Research Article

Optimization of Sampling Cell Size for Fingerprint Positioning

Wei Liu, 1 Yongle Chen, 2 Yongping Xiong, 1 Limin Sun, 3 and Hongsong Zhu 3

1 State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100093, China
2 College of Computer Science and Technology, Taiyuan University of Technology, Taiyuan 030024, China
3 Beijing Key Laboratory of IOT Information Security Technology, IIE, CAS, Beijing 100093, China

Correspondence should be addressed to Yongle Chen; chenyongle@tyut.edu.cn

Received 25 April 2014; Revised 20 June 2014; Accepted 21 June 2014; Published 2 September 2014

Academic Editor: Jianwei Niu

Copyright © 2014 Wei Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The fingerprint positioning based on received signal strength has a widespread development in many applications since the technology does not depend on additional hardware deployment. However, the site survey process which is executed before fingerprint positioning is time consuming and labor intensive. Sampling cell size directly affects the localization precision and the survey workload. We propose a sampling cell size optimization method to release the survey workload while holding the precision. We adopted the Gaussian process model to estimate the received signal strength based on some samples and used the inherent maximum Euclidean error of fingerprint vectors to compute the optimal cell size. We conduct some experiments to validate the efficiency of the mechanism. And the results show that our method can largely reduce the workload of the site survey process.

1. Introduction

Indoor positioning is a critical enabler for location based services and applications. In many scenarios (e.g., shopping malls, supermarkets, and airport terminals), the location can provide the shopping guide, commodity promotion, and path navigation. Since the GPS technology is not available in the building, various wireless signal technologies (e.g., Zigbee, RFID, Bluetooth, UWB, and WiFi) have been applied to indoor positioning. However, accuracy and the overhead limit the feasibility of indoor localization system. Various positioning systems can achieve a high accuracy at a cost of additional hardware. For example, the systems based on the time of arrival (ToA) measurement need accurate time synchronization or specific hardware to compute the arriving time.

The fingerprint positioning is a promising indoor localization method which requires no additional hardware. Therefore, fingerprint positioning has been widely applied in many applications [1–3]. However, fingerprint positioning needs an additional site survey process to record the RSS fingerprints at every point of the interested area to build the fingerprints database. Meanwhile, sampling cell size (i.e., cell area) also needs to be defined before the site survey. This process is time consuming and labor intensive. Furthermore, the fingerprint database needs to be updated with the changing of environments. Much research has focused on avoiding the tedious presurvey. In some work, the crowdsourcing method is adopted to assemble the fingerprint information from masses of location based services’ users [4, 5]. This can reduce the individual labor through many people’s cooperation. But the sampling cell size still needs to be determined.

Larger sampling cell size implies the fewer survey workload but the lower positioning accuracy. Thus, we need to find the optimal sampling cell size to achieve a balance between the survey workload and positioning accuracy. The traditional methods adopt the default empirical values (e.g., 1m × 1m or 2m × 2m) [6, 7] or the users’ stride in the crowdsourcing. In this paper, we adopted the Gaussian process model to estimate the received signal strength based on some samples. Then we used the inherent maximum Euclidean error of fingerprint vectors to compute the optimal cell size. The experimental results show the availability of Gaussian processes and the efficiency of optimal cell size.
2. Performance Analysis of Different Sampling Cell Sizes

During the fingerprint positioning, the interested area is first divided into some equal cells, where the RSSs from WiFi access points (APs) are sampled to construct the fingerprint database. Cell size affects the fingerprint database scale and localization performance. To study the impact of the sampling cell size on the localization performance, we conducted three experiments where the cell sizes are set to 1 m × 1 m, 2 m × 2 m, and 3 m × 3 m, respectively, in a 30 m × 24 m office room with 8 deployed APs. Obviously, the workload is increased as the cell size decreases.

2.1. Methodology. After dividing the interested area into small cells, we need to sample the RSSs of APs in each cell. Assume the WiFi fingerprint RSS values at location $x_i$ are denoted by $(r_{i1}, r_{i2}, \ldots, r_{im})$ in which $r_{im}$ is the RSS value collected from the $m$th AP. When the user has to be localized, the similarity between the user’s fingerprint and the fingerprints in the database is used as the localization metric. Generally, the Euclidean distance is used as the similarity of fingerprint vectors which is represented as follows:

$$E_{ij} = \sqrt{\sum_{k=1}^{m} \left( r_{ik} - r_{jk} \right)^2},$$

where the $E_{ij}$ represents the Euclidean distance between fingerprint vectors in locations $x_i$ and $x_j$.

In the localization process, the K-nearest neighbors (KNN) are selected to estimate the user’s location; for example, the centroid of the K-nearest neighbors is regarded as the user’s current location. In the above three experiments, we adopt the KNN method.

In various positioning applications, the mobile phone has grown up to be a primary localization facility. We use GS1011 WiFi module from American Gainspan company as the test platform. This is because GS1011 module has a low-power property which can support a long time test. Additionally, to avoid the existing obstacles such as the office tables and chairs in the test room, we just experimented in the corridor cells to sample the fingerprints. According to the above dividing results, 120, 60, and 40 samplings sites are generated, respectively. After creating the fingerprint database, we select 60 random locations to evaluate the localization performance, where some locations may be far away from the sampling sites. As shown in Figure 1, we can find the 2 m × 2 m scenario has the comparable localization accuracy with 1 m × 1 m’s. But when the sampling cell size is 3 m × 3 m, the localization performance declines dramatically. In other words, the localization accuracy has not always been improved with subdividing of sampling cell.

2.2. Root Cause Analysis. To pursue the cause of above experimental results, we designed another test for WiFi signal attenuation with 4 m, 8 m, and 12 m distance away from APs. In Figure 2, the received signal strength is reducing with the distance increasing, while the signal variation is augmenting as well. This indicates the RSSs measured in the same sampling cell have a certain error. This is because complex indoor environment will bring about shadowing and multipath of radio signals. Therefore, during the KNN matching, the Euclidean distance between fingerprint vectors has some fluctuation due to the existing of RSS measurement errors. Smaller sampling cell size will lead to more influence from the Euclidean distance fluctuation. Therefore, we have to find the balance between the RSS measurement error and the sampling cell size.

3. Sampling Cell Size Optimization

In order to find the balance between the RSS measurement error and the sampling cell size, we defined the Euclidean distance of fingerprint vectors between sampling cells as fingerprint granularity. The fingerprint granularity represents...
the fuzzy degree of two sampling fingerprints. The bigger fuzzy degree indicates a higher similarity of the fingerprints. Meanwhile, the RSS measurement error also affects the fingerprint Euclidean distance. For example, supposing the RSS measurement error follows normal distributions with zero mean and $\varepsilon$ error variance, then the fingerprints in the same sampling cell have an inherent maximum Euclidean error as follows:

$$e_{\text{max}} = \sqrt{\frac{m}{\sum_{i=1}^{m} (r_i + \varepsilon) - (r_i - \varepsilon))^2}} = 2\varepsilon\sqrt{m},$$

(2)

where $m$ represents the number of APs. In addition, according to the results in Figure 2, the value of $\varepsilon$ will be changing with the distance away from APs. When the fuzzy degree of any two adjacent sampling cell fingerprints is over the maximum Euclidean error $e_{\text{max}}$, the fingerprint positioning algorithm can not distinguish two adjacent sampling sites. With the further reduction of the sampling cell size, the localization performance will not be clearly improved. This is the critical point that the performance is improved through reducing the optimal sampling cell size.

However, we can not compute the fingerprint granularity between any two adjacent sampling cells before the site survey. Meanwhile, the survey of an interesting area first needs to divide the area into the same cells. So the sampling cell size needs to be obtained in advance. This becomes a classic chicken-and-egg problem. To solve this difficulty, the existing techniques mostly assume knowledge of the APs locations and then adopt the signal propagation model to estimate each cell signal strength based on the distance from the APs [8, 9]. But the estimated value has lots of inaccuracy due to the ambient interferences. In this paper, we introduce the Gaussian processes (GPs) for modeling signal strength measurements to address these limitations. Based on fewer precalibration data, we can use Gaussian processes to generate a fingerprint database corresponding to all cells. In the following, we will show how Gaussian processes work on signal strength measurements.

3.1. Preliminaries. GPs can be derived in the function-space ways described in [10]. Let $D = (x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ be a set of samples at all cells with a noisy process. $n$ is the number of cells. Consider

$$y_i = f(x_i) + \varepsilon,$$

(3)

where $y_i$ represents a fingerprint vector at location $x_i$ which is an input sample and the fingerprint vector $y_i$ is a target value. $\varepsilon$ is zero mean, additive Gaussian noise with known variance $\sigma^2_m$. For notational convenience, we aggregate the $n$ input values $x_i$ into a matrix $X$ with $n$ row and $d$ column, where $d$ is the coordinate dimensions, and target values $y_i$ into the vector denoted $Y$.

GPs use precalibration data as the input values to estimate posterior distribution of fingerprint vectors through the functions $f$. But a key idea underlying GPs is the requirement that the function values at different cells are correlated, where the covariance between two functions $f(x_p)$ and $f(x_q)$ depends on the input values, $x_p$ and $x_q$. The dependency can be specified via an arbitrary covariance function or kernel $k(x_p, x_q)$. The Gaussian kernel is widely used as a choice of kernel function as follows:

$$k(x_p, x_q) = \sigma^2_f \exp\left(\frac{-1}{2\ell^2} |x_p - x_q|^2\right),$$

(4)

where $k(x_p, x_q)$ represents covariance between two functions $f(x_p)$ and $f(x_q)$. The $\sigma^2_f$ is the signal variance which is different from the signal error variance $\varepsilon$. $l$ is a length scale that describes how strongly the dropping of the correlation between input values. These parameters control the smoothness of the functions estimated by GPs. Seen from (4), the larger distance between two cells will lead to the lower covariance value.

During this Gaussian processes model, we do not consider the specific function value. Conversely we just focus on the covariance between arbitrarily two functions. So it is necessary to consider the corresponding observation noise covariance. The observation noise covariance between both cells can be described as follows:

$$\text{cov}(y_p, y_q) = k(x_p, x_q) + \sigma^2_n \delta_{pq},$$

(5)

where $\sigma^2_n$ is the Gaussian observation noise. When the $p$ and $q$ are the same cell, the $\delta_{pq}$ is one and zero otherwise. For applying to the entire input values $X$, the above covariance can be further modified to a general formulation. Consider

$$\text{cov}(Y) = K + \sigma^2_n I,$$

(6)

Here, $K$ is the $n \times n$ covariance matrix of the input values $X$. For example, $K(p, q) = k(x_p, x_q)$.

3.2. Optimal Sampling Cell Design. Based on the input values $X$ and corresponding observation $Y$, we can construct the above covariance. That is, the calibration data $Y$ is jointly Gaussian with $Y \sim \mathcal{N}(0, K + \sigma^2_n I)$. Here, we can predict the function value at an arbitrary cell $x_s$, conditioned on the calibration data $X, Y$. Known from (4), the posterior over function values are Gaussian with mean $\mu_{x_s}$ at point $x_s$ [11]:

$$\mu_{x_s} = k_s^T(K + \sigma^2_n I)^{-1}Y,$$

(7)

where $k_s$ is the $n \times 1$ vector of covariances between $x_s$ and all $n$ input values $X$, and $K$ is the covariance matrix of $X$. We can find that (7) shows a linear combination of the calibration values $Y$. During the signal strength modeling, the input values $X$ represent the cell locations, and target value $Y$ corresponds to the fingerprint vector. So the above mean $\mu_{x_s}$ is the signal strength fingerprint vector at location $x_s$. Then the Euclidean distance of arbitrary two adjacent cells corresponding to fingerprint vector can be represented as shown below:

$$\mu_{x_s} - \mu_{x_q} = k_s^T(K + \sigma^2_n I)^{-1}Y - k_q^T(K + \sigma^2_n I)^{-1}Y$$

$$= (k_s - k_q)^T(K + \sigma^2_n I)^{-1}Y$$

$$\geq 2\varepsilon\sqrt{m},$$

(8)
where the $k_{x_i} - k_{x_j}$ is the weight of each observation $y$ constructing Euclidean distance of fingerprint vector between locations $x_i$ and $x_j$. When the above Euclidean distance is less than the maximum Euclidean error $2\varepsilon\sqrt{m}$, the two adjacent cells will not be distinguished by the fingerprint positioning algorithm. Meanwhile, the $k_{x_i} - k_{x_j}$ can be further expanded as follows:

$$W_{ij} = k_{x_i} - k_{x_j} = k(x_i, x_j) - k(x_j, x_i),$$

(9)

where $k$ is $1, 2, \ldots, n$ and $n$ is the number of inputs $X$. $k(x_i, x_k)$ is also denoted by physical distance between $x_i$ and $x_k$ which is a function of cell side length. According to (8), we can compute a bound value of cell side length, that is, the optimal cell size to obtain better localization performance and less sampling effort. Before computing the optimal cell size, we firstly learn of these parameters based on the calibration data. The parameters include $\sigma_n^2$, $l$, $\sigma_f^2$, denoted by the hyperparameters, which can be estimated by maximizing the log likelihood of the fingerprint vector $y$. Since the calibration data $Y$ is jointly Gaussian, we can maximize its log likelihood using conjugate gradient descent (LBFGS) [12].

4. Experiments Evaluation

In the experiments, we evaluate the performance of GPs and the optimal sampling cell size using WiFi fingerprint data in our lab office.

4.1. Setup of Experiments. Our test scenario is an open office room with 30 m × 24 m, which consists of the workstations, corridors, and cement columns. To collect calibration data, we still select GS1011 WiFi module as the measurement platform. When the surrounding APs send beacon packets, the GS1011 model can receive signals and measure the received signal strengths. The calibration data were collected by manual mobile survey with handheld GS1011 model. The number of APs is still 8. We sampled 72 data records at 72 cells in the test room. At every cell, we sampled 20 times to compute the mean of RSSs without considering the body interference. Additionally, we randomly selected some data records to estimate the hyperparameters of the GPs. Then we used the estimated hyperparameters to construct the GPs. According to the constructed GPs, we computed the optimal sampling cell size through (8).

4.2. Experimental Results. To validate the feasibility of GPs model, we selected 15 sampled records to compute hyperparameters and construct GPs. Then another 15 cells were selected to estimate the RSS fingerprint vectors according to the newly constructed GPs. On one AP, the measurement RSSs and estimated values at above 15 cells were shown in Figure 3. The box chart represents the measurement sample and the black rectangle represents the estimated value by GPs model. From this figure, we can find the measurement RSS values have a bigger variance when the RSS values become weak. It is consistent with our results in Figure 2. Meanwhile, the estimated value also deviates from measurement mean with the fading of signal strength. This indicates the GPs model is suitable for estimating the stable data. On the other hand, the experimental results also indicate that the GPs model can be implemented to signal strength modeling.

Then we changed the number of sample records to recompute hyperparameters and GPs model. The relation between sample numbers and estimated errors was shown in Figure 4. We selected 5, 10, 15, 20, 25, and 30 samples to test the estimated errors in two random cells. The means of measured RSSs in two cells are $-49$ dB and $-72$ dB, respectively. From the experimental results, we can find the estimated errors will reduce with the increasing of sample numbers. Meanwhile, for the same samples, the lower measured RSSs lead to a bigger estimated error. This is because the lower RSSs have a bigger variance.

The AP number is another main factor to affect the estimated errors from GPs model. We selected 2, 4, and 8 of the test APs to estimate the RSS corresponding to the 10, 20, and 30 samples, respectively. The results were shown in Figure 5. The increasing of AP numbers will lead to the
enlarging of the estimated mean errors about RSS. This is because more APs will accumulate the estimated errors. Besides, the samples also affect the estimated accuracy of GPs model corresponding to Figure 4.

However, less APs are likewise no better, because less APs will reduce the dimensions of fingerprint vector so as to increase the fingerprint fuzziness. We design another test to verify the APs affection for two adjacent sampling cell fingerprints. 30 adjacent fingerprints with $2 \times 2$ m cell size are selected to compute the Euclidean distance between each other. The computed results of 2, 4, 6, and 8 APs are shown in Figure 6. More APs will increase the mean of Euclidean distance. Then the fingerprint fuzziness will be reduced.

According to (8), we can compute the optimal cell size from GPs model based on some samples. In our experiments, we adopted the 30 samples and the optimal cell size is computed as $1.8 \times 1.8$ m. Then we compared the performance of cell size $1.8 \times 1.8$ m with the above three scenarios $1 \times 1$, $2 \times 2$, and $3 \times 3$ m as shown in Figure 7. We can find that the positioning performance of $1.8 \times 1.8$ m falls in between $1 \times 1$ m and $2 \times 2$ m. But the interest area is difficult to be divided into $1.8 \times 1.8$ m cells. For survey convenience, we usually select the $2 \times 2$ m cell to sample the received signal strengths as the fingerprint vectors.

5. Conclusions

In this paper, we first analyze the influence of the sampling cell size for fingerprint positioning performances and find the root cause of the similarity between $1 \times 1$ m and $2 \times 2$ m scenarios. In order to achieve the optimal sampling cell size, we adopted the Gaussian process model to estimate the received signal strengths based on a few samples. Then we used the inherent maximum Euclidean error of fingerprint vectors to compute the optimal cell size. The experiment results also show the availability of Gaussian processes and the efficiency of optimal cell size.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The work was supported by the National Natural Science Foundation of China (61202436, 61202066), National High Technology Research and Development Program of China (2013AA014002), “Strategic Priority Research Program” of the Chinese Academy of Sciences (XDA06040100), and Youth Foundation of Taiyuan University of Technology (2013Z060).

References

[1] J. Niu, B. Lu, L. Cheng, Y. Gu, and L. Shu, “ZiLoc: energy efficient WiFi fingerprint-based localization with low-power...
radio,” in Proceeding of the IEEE Wireless Communications and Networking Conference (WCNC ’13), pp. 4558–4563, Shanghai, China, April 2013.

[2] J.-H. Kim and W.-Y. Yeo, “A coherent data filtering method for large scale rf fingerprint wi-fi positioning systems,” EURASIP Journal on Wireless Communications and Networking, vol. 2014, article 13, 2014.

[3] Y. Gao, J. Niu, R. Zhou, and G. Xing, “ZiFind: exploiting cross-technology interference signatures for energy-efficient indoor localization,” in Proceedings of the IEEE Conference on Computer Communications (INFOCOM ’13), pp. 2940–2948, Turin, Italy, April 2013.

[4] C. Wu, Z. Yang, Y. Liu, and W. Xi, “WILL: wireless indoor localization without site survey,” IEEE Transactions on Parallel and Distributed Systems, vol. 24, no. 4, pp. 839–848, 2013.

[5] A. Rai, K. K. Chintalapudi, V. N. Padmanabhan, and R. Sen, “Zee: zero-effort crowdsourcing for indoor localization,” in Proceedings of the 18th Annual International Conference on Mobile Computing and Networking (MobiCom ’12), pp. 293–304, ACM, August 2012.

[6] G. Tsagkatakis and P. Tsakalides, “Efficient high dynamic range imaging via matrix completion,” in Proceedings of the 22nd IEEE International Workshop on Machine Learning for Signal Processing (MLSP ’12), pp. 1–6, September 2012.

[7] C. Wu, Z. Yang, Z. Zhou, Y. Liu, and M. Liu, “DorFin: WiFi fingerprint-based localization revisited,” http://arxiv.org/abs/1308.6663.

[8] P. Barsocchi, S. Lenzi, S. Chessa, and G. Giunta, “Virtual calibration for RSSI-based indoor localization with IEEE 802.15.4,” in Proceedings of the IEEE International Conference on Communications (ICC ’09), pp. 1–5, June 2009.

[9] I. Bisio, M. Cerruti, F. Lavagetto et al., “A trainingless wifi fingerprint positioning approach over mobile devices,” 2014.

[10] C. E. Rasmussen, Gaussian Processes for Machine Learning, 2006.

[11] B. Ferris, D. Haehnel, and D. Fox, “Gaussian processes for signal strength-based location estimation,” in Proceedings of the Robotics Science and Systems, Citeseer, 2006.

[12] A. M. Ladd, K. E. Bekris, A. Rudys, L. E. Kavraki, and D. S. Wallach, “Robotics-based location sensing using wireless ethernet,” Wireless Networks, vol. 11, no. 1-2, pp. 189–204, 2005.
