Journal of Network and Computer Applications 62 (2016) 128–136

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

journal homepage: www.elsevier.com/locate/jnca

Accuracy-aware wireless indoor localization: Feasibility and applications

Chengwen Luo [a,c,*], Hande Hong [c], Long Cheng [b], Mun Choon Chan [c], Jianqiang Li [a], Zhong Ming [a]

[a] College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China
[b] VirginiaTech, Blacksburg, VA 24061, USA
[c] School of Computing, National University of Singapore, Singapore 117417, Singapore

ARTICLE INFO

Article history:
Received 22 May 2015
Received in revised form
22 October 2015
Accepted 21 December 2015
Available online 6 January 2016

Keywords:
Accuracy awareness
Indoor localization
Gaussian process

ABSTRACT

Fingerprint-based indoor localization has attracted extensive research efforts due to its potential for deployment without extensive infrastructure support. However, the accuracies of these different systems vary and it is difficult to compare and evaluate these systems systematically. In this work, we propose a Gaussian process based approach that takes the radio map and the localization algorithm as an input, and outputs the expected accuracy of the localization system. With an efficient error estimation algorithm, many applications such as landmark detection, localization algorithm selection and access point subset selection can be performed. Our evaluations show that our approach provides sufficient accuracy and can serve as a useful tool for system evaluation and performance tuning when developing fingerprint-based indoor localization systems.

1. Introduction

Accurate indoor localization is the fundamental building block for mobile pervasive computing. With the proliferation of smartphones and the rise of location-based services, indoor localization has attracted extensive research efforts (Yang et al., 2012; Wang et al., 2012; Bahl and Padmanabhan, 2000; Youssef and Agrawala, 2005; Shen et al., 2013; Rai et al., 2012; Haque, 2014; Zhang and Wong, 2012).

Among different categories of wireless indoor localization approaches, *fingerprint-based indoor localization* (Bahl and Padmanabhan, 2000) is one of the most popular due to the widespread availability of WiFi access points (AP). WiFi fingerprints consist of the received signal strength (RSS) of WiFi APs, and are used as unique signatures of different locations to determine the location of mobile devices in the localization process. State-of-the-art research on fingerprint-based indoor localization either focuses on improving the accuracy of the location estimation (Youssef and Agrawala, 2005; Liu et al., 2012; Sun et al., 2013), or reducing the time and effort in constructing the fingerprint database (Yang et al., 2012; Wang et al., 2012; Shen et al., 2013; Luo et al., 2014). While all these approaches have addressed various shortcomings in existing indoor localization systems, the configurations of these systems vary. Each of these localization systems is evaluated in settings with different physical layout and environmental effects, making it difficult to compare and evaluate them systematically. One of the key objectives of our work is to make systematic comparison feasible.

The main idea of this paper is as follows: given a set of radio signal fingerprints collected, a Gaussian process (GP) (Rasmussen, 2006) approach is used to model the signal distribution of access points that cover the area of interest. Using the signal distribution model derived, random sampling is performed to simulate the collection of fingerprint values collected at each location of interest during localization. Given a particular localization algorithm, the mapped location in the system can be determined. The average localization error of each location in the area of interest can now be estimated.

By decoupling radio map construction and localization, and with the ability to estimate the accuracy of the localization system over the area of interest, our system can achieve the following:

- It is now possible to systematically compare different localization algorithms under different environmental settings.
- Landmarks, or locations with high localization confidence, can be easily identified and used to further improve the accuracy.
- The set of APs that can provide better accuracy for the entire area of interest can be identified, as opposed to using all APs

* Corresponding author.
E-mail addresses: chluo@comp.nus.edu.sg (C. Luo), honghand@comp.nus.edu.sg (H. Hong), chengl@vt.edu (L. Cheng), chanmc@comp.nus.edu.sg (M.C. Chan), lijq@szu.edu.cn (J. Li), mingz@szu.edu.cn (Z. Ming).

http://dx.doi.org/10.1016/j.jnca.2015.12.008
1084-8045/

available or a set of APs that may be good locally but not for the entire area.

To the best of our knowledge, this paper is the first systematic study on accuracy awareness of fingerprint-based localization systems. We believe that it has the potential to be integrated into future fingerprint-based localization systems to provide direct feedback about the accuracy levels of the system in use, and guidelines to achieve better accuracy.

To validate our approach, we evaluate the system in two different indoor environments covering more than 300 m^2. In both environments, point-level, region-level and floor-level error estimation are evaluated with 3 different localization metrics and more than 20,000 testing data points. For point level accuracy, the evaluation results show that the difference between GP estimation and ground truth is small, demonstrating that accuracy awareness provides an accurate and practical assessment to fingerprint-based localization systems. In addition, we are able to successfully identify 5 landmarks with high localization confidence in the area localized and find the minimum AP subsets that should be selected to achieve better accuracy.

The rest of the paper is organized as follows. We discuss the Gaussian process for modeling signal strength of access points in Section 2. Section 3 explains the concepts, algorithms and applications for point-level, region-level and floor-level accuracy awareness. The evaluation results are given in Section 4 and related works are discussed in Section 5. Finally, we conclude in Section 6.

2. Preliminaries

The received signal strength of the wireless access point at each location has been characterized in the literature as a Gaussian distribution (Haeberlen et al., 2004; Youssef and Agrawala, 2005; Kaemarungsi and Krishnamurthy, 2004; Ferris et al., 2007). To model the signal strength propagation continuously over the whole field, Gaussian process (GP) (Rasmussen, 2006) is used to capture the spatial correlation existed in signal strength distribution (Ferris et al., 2006, 2007; Xu et al., 2014). Gaussian process is a Bayesian non-parametric model that performs non-linear regression on the training data $\mathcal{D} = \{(\mathbf{x_i}, y_i) | i = 1, \dots, n\}$ to estimate the distribution over functions f that generate the data. That is,

$$y_i = f(\mathbf{x_i}) + \varepsilon \tag{1}$$

where $\mathbf{x_i} \in \mathbb{R}^d$ is a d dimensional input value, y_i is the observation value, and ε is a zero-mean noise term with known covariance ${\sigma_n}^2$. Gaussian processes allow spatial correlation between measurements and are fully specified by GP priors. Therefore, function $f \sim \mathcal{GP}(\mu(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$ is a GP with mean function $\mu(\mathbf{x})$ and covariance function, or kernel, $k(\mathbf{x}, \mathbf{x}')$, where

$$\mu(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})] \tag{2}$$

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - \mu(\mathbf{x}))(f(\mathbf{x}') - \mu(\mathbf{x}'))] \tag{3}$$

The choices of the kernel function characterize the property of GP, and the most widely used kernel is the *squared exponential* function (Ferris et al., 2007):

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left(-\frac{1}{2l^2}|\mathbf{x} - \mathbf{x}'|^2\right) \tag{4}$$

where σ_f^2 is the variance of observation value and l is the length scale that decides how strongly the correlation between different points drops off (Ferris et al., 2007). Assuming additive independent identically distributed Gaussian noise ε and noise covariance ${\sigma_n}^2$ (Rasmussen, 2006), the covariance between observations becomes:

$$\text{cov}(f(\mathbf{x}), f(\mathbf{x}')) = k(\mathbf{x}, \mathbf{x}') + \sigma_n^2 \delta_{\mathbf{x},\mathbf{x}'} \tag{5}$$

Here $\delta_{\mathbf{x},\mathbf{x}'} = 1$ if $\mathbf{x}$ and $\mathbf{x}'$ are the same point, and 0 otherwise. After the prior is specified, the Gaussian process posterior is obtained from the training data $\mathcal{D}$. Therefore with GP priors and training data, prediction of the unobserved function value at any arbitrary location $\mathbf{x}^*$ can be made (Xu et al., 2014):

$$\mu_{\mathbf{x}^*|\mathcal{D}} = \mu_{\mathbf{x}^*} + \Sigma_{\mathbf{x}^*\mathcal{D}} \Sigma_{\mathcal{D}\mathcal{D}}^{-1} (y_{\mathcal{D}} - \mu_{\mathcal{D}}) \tag{6}$$

Here $\mu_{\mathbf{x}^*}$, $\mu_{\mathcal{D}}$ are the mean values of the data points and are specified by the GP prior $\mu(\mathbf{x})$. $\Sigma_{\mathbf{x}^*\mathcal{D}}$ is the $1 \times n$ vector of covariance between x^* and the n training data $\mathcal{D}$, and $\Sigma_{\mathcal{D}\mathcal{D}}^{-1}$ is the $n \times n$ covariance matrix of the training data. Both $\Sigma_{\mathbf{x}^*\mathcal{D}}$ and $\Sigma_{\mathcal{D}\mathcal{D}}^{-1}$ are calculated using Eq. (5). With this formulation, the observation value at any arbitrary location in the field can be predicted conditionally on the training data.

To model the signal strength distribution of the access points covering a certain area, input $\mathbf{x} = (x_h, x_v)$ is a two dimensional vector specifying the horizontal and vertical coordinates of the location. The observation value y_i is the signal strength received at the given location. Note that the input data $\mathcal{D}$ here can be obtained from the fingerprint database, or radio map, which is generally required and constructed by any fingerprint-based localization systems in the offline calibration phase in order to perform localization.

The radio map contains a sequence of records $(\mathbf{x}, \mathbf{fp})$ which associates wireless fingerprints $\mathbf{fp}$ to each location $\mathbf{x}$. Fingerprint $\mathbf{fp} = (BSSID_i, r_i | i = 1, \dots, k)$ consists of signal strength readings r of all k WiFi BSSIDs (MAC addresses of access points) observable. Hence for each BSSID in the system, the training data $\mathcal{D} = \{(\mathbf{x_i}, r_i) | i = 1, \dots, n\}$ is available. With the availability of the training data, Gaussian processes can be applied to characterize the signal strength distribution of the whole area.

The squared exponential kernel in Eq. (4) assumes the same length scale in all input dimensions. However, in practice the effect of horizontal or vertical dimensions on signal strength can be different due to the physical settings. For example, there can be a wall in the horizontal dimension blue, resulting in the fast decay of signal strength in only this dimension. To model this effect, we use separate length scale l_h and l_v in each dimension in modeling the signal strength:

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left[-\frac{1}{2}\left(\frac{(x_h - x_h')^2}{l_h^2} + \frac{(x_v - x_v')^2}{l_v^2}\right)\right] \tag{7}$$

The mean function and the covariance function characterize the signal strength model. To handle the mean shift problem, we set mean function $\mu(\mathbf{x}) = -100$, so that those locations that are not able to receive any signal strength of certain access point will converge to mean -100 dbm in its model. The covariance function contains four parameters $\theta = \langle \sigma_n, \sigma_f, l_h, l_v \rangle$.

One advantage of the Gaussian process is that it is a non-parametric model, and therefore no parameters need to be specified beforehand: all parameters are learned from the training data by maximizing the log likelihood using the conjugate gradient decent algorithm (Ferris et al., 2006). Fig. 1(a) shows the GP estimation of the mean signal strength value for one access point covering a 20×12 m^2 indoor area. Note that even though the GPs also provide uncertainty measurement for Eq. (6) (e.g., the variance of the predicted $\mu_{\mathbf{x}^*|\mathcal{D}}$ (Ferris et al., 2006)), it only measures the "spatial uncertainty" of the predicted mean. This uncertainty is different from the "temporal uncertainty", which is the variance of signal strength at each location at various times. The temporal uncertainty provides the likelihood measurement for the signal strength. To model temporal uncertainty, we treat variance as the second variable and train a second GP for the same access point using mean function $\mu(\mathbf{x}) = 0$ and the same covariance function (7)