Computer Communications 83 (2016) 56–63

Contents lists available at ScienceDirect

Computer Communications

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

Reducing fingerprint collection for indoor localization[☆]

Zhuan Gu[a], Zeqin Chen[a], Yuexing Zhang[a], Ying Zhu[a], MingMing Lu[b], Ai Chen[a,*]

[a] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, PR China
[b] Central South University, Changsha 410083, PR China

ARTICLE INFO

Article history:
Received 8 February 2015
Revised 27 August 2015
Accepted 21 September 2015
Available online 30 September 2015

Keywords:
Compressive sensing
Fingerprint collection
Indoor localization
Interpolation
Merging matrix

ABSTRACT

A typical WiFi-based indoor localization technique estimates a device's location by comparing received signal strength indicator (RSSI) against stored fingerprints and finding the closest matches. However, the collection of fingerprints is notoriously laborious and costly. It is challenging to reduce fingerprint collection and recover missing data without introducing significant errors. In this article, a novel approach based on compressive sensing is presented for recovering absent fingerprints. The hidden structure and redundancy characteristics of fingerprints are revealed in a merging matrix. The spatial and temporal correlations of fingerprints result in a small rank of the merging matrix. The *Sparsity Rank Singular Value Decomposition* (SRSVD) method is used to effectively reduce the interference caused by the multipath effect of the WiFi signal. We further propose to combine SRSVD with the *K*-Nearest Neighbor (KNN) algorithm to deal with missing columns or rows in the matrix. Experimental results show that with only half of the fingerprints, our approach can recover all the fingerprint information with error rate below 6.6%. Even with only 5% of the data, the approach can recover the information with error rate below 14%, without loss of localization accuracy.

1. Introduction

Numerous indoor localization systems are based on wireless local area networks (WLANs), which are ubiquitously deployed in public places [2]. The localization techniques underlying these systems can be generally classified into two categories: deterministic [3–5] and probabilistic [6,7]. In these systems, one has to measure received signal strength indicator (RSSI) values from surrounding access points (APs) at each reference location to construct a fingerprint database, which is a tedious and time-consuming process. For example, in our experiments, it takes 10 h for us to collect the fingerprint data of an office area of 1000 m^2. This problem seriously affects the application of indoor localization systems. In order to reduce the cost, several approaches [8,9] have been proposed. However, these approaches only attempt to reduce the number of reference points at which fingerprints are collected. At each point, the same amount of time still needs to be spent on collecting a stable RSSI result as in the naive approach.

In this article, we focus on reducing fingerprint measurements in both the time and space domains, and recovering the absent data faithfully. The main challenge is to keep a small localization error while recovering the absent data. Intuitively, we can consider the simple interpolation approach, such as the *K*-Nearest Neighbors (KNN) method. However, the relation between an absent data point and its neighbors is not easy to identify. Moreover, if a large number of data points are missed, the KNN method will perform poorly because a point may not be able to find enough neighbors in range for proper estimation.

We find that a compressive sensing based algorithm, namely *Sparsity Rank-Singular Value Decomposition* (SRSVD) [10] can solve the problem of recovering absent data. The challenge is how to model our problem and leverage the hidden structure and redundancy of the collected data. We use the merging matrix [10] to merge and arrange all the RSSI values collected at different locations and at different times. In order to simplify the analysis, we use the rank of a matrix to judge the sparsity in compressive sensing. We formulate the problem as an optimization problem and try to find solutions.

SRSVD is a mathematical method to sparsity matrices. However, the characteristic structure of a merging matrix tends to be uncertain, due to the complex indoor environments. In reality, it is possible that several columns or rows are absent in the matrix at the same time, in which case SRSVD delivers quite low performance. We combine the SRSVD algorithm with the *K*-Nearest Neighbor (KNN) algorithm to address the problem. The approach interpolates only one element in an absent column or row and recovers the rest of absent data in the sparse matrix.

The major contributions of this paper are as follows. First, we propose to use the merging matrix to find the hidden structure and redundancy characteristics of the fingerprints. Second, we propose a

☆ A preliminary version of this article appeared in *Proceedings of the Eleventh IEEE Wireless Communications and Networking Conference (IEEE WCNC 2013)* [1].

* Corresponding author. Tel.: +86 755 8639 2325.
E-mail address: ai.chen@siat.ac.cn (A. Chen).

http://dx.doi.org/10.1016/j.comcom.2015.09.022
0140-3664/

novel approach based on compressive sensing to recover the absent data in the merging matrix. Third, we combine the SRSVD algorithm with the K-Nearest Neighbor (KNN) algorithm to deal with missing columns or rows in the matrix. Finally, we conduct experiments to evaluate the performance of the methods. The results show that given only half of the data, our approach can fully recover the fingerprint information with error rate below 6.6%. With merely 5% of the data, the approach can still recover the information with error rate below 14% and without loss of localization accuracy.

Organization: Section 2 introduces prior work on reducing fingerprint collection; Section 3 formulates the problem; Section 4 analyzes the problem and describes the solution; experiments are conducted in Section 5; and Section 6 concludes the paper.

2. Related work

One of the main methods of WiFi localization is based on RSSI. The related techniques can be classified into two categories: *propagation model based* and *fingerprint based*.

The propagation model based approach does not need signal fingerprints. Ubicarse [11] leverages Synthetic Aperture Radar (SAR) on hand-held devices that are twisted by their users along unknown paths. Ubicarse combines RF localization with stereo-vision algorithms to localize common objects with no RF source attached to them. The method in [12] leverages such a model to estimate the distance between the user and APs. Then it uses the extended Kalman filter to transform the distance to the user's position. The technique in [13] allows an organic positioning system to maintain its accuracy over time, based on outlier detection through clustering. A novel technique is proposed in [14], which uses the Gaussian Process Latent Variable Model (GPLVM) to relate RSS fingerprints and models of human movements (displacement, direction, etc.) as hidden variables. Utilizing a probabilistic RSS model derived from indoor signal propagation models that explicitly consider the effect of intervening walls and the building plan, the scheme in [15] first estimates the distance of the client to each of the APs and then obtains a location estimate through trilateration.

Radio propagation models are not very accurate for distance and position estimation, due to the multi-path effect and environmental interference. In comparison, the fingerprint based algorithms (e.g., [3,16]) normally have higher localization accuracy. In this approach, fingerprints collected with coordinate information are called labeled data and those without coordinates unlabeled data. The Label Propagation algorithm (LP-algorithm), by using semi-supervised learning in [8], tries to reduce the effort of collecting labeled data. In summary, these algorithms can reduce the work of labeling data, but still require collecting a large amount of unlabeled data.

In [17], a technique called the Signal-Distance Map (SDM) is proposed. SDM uses a truncated singular value decomposition technique to relate RSSI with geographical distance to the APs. Zee [18] and UnLoc [19] utilize WiFi and inertial sensors readings crowdsourced from users to build the fingerprint training set. In [20], a method called Walkie-Markie generates indoor pathway map by leveraging the locations of users when they pass WiFi marks. Furthermore, it uses the direction and distance information retrieved from the user trajectories to place the WiFi marks at real locations. Phaser [21] makes phased array signal processing practical on many WiFi access points deployed in the real world. In contrast, SpotFi [22] deployed on commodity WiFi infrastructure is able to achieve accuracy of 40 cm by calculating AoA of multipath components.

Compressive sensing techniques have been considered for reducing fingerprinting effort in a number of previous researches. In [23,24], the authors use l_1-minimization to solve the sparse signal recovery problem and use the map-adaptive Kalman filter to improve accuracy. Bayesian Compressive Sensing (BCS) based compressive sensing is used in [25,26]. In these techniques, the systems make full use of the relationship between the collected signals in the space. In [27], a multivariate Gaussian model is used to average the measurements of RSSI in the first step, and then compressive sensing is used to reduce the amount of information transmitted from a device in the second step. The Matrix Completion (MC) framework in [28] minimizes the number of RSSI fingerprints by sensing a subset of the available channels in a WiFi network. It provides a paradigm for reconstructing low-rank data matrices from a small number of randomly sampled entries. In the project, Environmental Space Time Improved Compressive Sensing (ESTI-CS) [29], real sensory data from the Intel Indoor, GreenOrbs, and Ocean Sense data sets are analyzed using the Multi-Attribute Assistant (MAA) component for data reconstruction.

Apart from RSSI-based methods, there are some other techniques of localization that are often used in combination with WiFi. Proximity detection is perhaps the simplest localization method. In such a method, the device estimates its location by simply detecting nearby radio sources [30]. The triangulation method provides improved localization accuracy by measuring the device's distance to multiple reference points [31,32]. When radio signal is missed during the user's navigation, dead reckoning is often used to fill in the gaps. Dead reckoning is a process of estimating the current position based on last determined position and incrementing that position based on known or estimated speeds over elapsed time [33]. For improved accuracy on real maps, the map matching techniques can be used. They include topological analysis, pattern recognition, or advanced techniques such as hierarchical fuzzy inference algorithms [19,34,35].

3. Problem formulation

This section describes the localization model and formulates the problem of recovering absent fingerprint data.

3.1. The system model

Normally, a WiFi-based localization system works in two phases: offline phase and online phase.

Offline phase: The indoor region of interest is divided into small grids. At the center point of each grid (reference location), an RF receiver collects the RSSI of each pre-deployed AP. At each reference location, the ID, coordinates, as well as each AP's RSSI, are recorded. These three elements together are called a *fingerprint*).

Online phase: The localization system estimates a device's location by comparing the measured RSSI against the fingerprints, and finding the closest matches.

3.2. Problem statement

In the offline phase, the device collects W RSSI values at each of the N reference location. Every time it measures the RSSI of M APs. Let $\mathbf{X}$ be the merging matrix of fingerprints with dimensions $W \cdot M \times N$. Also $\mathbf{X} = [\mathbf{E}_1\, \mathbf{E}_2 \ldots \mathbf{E}_W]^{\mathrm{T}}$, where $(\bullet)^{\mathrm{T}}$ is the transpose of a matrix, $\mathbf{E}_w$ the wth $(w = 1, 2, \ldots, W)$ sub-matrix of $\mathbf{X}$, and $\mathbf{E}_w(m, n)$ the RSSI of the mth $(m = 1, 2, \ldots, M)$ AP at the nth $(n = 1, 2, \ldots, N)$ location. We use an indicate matrix, $\mathbf{A}$, and a measurement matrix, $\mathbf{B}$, to represent the problem.

$$\mathbf{B} = \mathbf{A} .* \mathbf{X}$$
$$\mathbf{A} = [\mathbf{A}(i, j)] = \begin{cases} 0, & \text{if } \mathbf{B}(i, j) \text{ is absent} \\ 1, & \text{otherwise} \end{cases} \tag{1}$$

where the symbol '.∗' denotes the dot multiplication, which means the multiplication of the corresponding elements in the matrix. The zero elements in matrix $\mathbf{A}$ mean that the corresponding data elements in matrix $\mathbf{B}$ are absent.

Designing an algorithm to recover the fingerprint matrix $\mathbf{X}$ based on the measurement matrix $\mathbf{B}$ with the absent elements can reduce the efforts of data collection.