracking Analysis in Location Fingerprinting. Proc. IEEE Int'l Conf. Wireless and Mobile Computing. 2008: 214-220. E.C.L. Chan and G. Bacieu: wireless tracking analysis.
[3] S Shah, S Srirangrajan, A Tewfik. Implementation of a directional beacon-based position location algorithm in a signal processing framework. IEEE Trans. Wireless Commun. 2010; 9(3): 1044-1053.
[4] S Al-Jazzar, J Caffery, HR You. Scattering-model-based methods for TOA location in NLOS environments. IEEE Trans. Veh. Technol. 2007; 56: 583-593.
[5] A Küpper, Location-based services: Wiley Online Library. 2005.
[6] H Laitinen, J Lahteenmaki, T Nordstrom. Database correlation method for GSM location. Vehicular Technology Conference. 2001; 4: 2504-2508.
[7] CK Seow, SY Tan. Non-line-of-sight location in multipath environments. IEEE Trans. Mobile Comput. 2008; 7(5): 647-660
[8] Heikki Laitinen, Jaakko Lahteenmaki, Tero Nordstrom. Database Correlation Method for GSM Location. Vehicular Technology Conference. 2001; 4.
[9] Paul Kemppi, Database Correlation Method for Multi-system Location, Master’s Thesis, Department of Electrical and Communications Engineering, Helsinki University of Technology. 2005.
[10] P Bahl, VN Padmanabhan. RADAR: an in-building RFbased user location and tracking system. Proceedings of the 19th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '00). 2000: 2: 775–784.
[11] JB Zhu, XL Luo, D Chen. Maximum likelihood scheme for fingerprinting positioning in LTE system. Proceedings of the IEEE 14th International Conference on Communication Technology (ICCT ’12). IEEE, China, 2012: 428–432.

[12] Takenga CM, Wen Q, Kyamakya K. On the accuracy improvement issues in GSM location fingerprinting. Vehicular Technology Conference. VTC-2006 Fall. IEEE, 2006: 1-5.

[13] Ibrahim M, Youssef M. Cellsense: A probabilistic rssi-based gsm positioning system. Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE. 2010: 1-5.

[14] Rubi R, Arockiam L. Positional_LCS: A position based algorithm to find Longest Common Subsequence (LCS) in Sequence Database (SDB). Computational Intelligence & Computing Research (ICCIC), 2012 IEEE International Conference on. IEEE. 2012: 1-4.

[15] Ibrahim M, Youssef M. CellSense: An accurate energy-efficient GSM positioning system. Vehicular Technology. IEEE Transactions on. 2012; 61(1): 286-296.

[16] Chen MY, Sohn T, Chmelev D, et al. Practical metropolitan-scale positioning for gsm phones. UbiComp 2006: Ubiquitous Computing. Springer Berlin Heidelberg. 2006: 225-242.