UNIVERSITÉ PARIS SUD XI - DÉPARTEMENT D' INFORMATIQUE

UNIVERSITY OF CRETE - DEPARTMENT OF COMPUTER SCIENCE

M A S T E R T H E S I S

to obtain the title of

Master of Science

of the University of Paris Sud XI - Paris, France
Specialty : Computer Science

Defended by

Dimitrios Milioris

Low-dimensional Signal-Strength Fingerprint-based Positioning in Wireless LANs

Thesis Advisor: Philippe Jacquet

prepared at INRIA Rocquencourt, Hipercom Team

defended on June 15, 2011

Submitted to the
Department of Computer Science
in partial fulfillment of the requirements for the degree of
Master of Science

Author:

Dimitrios MILIORIS
Department of Computer Science

Committee

Supervisor

Philippe JACQUET
Professor, École Polytechnique

Member

Véronique BENZAKEN
Professor, Université Paris Sud XI

Member

Yannis MANOUSSAKIS
Professor, Université Paris Sud XI

Accepted by:

Chairman of the
Graduate Studies Committee

Yannis STYLIANOU
Associate Professor, University of Crete

Paris, June 2011

Acknowledgments

It is a pleasure to thank the many people who made this thesis possible.
I would like to express my gratitude to my supervisors, Dr. Philippe Jacquet and Prof. Panagiotis Tsakalides, who were abundantly helpful and offered invaluable assistance, support and guidance.

Deepest gratitude are also due to Dr. George Tzagkarakis without whose knowledge and assistance this study would not have been successful.

I would like also to thank Prof. Maria Papadopouli. Through our cooperation she taught me "It's not only research that matters - Life is more important".

I am grateful to the secretary in the Computer Science department of University of Crete, for helping the department to run smoothly and for assisting me in many different ways. Mrs. Rena Kalaitzaki, deserves special mention.

I wish to offer my regards and blessings to all of my best friends in undergraduate and graduate level for helping me get through the difficult times, and for all the emotional support, camaraderie, entertainment, and caring they provided.

I would like also to convey thanks to the Faculty of University of Crete, Paris-Sud XI University, FO.R.T.H - ICS and I.N.R.I.A - Hipercom Team for providing the financial means and laboratory facilities.

Lastly, and most importantly, I wish to thank my sister, Maria, and my parents, Milioris Spyridonas and Anastasiou Magdalini. They bore me, raised me, supported me, taught me, and loved me. To them I dedicate this thesis.

Milioris S. Dimitrios

Εντοπισμός θέσης σε ασύρματα δίκτυα βασισμένος σε χαμηλής διάστασης αποτυπώματα

Περίληψη: Ο ακριβής εντοπισμός θέσης σε εσωτερικό περιβάλλον αποτελεί μια σημαντική διαδικασία για πολλά διάχυτα υπολογιστικά συστήματα, με πληθώρα εφαρμογών που βασίζονται στο IEEE 802.11, Blutetooth, τεχνολογίες υπερήχων και σε τεχνολογίες υπερήθρων.

Η έμφυτη αραιότητα που παρουσιάζεται στο πρόβλημα του εντοπισμού θέσης μας παρακινεί να χρησιμοποιήσουμε την πρόσφατη θεωρία της συμπιεστικής δειγματοληψίας (Compressive Sensing - CS), η οποία ορίζει οτι εάν έαν σήμα έχει αραιή αναπαράσταση σε μια κατάλληλη βάση τότε μπορεί να ανακατασκευαστεί με μεγάλη ακρίβεια από ένα μικρό αριθμό τυχαίων γραμμικών προβολών.

Η παρούσα μεταπτυχιακή εργασία εισάγει αρχικά μια νέα τεχνική εύρεσης θέσης μοντελοποιώντας τις μετρήσεις της ισχύς του σήματος - που έχουν ληφθεί από διάφορα σημεία πρόσβασης (Access Points - APs) - με πολυδιάστατη Γκαουσιανή κατανομή. Χρησιμοποιείται επίσης η θεωρία της συμπιεστικής δειγματοληψίας για την ακριβή εκτίμηση εύρεσης θέσης βασισμένη στις μετρήσεις της ισχύς του σήματος, ενώ μειώνεται σημαντικά η ποσότητα της πληροφορίας που εκπέμπεται από μια ασύρματη συσκευή με ελάχιστους πόρους ενέργειας, δυνατότητα αποθήκευσης και επεξεργασίας, σε μια κεντρική μηχανή (server). Η δυνατότητα της θεωρίας της συμπιεστικής δειγματοληψίας να συμπεριφέρεται σαν αδύναμη διαδικασία κρυπτογράφησης αποδεικνύεται από το γεγονός ότι η προτεινόμενη προσέγγιση παρουσιάζει αυξημένη ανθεκτικότητα σε πιθανές εισβολές από κακόβουλους χρήστες. Τέλος, παρουσιάζεται ένα υβριδικό μοντέλο εκτίμησης πορείας, το οποίο χρησιμοποιεί την αποδοτικότητα του φίλτρου Kalman σε συνδυασμό με την θεωρία της συμπιεστικής δειγματοληψίας και το πολυδιάστατο Γκαουσιανό μοντέλο.

Τα πειράματα αποκαλύπτουν αυξημένη επίδοση εκτίμησης θέσης, ενώ τα επίπεδα υπολογιστικής πολυπλοκότητας παραμένουν χαμηλά, σε σύγκριση με προηγούμενες στατιστικές μεθόδους.

Λέξεις - κλειδιά: Συμπιεστική δειγματοληψία (CS), αραιή αναπαράσταση, πολυδιάστατο Γκαουσιανό μοντέλο, απόκλιση Kullback-Leibler, μετρήσεις λαμβανόμενης ισχύς του σήματος, φίλτρο Kalman

Low-dimensional Signal-Strength Fingerprint-based Positioning in Wireless LANs

Abstract: Accurate indoor localization is a significant task for many ubiquitous and pervasive computing applications, with numerous solutions based on IEEE 802.11, Bluetooth, ultrasound and infrared technologies being proposed.

The inherent sparsity present in the problem of location estimation motivates in a natural fashion the use of the recently introduced theory of compressive sensing (CS), which states that a signal having a sparse representation in an appropriate basis can be reconstructed with high accuracy from a small number of random linear projections.

This thesis proposes a novel localization technique based on a multivariate Gaussian modeling of the signal strength measurements collected from several access points (APs) at different locations. It is also exploited the framework of CS to perform accurate indoor localization based on signal-strength measurements, while reducing significantly the amount of information transmitted from a wireless device with limited power, storage, and processing capabilities to a central server. Equally importantly, the inherent property of CS acting as a weak encryption process is demonstrated by showing that the proposed approach presents an increased robustness to potential intrusions of an unauthorized entity. Finally, a hybrid path tracking system is presented, which exploits the efficiency of a Kalman filter in conjunction with the power of compressive sensing to represent accurately sparse signals and a region-based multivariate Gaussian model.

The experimental evaluation reveals an increased localization performance, while maintaining a low computational complexity when compared with previous statistical fingerprint-based methods.

Keywords: Compressive sensing, sparse representation, multivariate Gaussian model, Kullback-Leibler divergence, RSS measurements, Kalman filter

Low-dimensional Signal-Strength Fingerprint-based Positioning in Wireless LANs

Résumé: La localisation précise en intérieur est un élément important pour nombre d'applications pervasives et ubiquitaires et met en oeuvre des solutions diverses basées sur IEEE 802.11, Bluetooth, les ultrasons, et les infrarouges.

La faible densité du signal présenté dans le problème de la localisation est une propriété exploitée dans la récente théorie du "Compressive Sensing" (CS), qui établit qu'un signal qui a une faible densité dans une base donnée peut être reconstruit avec grande précision a partir d'un petit nombre de projections linéaires.

Cette thèse propose une approche innovante basée sur un modèle Gaussien à plusieurs dimensions des mesures des signaux collectes sur des points d'accès (APs) à des positions différentes. La théorie du "Compressive Sensing" (CS) est aussi utilisée afin de réaliser une localisation précise en intérieur, basée sur la mensuration de la force du signal. Le résultat de l' application de cette théorie est la réduction significative du pourcentage des informations transmises par un point d' accès sans fil (PDA, ordinateur portable etc.) avec une puissance, capacité d' enregistrement et capacité du processeur limités dans un serveur central. La propriété de la théorie du CS de donner un résultat faible concernant le processus de chiffrement est prouvé par le fait que l' approche proposé présente une résistance croissante aux intrusions potentiels des utilisateurs sans autorisation d' entrée dans le système. Enfin, un système de tracking est présenté, utilisant d' un côté l' efficacité du filtre Kalman et de l' autre côté l' efficience de la théorie du CS en combinaison avec le modèle Gaussien multi-varié, basé sur des régions.

L' évaluation expérimentale de la méthode révèle une performance croissante en ce qui concerne la localisation. En même temps, la méthode utilisée maintient une complexité computationnelle basse, comparée à d' autres méthodes statistiques d' empreinte digitales.

Mots-clés: Compressive sensing, faible densité, modèle Gaussien multi-varié, divergence Kullback-Leibler, RSS mesures, filtre Kalman

Publications

D. Milioris, G. Tzagkarakis, A. Papakonstantinou, P. Tsakalides and M. Papadopouli, "Low-dimensional Signal-Strength Fingerprint-based Positioning in Wireless LANs", in Ad Hoc Networks Journal, Elsevier (Under review).

D. Milioris, G. Tzagkarakis, P. Jacquet and P. Tsakalides, "Indoor Positioning in Wireless LANs using Compressive Sensing Signal-Strength Fingerprints", in Proc. 19th European Signal Processing Conference (EUSIPCO '11), Barcelona, Spain, August 29–September 2, 2011.

D. Milioris and P. Jacquet, "A Statistical Region-based Compressive Sensing Path-Tracking System", in Proc. 12th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc '11), Paris, France, May 16–19, 2011.

D. Milioris, L. Kriara, A. Papakonstantinou, G. Tzagkarakis, P. Tsakalides and M. Papadopouli, "Empirical Evaluation of Signal-Strength Fingerprint Positioning in Wireless LANs", in Proc. 13th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM '10), Bodrum, Turkey, October 17–21, 2010 **(2nd Best Paper Award)**.

Contents

Chapter 1

Introduction

Contents

Location estimation systems have a great potential in several distinct areas, such as in navigation, transportation, medical community, security, and entertainment. With the wide deployment of mobile wireless systems and networks, location-based services are made possible on mobile devices.

Location-sensing systems can be classified according to their dependency on and use of: (a) specialized infrastructure and hardware, (b) signal modalities, (c) training, (d) methodology and/or use of models for estimating distances, orientation, and position, (e) coordination system (absolute or relative), scale, and location description, (f) localized or remote computation, (g) mechanisms for device identification, classification, and recognition (h) accuracy and precision requirements. The distance can be estimated using time of arrival (e.g., GPS, PinPoint [35]) or signal-strength measurements, if the velocity of the signal and a signal attenuation model for the given environment, respectively, are known. Positioning systems may employ different modalities, such as, IEEE 802.11 (Radar [7, 15], Ubisense, Ekahau [2]), infrared (Active Badge [33]), ultrasonic (Cricket [25, 26], Active Bat), Bluetooth [8, 13, 27, 5, 15], 4G [28], vision (EasyLiving), and physical contact with pressure (Smart Floor), touch sensors or capacitive detectors. They may also combine multiple modalities to improve the localization, such as optical, acoustic and motion attributes (e.g.,[6]).

Most of the signal-strength based localization systems can be classified into the following two categories, namely *signature- or map*-based and *distance-prediction*-based techniques. The first type creates a signal-strength signature or map of the physical space during a training phase and compares it with the signature generated at runtime (at the unknown position) [7, 20, 34]. To build

such signatures, signal-strength data is gathered from beacons received from APs. During a training phase, at various predefined positions (of the map), such measurements are collected, and signatures are generated that associate the corresponding positions of the physical space with statistical measurements based on signal-strength values acquired at those positions. Such maps can be formed with data from different sources or signal modalities, such as ultrasound from deployed sensors to improve location-sensing [25, 15]. The distance-prediction-based techniques use the signal-strength values and radio-propagation models to predict the distance of a wireless client from an AP (or any landmark) or even between two wireless clients (peers) with estimated position (such as CLS [31]). In situations where a deployment of a wireless infrastructure may not be feasible, positioning mechanisms may exploit cooperation by enabling devices to share positioning estimates [29, 9, 21, 31, 14, 11, 12, 35]. A survey of positioning systems can be found in [17].

The IEEE 802.11 infrastructure does not require any specific hardware or installation costs. However, due to the nature of the indoor environment, transient phenomena, such as shadowing and multipath fading, lead to radio channel obstructions and variations of the RSS. Most of the fingerprint-based systems have increased computational cost. This makes the design of accurate positioning systems a difficult task and location estimation a challenging area of research.

On the other hand, the inherent sparsity of the physical space motivated in a natural fashion the use of the recently introduced theory of *compressive sensing* (CS) [68, 69] in the problem of target localization [71]. CS states that signals which are sparse or compressible in a suitable transform basis can be recovered from a highly reduced number of incoherent random projections, in contrast to the traditional methods dominated by the well-established Nyquist-Shannon sampling theory.

1.1 RSS-based WLAN Positioning Systems

The WLAN IEEE 802.11b/g is a standard used for providing wireless internet access for indoor areas. It is operated at 2.4 GHz Industrial, Scientific and Medical (ISM) band within a range of 50-100. As mentioned earlier, the RSS can be easily obtained by using any WLAN-integrated device, thus it is used by most of the WLAN positioning systems.

1.1.1 Location-Sensing Techniques

There are three major techniques to obtain the location estimate from the RSS [76, 38]. They are listed as follows:

1. Triangulation: The RSS can be translated into distance from the particular AP according to a theoretical or empirical signal propagation model. Then,

with distance measurements from at least 3 APs with known positions, lateration can be performed to estimate the locations. This approach does not give accurate estimate, as the indoor radio propagation channel is highly unpredictable and thus the use of the propagation model is not reliable.

2. Proximity: This method finds the strongest RSS from a specific AP and determines the location to be the region covered by this AP. This method only gives a very rough position estimate but it is easy to be implemented.

3. Scene Analysis: This method first collects RSS readings at known positions, which are referred to as fingerprints, in the area of interest. Then, it estimates the locations by comparing the online measurements with the fingerprints through pattern recognition techniques. This method is used by most WLAN positioning systems, as it is able to compute accurate location estimates. This is the approach used by the positioning and tracking system proposed in this thesis.

1.1.2 Existing Positioning Systems

Some WLAN positioning systems that can be accessible to the public, show that the use of fingerprinting achieves the best accuracy in indoor areas. Although the Ekahau [2] attains the best accuracy, it uses the probabilistic method to compute the estimated positions and thus requires a more comprehensive survey of RSS readings in the region of interest. In addition, its position calculation is computed at the server as the complexity of the probabilistic method is too high to be performed on the mobile devices. This raises additional issues when using these systems. First, the devices must be connected to the same network as the server to obtain position estimates. Second, positions obtained from the server must be encrypted before it is transmitted to the mobile devices, in order to protect the privacy of the users. The aim of this thesis is to design an indoor positioning and tracking system that can provide accurate position estimate with relatively low computational complexity, so that it can be computed on mobile devices. This solution may have a database server to keep track of the fingerprints database collected, and when downloaded to the devices, they are connected to the server to obtain position estimates.

1.2 Problem Statement and Objectives

A typical WLAN indoor tracking scenario consists of 1) a mobile device equipped with a WLAN adapter, which is carried by a user and collects RSS from detectable access points for localization 2) access points (APs), which can be commonly found in most buildings and their exact positions are not necessarily known to the localization systems, as they may belong to different network groups and possibly 3) a database server, which stores the fingerprints collected by the mobile device. The WLAN-enabled device can extract information, such as MAC address, SSID and received signal strength (RSS) about these APs by receiving messages broadcasted

from them. This thesis focuses on the WLAN localization and tracking problem using RSS values for positioning. The mobile device carried by the user collects the RSS from L different APs whose unique MAC addresses are used for identification. Then, the system determines the current position based on this RSS measurements and previously collected fingerprint database. The goal of this thesis is to propose a WLAN positioning and tracking system that can give accurate position estimate. In the context of this thesis, the mobile devices refer to the handheld devices (or netbooks/laptops), which have degraded WLAN antennas, limited power, memory and computation capabilities, thus a light-weight algorithm is required to allow these devices to have real-time and accurate performance.

The localization problem is defined as follow. First, the device collects online RSS readings from available APs periodically at a time interval Δt, which is limited by the device's network card and hardware performances. These online RSS readings can be denoted as $r(t) = [r_1(t), r_2(t), ..., r_L(t)], t = 0, 1, 2, ...$, where $r_l(t)$ refer to the RSS reading collected from AP l at time t. Then, the proposed positioning and tracking system uses $r(t)$ to compute the position estimate, denoted as $\hat{p}(t) = [\hat{x}(t), \hat{y}(t)]^T$, where $(\hat{x}(t), \hat{y}(t))$ are the Cartesian coordinates of the estimated position at time t.

1.3 Technical Challenges

The unpredictable variation of RSS in the indoor environment is the major technical challenge for the RSS-based WLAN positioning systems. There are three main reasons that lead to the variation of RSS.

1. First, due to the structures of the indoor environment and the presence of different obstacles, such as walls and doors or human behavior, the WLAN signals experience severe multi-path and fading and the RSS varies over time even at the same location.

2. Secondly, since the WLAN uses the licensed-free frequency band of 2.4GHz, the interference on this band can be very large. Example sources of interference are the cordless phones, BlueTooth devices and microwave. Moreover, the presence of human bodies also affects the RSS by absorbing the signals [37], as human bodies contain large amount of water, which has the same resonance frequency as the WLAN.

3. Finally, the orientation of the measuring devices also affects the RSS, as orientation of antenna affects the antenna gain and the signal is not isotropic in real indoor environment.

All of the above reasons make it infeasible to find a good radio propagation model to describe the RSS-position relationship. Thus, a fingerprinting method is often used instead to characterize the RSS-position relationship. This method computes the

position estimate by matching the online RSS readings to the fingerprints collected during training phase. This pattern matching process is a non-trivial problem as there are derivations between the online RSS readings to the fingerprint RSS readings due to the time-varying characteristics of the indoor radio propagation channel. In addition, the movement of objects, including the movement of the user who carries the mobile device, also affects the RSS readings. This type of variation of RSS is needed to be addressed by the fingerprinting-based positioning systems, in order to provide accurate position estimate.

Another challenge relates to the computational capabilities of the mobile devices. PDAs have very limited computation speed and memory when comparing to a laptop. Thus, some of the positioning systems that can be implemented on the laptop may not be able to be used by the PDA. The computational complexity and the use of memory must be taken into consideration when designing positioning and tracking systems.

1.4 Scope

In this thesis, an indoor RSS-based WLAN positioning and tracking system is proposed and implemented. Such system is able to address the challenges mentioned in the previous section. The structure of this thesis is organized as follows.

First, Chapter 2 overviews related positioning systems for mobile computing. It also describes five fingerprinting based methods: K-nearest neighbor (KNN), confidence interval, percentiles, empirical distribution and spatial sparsity, which are used as performance benchmarks to the proposed positioning system, and they are implemented in Chapter 6 to compare the performance of the proposed positioning system. In addition, it presents different ways to improve these positioning methods, such as the determination of region of interest, according to their probability densities, evaluation of the impact of the number of APs and the use of filters with inputs of previous estimate. Finally, the Compressive Sensing (CS) theory is introduced and briefly summarized in Section 2.4.

In Chapter 3, the Multivariate Gaussian (MvG) algorithm is introduced. This chapter presents how the system is operated to estimate the user's position. It describes how the construction of the database is done using the collected fingerprint by applying the algorithm during the training phase. At runtime the system compares the signature at the unknown position with the signature of each cell using the Kullback-Leibler Divergence estimation (KLD) between their corresponding probability densities. The algorithm reduces the area of interest by choosing a few regions of training cells, whose RSS readings from the database are best-matched to the online RSS readings.

Then, the CS-based positioning system is introduced in Chapter 4. The algorithm converts the localization problem into sparse signal recovery problem, so that CS theory can be applied. The inherent property of CS acting as a weak encryption process is demonstrated by showing that the proposed approach presents an increased robustness to potential intrusions of an unauthorized entity. The interactions between the mobile device and the server are also explained.

In Chapter 5, a hybrid path tracking system is presented, which exploits the efficiency of a Kalman filter in conjuction with the power of compressive sensing to represent accurately sparse signals and a region-based multivariate Gaussian model. First, the path tracking model employs a region-based multivariate Gaussian (MvG) model and then, for each region, a CS approach is applied as a refinement step.

Chapter 6 describes in detail all the experimental results conducted in three experimental sites and also on a simulated dataset. The experiments done at the FORTH-ICS, CretAquarium and INRIA-Rocquencourt in Paris, focused on the evaluation of the proposed positioning system, whereas the performance of the proposed tracking system was evaluated using the data collected from INRIA.

Finally, Chapter 7 presents the concluding remarks and gives directions for future work.

Chapter 2

Background and Related Work

Contents

2.1 Introduction

In this section, a brief overview of RSS-based WLAN positioning and tracking techniques is given. The five fingerprinting-based methods, namely KNN, confidence interval, percentiles, empirical distribution and spatial sparsity are summarized in Sections 2.3.1, 2.3.2, 2.3.3, 2.3.4 and 2.3.5, as they are implemented in Chapter 6 to compare the performance of the proposed positioning system.

There are two additional concepts used on this thesis to develop the proposed positioning and tracking system using the fingerprinting approach. Chapter 3 describes the operation of the MvG algorithm, which generates regions of interest. Section 2.4 summarizes the compressive sensing theory which can be applied on the localization problem to estimate the user's location, described extensively in Chapter 4.

2.2 Indoor RSS-based WLAN Positioning Techniques

The main problem for the indoor RSS-based positioning systems is to identify the RSS-position relationship, so that the user's location can be estimated based on the RSS collected at that location. There are two approaches in dealing with this relationship [39]: use of signal propagation models [40, 41] and the location fingerprinting methods [42, 7, 43].

2.2.1 Signal Propagation Modeling

This technique uses the RSS readings collected by the mobile device to estimate the distances of the device from at least two APs, whose locations are known, based on a signal radio propagation model. Then triangulation is used to obtain the device's position [76].

The accuracy of this technique depends heavily on finding a good model that can best describe the behavior of the radio propagation channel. However, the indoor radio propagation channel is highly unpredictable and time-varying, due to severe multipath in indoor environment, shadowing effect arising from reflection, refraction and scattering caused by obstacles, walls and interference with other devices operated at the same frequency (2.4GHz) as the IEEE 802.11b/g WLAN standard, such as cordless phones, microwaves and BlueTooth devices. There are two models that are often used for the indoor radio propagation channel:

1. The log-distance path loss model is a radio propagation model that predicts the path loss a signal encounters inside a building or densely populated areas over distance [64]. Experiments have shown that in different distance "d", the path loss PL(d) in a specific position are randomly and distributed logarithmic-normal (normal in dB) around the dependent distance mean value [65, 66]. The model is formally expressed as:

$$P_R = P_T - PL_0 - 10n \log 10\frac{d}{d_0} - X_\sigma \tag{2.1}$$

where P_R is the received power in dBm, P_T is the transmitted power in dBm, PL_0 is the path loss at the reference distance "d" (dB), d is the length of the path, d_0 is the reference distance (in our case, for indoor positioning is 1 m) and n is the path loss exponent.

X_σ is a normal (or Gaussian) random variable with zero mean, reflecting the attenuation (in dB) caused by flat fading. In case of no fading, this variable is equal to 0. In case of only shadow fading or slow fading, this random variable may have a Gaussian distribution with σ standard deviation in dB, resulting in a log-normal distribution of the received power in Watts. In case of only fast fading caused by multipath propagation, the corresponding gain in times

$F_\sigma = -10 \log 10 X_\sigma$ may be modelled as a random variable with Rayleigh or Ricean distribution.

2. The Wall Attenuation Factor model [7] includes the effects of obstacles or walls between the transmitter and receiver. The received power can be obtained by:

$$P_r = P_0 - 10\gamma log10\frac{d}{d_0} - \begin{cases} n_w \cdot WAF, \ n_w < C \\ C \cdot WAF, \ n_w \geq C \end{cases} \tag{2.2}$$

where n_w is the number of obstacles or walls between the transmitter and receiver, C is a threshold up to which no significant attenuation can be observed and WAF is the wall attenuation factor. The two empirical models require the calibration of the parameters, such as the path loss exponent, which vary depending on different environments. This often requires a comprehensive survey of the RSS distributions over the environment, which is a time consuming process. In addition, the models assume the RSS is distributed isotropically from the transmitter. This is often not the case for indoor environments due to the presence of obstacles. The orientation of the antenna of the mobile device also affects the RSS [37], but it is not reflected in the two models. Finally, the locations of the APs may not be known in the real scenario, as these APs may be installed and owned by different vendors. All of these make the models inadequate to describe the RSS-position relationship in real situation and lead to errors in estimating the user's location.

2.2.2 Location Fingerprinting

A wireless device that listens to a channel receives the beacons sent by APs (at that channel) periodically and records their RSS values. Wireless devices that run fingerprint-based positioning systems acquire such measurements and generate statistical fingerprints for a position using these measurements. The statistical-based generation of fingerprints can take place using various methods. The physical space is represented as a grid of cells with fixed size and well-known coordinates. During a training phase, at known positions of the physical space such measurements are collected by a wireless client (*training measurements*). At each position, the wireless client scans all the available channels and listens for beacons from APs. During the runtime phase, the system also records the RSS values from the received beacons (*runtime measurements*). As in the case of training, the wireless client scans all the available channels.

A statistical-based signature is constructed for each cell of the grid using the signal-strength measurements collected during the training phase (training signatures). Similarly, applying the same statistical method, at runtime, a statistical-based signature is also generated using the runtime measurements

on-the-fly (*runtime signature*). The runtime signature is then compared with all the training signatures. The cell with a training signature that has the smallest distance from the runtime signature is reported as the estimated position. The *fingerprint of a cell* is a vector of training signatures. Each entry of the vector corresponds to one AP. The fingerprint of the unknown position is the corresponding vector of the runtime signatures.

A location fingerprinting method is often used instead of the radio propagation model, as it can give better estimates of the user's locations for indoor environments. This method is divided into two phases, namely the training and runtime phases. During the training phase, which is also referred to as the offline phase, the RSS readings from different APs are collected by the WLAN-integrated mobile device at known positions, which are referred to as the training cells, to create a fingerprint database, also known as the radio map. The actual positioning takes place in the runtime phase. The mobile device, which is carried by the user collects RSS readings from different APs at an unknown position. Then, these RSS online measurements are compared to the fingerprint database to estimate the user's location by using different methods described in the next section.

The accuracy of the estimated position of the user depends highly on the number of training cells collected in the fingerprint database. If there are more training cells, then the radio map has a finer resolution and thus allows a better estimation [43]. In addition, since the RSS varies over time, collecting more time samples of RSS readings at the same training cell also improves the position estimation. Thus, this fingerprint database collection is a time consuming and labour-intensive process. [44] uses the spatial correlation of adjacent training cells to generate the database by interpolation from a small number of training cells and this method is able to reduce the labour effort and time required for the offline phase.

Another disadvantage of this fingerprinting approach is the maintenance of such databases. Since the RSS propagation environment varies with time, the accuracy of using the database degenerates over time, as the current RSS readings slowly deviate from the readings in the database. The database may even be rendered useless, if the environment changes significantly. This requires the fingerprint database to be rebuilt periodically, in order to ensure the accuracy of the positioning system. In [45] a novel method is presented to update the radio map using the online RSS readings, which can efficiently update the fingerprint database without the labour and time overhead cost as required by rebuilding such database from scratch.

As shown in [46], the RSS readings collected by different network cards are different, which can vary up to -25dBm. This indicates that the same fingerprint database cannot be used by different mobile devices, which are equipped with different WLAN network cards. That means that the fingerprint collection process must be done on each device and lead to very high labour and time costs. Another

method is to use the signal strength difference (SSD) between APs instead of the RSS as the fingerprint [47].

Although there are limitations to the location fingerprinting, it is a simple and effective method to be used by indoor positioning systems. This thesis also uses this approach to estimate the user's location.

2.3 Fingerprinting-based Positioning Methods

There are two approaches to estimate the user's location based on the online RSS measurements and the fingerprint database [48, 74]. The deterministic approach uses only the average of the RSS time samples from each training cell to estimate the location, whereas the probabilistic approach incorporates all the RSS time samples for the computation. In the following sections we will analyse the KNN, confidence interval, percentiles, empirical distribution and spatial sparsity methods.

2.3.1 K-Nearest Neighbour method (KNN)

Another common approach in location estimation problems is the use of the k-Nearest Neighbor algorithm (kNN) [74], where an RSS map is constructed by averaging separately the signal-strength values received from each AP. Let $\mu_R = [\mu_1, \ldots, \mu_P]$ be the signature vector of the unknown runtime cell c_R, where μ_i is the average RSS received from the i-th AP ($i = 1, \ldots, P$). Similarly, let $\nu_{T,c} = [\nu_1^c, \ldots, \nu_P^c]$ be the signature vector of the cell c extracted during the training phase. Then, the algorithm computes the Euclidean distance between the runtime and all the training cells, $d(c_R, c) = \|\mu_R - \nu_{T,c}\|_2$ ($c = 1, \ldots, C$), and reports the k closest neighbors by sorting the distances in increasing order. The final estimated position is given by computing the centroid of these k closest neighbors.

2.3.2 Confidence intervals

In the confidence intervals' method the signature is a vector of confidence intervals, each corresponding to an AP. Each confidence interval is generated using the RSS values of the beacons received from the corresponding AP. Let us denote as $[T_i^-(t), T_i^+(t)]$ the confidence interval for AP i at cell t during the training phase. The fingerprint of a cell is the vector of these confidence intervals (for all APs) at that cell. Similarly, at runtime, at the unknown position, the system records the RSS values from a number of beacons sent by the APs and computes a confidence interval for each AP. For example, the runtime confidence interval for AP i is the $[R_i^-, R_i^+]$. The runtime fingerprint is a vector composed by all confidence intervals formed at runtime from all APs. An AP i participates in this technique by assigning a *vote (weight) for a cell* t that indicates the similarity of its training confidence interval $[T_i^-(t), T_i^+(t)]$ with the runtime confidence interval $[R_i^-, R_i^+]$ according to

the following rule:

$$w(t) = \begin{cases} \frac{T_i^+(t) - R_i^-}{R_i^+ - T_i^-(t)} & \text{if } T_i^-(t) < R_i^- < T_i^+(t) < R_i^+ \\ \frac{R_i^+ - T_i^-(t)}{T_i^+(t) - R_i^-} & \text{if } R_i^- < T_i^-(t) < R_i^+ < T_i^+(t) \\ 1 & \text{if } R_i^- \leq T_i^-(t) < T_i^+(t) \leq R_i^+, \\ & \text{or } T_i^-(t) \leq R_i^- < R_i^+ \leq T_i^+(t) \\ 0 & \text{if } R_i^- < R_i^+ \leq T_i^-(t) < T_i^+(t), \\ & \text{or } T_i^-(t) < T_i^+(t) \leq R_i^- < R_i^+ \end{cases} \tag{2.3}$$

By adding these votes, the confidence-interval based approach computes a weight for that cell that indicates its likelihood to be the unknown position (at which the corresponding runtime measurements were collected).

At the start of the runtime phase, each cell has a zero weight. For each cell, the training confidence interval of each AP is compared with the corresponding (for that AP) runtime confidence interval at the unknown cell c. The algorithm assigns a weight at cell c, $w(c)$, that indicates the likelihood that this cell is the position of the device. Each AP participates by assigning a vote to that cell. Specifically, the weight of that cell is increased by a specific value, indicated by the following criteria: In the case that the training confidence interval is included in the runtime confidence interval or the runtime confidence interval is included in the training confidence interval, the weight of that cell is increased by one. In the case of partial overlap of these two confidence intervals, the value corresponds to the ratio of this overlap. The cell with the maximum weight is reported as the estimated position.

A main drawback of the weight as defined in (2.3) is its sensitivity to the relative position of the endpoints (boundaries) of the confidence interval. Even a small displacement of an endpoint (in the runtime confidence interval relative to the training confidence interval) may affect significantly the value of the weight. Furthermore, there are cases where the rule in (2.3) may result in equal weights between the unknown runtime cell c and two completely distinct training cells t_1, t_2. In the following, a typical example is presented for each case. For convenience, consider the simplified scenario of a single AP and the signatures of one runtime and two training cells, $c \mapsto [R^-(c), R^+(c)]$, $t_1 \mapsto [T^-(t_1), T^+(t_1)]$, $t_2 \mapsto [T^-(t_2), T^+(t_2)]$.

- ***Example 1***: the endpoints of the confidence intervals of the cells c and t_1 satisfy the following relations,
  case A: $T^-(t_1) < R^-(c) < R^+(c) < T^+(t_1)$
  case B: $T^-(t_1) < R^-(c) < T^+(t_1) < R^+(c)$ with $R^-(c) = T^+(t_1) - \varepsilon$, $R^+(c) = T^+(t_1) + \varepsilon$
  that is, the two cases differ by a small displacement of the runtime confidence interval by ε. Although the two cases are not so different with respect to the RSS measurements, however, in case A the rule (2.3) gives a weight equal to

1, while in case B the weight is equal to

$$\frac{T^+(t_1) - R^-(c)}{R^+(c) - T^-(t_1)} = \frac{T^+(t_1) - T^+(t_1) + \varepsilon}{T^+(t_1) + \varepsilon - T^-(t_1)} = \frac{\varepsilon}{T^+(t_1) - T^-(t_1) + \varepsilon} \ .$$

Clearly, when $\varepsilon \to 0$ the weight in case B is equal to 0. That is, even if the unknown runtime cell c coincides with the training t_1 a small variation of the RSS measurements may affect significantly the corresponding weight.

- ***Example 2***: the endpoints of the confidence intervals of the cells c, t_1 and t_2 are related as follows,
  $T^-(t_1) < R^-(c) < T^+(t_1) < R^+(c)$, $R^-(c) < T^-(t_2) < R^+(c) < T^+(t_2)$ with
  $|T^-(t_1) - R^-(c)| = |R^-(c) - T^-(t_2)|$
  $|R^-(c) - T^+(t_1)| = |T^-(t_2) - R^+(c)|$
  $|T^+(t_1) - R^+(c)| = |R^+(t_2) - T^+(c)|$
  By substituting in (2.3), the corresponding weights, that is, the values of the first two ratios are equal, and thus the confidence interval method would assign the same weight to both training cells, being unable to distinguish between t_1 and t_2.

2.3.3 Percentiles

This approach is similar to the confidence-interval one. However, instead of using confidence intervals for constructing the fingerprints, percentiles are employed. A set of percentiles can capture more detailed information about the signal strength distribution than confidence intervals, and thus, resulting to more accurate fingerprints. The weight of a cell c, w(c), is computed as follows:

$$w(c) = \sum_{i=1}^{N} \sqrt{\sum_{j=1}^{p} (R_j^i - T_j^i(c))^2} \tag{2.4}$$

where N is the number of APs, p the number of percentiles, R_j^i the j-th percentile of runtime measurements from the i-th AP and $T_j^i(c)$ the j-th percentile using the training measurements from the i-th AP at the cell c.
As in the confidence interval case, the cell with the maximum weight is reported as the estimated position. In the case of the *top 5 weighted percentiles* approach, the centroid of the top five cells with the largest weight is reported as the estimated position.

2.3.4 Empirical distribution

The signature of a cell is a vector of size equal to the number of APs that appear in both the training and runtime measurements. Each entry of a training (runtime) signature corresponds to the complete set of RSS values collected during the training (runtime) phase, respectively. This method creates a signature based on the set of signal-strength measurements collected at each cell from all APs. At runtime, at an

unknown position, each cell is assigned a *weight* which corresponds to the *average* empirical KLD distance of each AP (at that cell) from the runtime measurements collected at the unknown position from the same AP. The cell with the smallest weight is reported as the position.

2.3.5 Spatial sparsity

In a recent work [75], the problem of location estimation was treated in a framework that also takes advantage of the spatial sparsity. In particular, the location estimation is formulated as a constrained ℓ_1-norm minimization problem based on a suitably learned dictionary. A signature is associated to each AP by averaging the RSS measurements which would be received by the AP from each potential cell of the discrete spatial domain. Then, the system builds the dictionary by concatenating the signatures from all APs. A similar signature is generated at the unknown runtime cell, which is then projected on the dictionary to form the vector of measurements. However, the lack of a random measurement matrix required when working in the framework of CS may decrease the system's performance under unpredictable environmental conditions, while also the communication of the projected measurements from the wireless device to the APs, where the localization takes place, could pose undesired security issues.

2.4 Compressive Sensing Theory

Compressive sensing theory allows compressible signals to be recovered by fewer samples than traditional methods, which according to the Nyquist sampling theory requires the sampling rate to be at least twice the maximum bandwidth. This is possible when signals of interest are sparse and are sampled incoherently. The compressive sensing problem can be formulated as follow [67]:

Consider a discrete-time signal x as a $N \times 1$ vector in $\mathbb{R}^N$. Such signal can be represented as a linear combination of a set of basis $\{\psi_i\}_{i=1}^N$. Constructing a $N \times N$ basis matrix $\mathbf{\Psi} = [\psi_1, \psi_2, \ldots, \psi_N]$, the signal x can be expressed as

$$x = \sum_{i=1}^{N} s_i \psi_i = \mathbf{\Psi} s \tag{2.5}$$

where s is a $N \times 1$ vector and is an equivalent representation of x in a basis Ψ. A signal is K-sparse when it can be represented as a linear combination of $K << N$ basis vectors. This means that there is only K nonzero entries for the vector s.

The overall compressive sensing measurement model can be expressed as

$$y = \mathbf{\Phi} x = \mathbf{\Phi}\mathbf{\Psi} s = \mathbf{\Theta} s \tag{2.6}$$

where $\mathbf{\Phi}$ is a $M \times N$, $M < N$ measurement sensing matrix for sensing the signal x, $\mathbf{\Theta} = \mathbf{\Phi}\mathbf{\Psi}$ is an $M \times N$ matrix, and y is a $M \times 1$ observation vector collected as a result of this sensing process. This problem can be referred to as incoherent sampling

if the largest correlation between the sensing matrix $\mathbf{\Phi}$ and the representation basis $\mathbf{\Psi}$,

$$\mu(\mathbf{\Phi}, \mathbf{\Psi}) = \sqrt{N} \cdot \max_{1 \leq i,j \leq N} | < \phi_i, \psi_j > | \tag{2.7}$$

is small. Compressive sensing theory requires both the sparsity and incoherent sampling, so that the signal can be recovered exactly with high probability. If $M \geq cK \log(N/K) << N$, where c is a small constant, the signal can be reconstructed by solving the following ℓ_1 norm minimization problem:

$$\hat{s} = arg \min_{s \in \mathbb{R}^N} ||s||_1 \ s.t. \ \mathbf{\Theta} s = y. \tag{2.8}$$

This is a convex optimization problem that can be easily converted into a linear program, known as basis pursuit, through primal-dual method [69, 68]. Additional algorithms to solve this optimization problem can also be found in [68].

2.5 Indoor Positioning and Tracking

Most of the indoor tracking methods use past position estimates and motion dynamics to refine the current position estimate determined by the above positioning methods. In addition, the dynamic motion model can also be used in conjunction with the current position estimate to predict the future possible locations. The motion dynamics can be modeled by a general Bayesian tracking model and a filter is then derived to refine the position estimates [49]. There are two filters that are used commonly to improve the accuracy of positioning systems, Kalman filter and Particle filter.

2.5.1 Kalman filter

By assuming the Gaussian tracking noise model and linear motion dynamics, the general filter becomes a Kalman filter, whose optimal solution is a minimum mean square error (MMSE) estimate. Although the assumption of Gaussian RSS-position relationship is not often the case [37], the application of the Kalman filter as the post-processing step is able to improve the accuracy of the positioning systems [49, 50, 51, 52]. The parameters of the Kalman filter are needed to be found experimentally. In [53] are provided some guidelines on how to set the parameters for each update steps based on the map information.

2.5.2 Particle filter

The particle filter is a sequential Monte Carlo method that generates random samples, known as particles, according to a motion models and estimates their probability densities [54, 55]. Unlike the Kalman filter, the particle filter can be applied on non-Gaussian and non-linear models. In addition, map information can be used to further improve the performance of the particle filter by assigning zero

weights to the invalid particles, such as those across the wall [56, 57]. Backtracking based on the map information is also proposed in [58]. Moreover, information obtained from accelerometers and inertial measurement units (IMU) can also be used to refine the motion models and let the filter to generate particles that are more relevant and hence improve the tracking accuracy [59, 60].

However, the major drawback of the particle filter is its high computation complexity. For example, 1600 particles are needed for each filter update for a $40m \times 40m$ experimental area to achieve the best performance [57]. This large computation workload can not be handled by the mobile devices to give real-time updates to the user. Hence, this thesis chooses the Kalman filter to post-process the estimates instead of the particle filter, which may severely hinder the operations of the mobile devices.

2.5.3 Other fingerprinting methods

Besides the use of the above filters, several other methods are also used for the indoor positioning and path tracking. Recently significant work has been published in the area of location-sensing using RF signals. Like CLS, Radar [7] employs signal-strength maps that integrate signal-strength measurements acquired during the training phase from APs at different positions with the physical coordinates of each position. Each measured signal-strength vector is compared against the reference map and the coordinates of the best match will be reported as the estimated position. Bahl *et al.* [23] improved Radar to alleviate side effects that are inherent properties of the signal-strength nature, such as aliasing and multipath.

Ladd *et al.* [20] proposed another location-sensing algorithm that utilizes the IEEE 802.11 infrastructure. In its first step, a host employs a probabilistic model to compute the conditional probability of its location for a number of different locations, based on the received signal-strength measurements from nine APs. The second step exploits the limited maximum speed of mobile users to refine the results and reject solutions with a significant change in the location of the mobile host.

Kung *et al.* [19] propose a method for evaluating the impact of the IEEE 802.11 APs on positioning in order to strengthen the role/contribution of a “good" AP while the emphasizing the role of the “bad" APs. The “goodness" of an AP indicates the capability of that AP to estimate accurately its distance from the others.

Youssef *et al.* [34] substantially improved the accuracy (e.g., reporting an 1.3m error in 90% in their experiments conducted in their department) by employing an autoregressive model that captures the autocorrelation in signal strength measurements of the same AP at a particular location. Specifically, the time series generated from signal strength measurements collected from an AP is represented by a first-order autoregressive model. The fingerprints are formed for

each cell and AP, including three values, namely the degree of autocorrelation, the mean and the variance of the empirical measurements collected from that AP at that cell. Finally, an interesting approach propose fingerprints based on attributes that characterize the effects of multipath (e.g., channel response) in order to detect changes of the positions of wireless hosts were presented in [24, 36].

Niculescu and Badri Nath [22] designed and evaluated a cooperative location-sensing system that uses specialized hardware for calculating the angle between two hosts in an ad-hoc network. This can be done through antenna arrays or ultrasound receivers. Hosts gather data, estimate their position, and propagate them throughout the network. Previously, these authors [21] introduced a cooperative location-sensing system in which position information of landmarks is propagated towards hosts that are further away, while at the same time, closer hosts enrich this information by determining their own location. Another location-sensing system in ad-hoc networks performs positioning without the use of landmarks or GPS and presents the tradeoffs among internal parameters of the system [9]. The location-sensing systems presented in [29] and [16] are the closest to CLS and are compared in detail in [14].

Active Badge [32] uses diffuse infrared technology and requires each person to wear a small infrared badge that emits a globally unique identifier every ten seconds or on demand. A central server collects this data from fixed infrared sensors around the building, aggregates it and provides an application programming interface for using the data. The system suffers in the case of fluorescent lighting and direct sunlight, because of the spurious infrared emissions these light sources generate. A different approach, SmartFloor [3], employs a pressure sensor grid installed in all floors to determine presence information. It can determine positions in a building without requiring users to wear tags or carry devices. However, it is not able to specifically identify individuals.

Examples of localization systems that combine multiple technologies are UbiSense [4], Active Bats [1] and SurroundSense [6]. UbiSense can provide a high accuracy using a network of ultra wide band (UWB) sensors installed and connected into a building existing network. The UWB sensors use Ethernet for timing and synchronization. They detect and react to the position of tags based on time difference of arrival and angle of arrival. An RFtag is a silicon chip that emits an electronic signal in the presence of the energy field created by a reader device in proximity. Location can be deduced by considering the last reader to see the card. RFID proximity cards are in widespread use, especially in access control systems. The Active Bats architecture consists of a controller that sends a radio signal and a synchronized reset signal simultaneously to the ceiling sensors using a wired serial network. Bats respond to the radio request with an ultrasonic beacon. Ceiling sensors measure time-of-flight from reset to ultrasonic pulse. Active Bat applies statistical pruning to eliminate erroneous sensor measurements caused by a sensor

hearing a reflected pulse instead of one that travelled along the direct path from the Bat to the sensor. A relatively dense deployment of ultrasound sensors in the ceiling can provide within 9 cm of the true position for 95% of the measurements. SurroundSense runs on a mobile phone to provide logical localization by generating fingerprints using sound, accelerometers, cameras and IEEE 802.11. Tesoriero *et al.* [30] propose a passive RFID-based indoor location system that is able to accurately locate autonomous entities, such as robots and people, within a physical-space.

Ariadne [18] is an automated location determination system. It uses a two dimensional construction floor plan and only a single actual strength measurement. It generates an estimated signal strength map comparable to those generated manually by actual measurements. Given the signal measurements for a mobile, a proposed clustering algorithm searches that signal strength map to determine the current mobile's location.

In a recent work [73], the physical space is discretized as a grid consisting of cells with known coordinates. Then, a statistical signature is extracted for each cell in the training phase by modeling the RSS values received from a set of APs using a multivariate Gaussian distribution (MvG). In the runtime phase, a similar statistical signature is generated at the unknown position, which is then compared with the training signatures by means of a statistical similarity measure, namely, the Kullback-Leibler divergence (KLD). The major benefit of the MvG proposed algorithm, when compared with other methods, is the significantly reduced computational complexity and implementation simplicity, as well as the high accuracy in several specific environments (obstacle-free, robust measurements) as it was revealed by the experimental evaluation.

The problem of indoor location estimation is also treated in a probabilistic framework using Hidden Markoc Models [10]. In particular, a reduced number of locations sampled to construct a radio map is employed in conjunction with an interpolation method, which is developed to effectively patch the radio map. Furthermore, a Hidden Markov Model (HMM) that exploits the user traces to compensate for the loss of accuracy is employed to achieve further improvement of the radio map due to motion constraints, which could confine possible location changes. Both the proposed multivariate Gaussian model-based algorithm [73] and the HMM-based approach belong to the class of the probabilistic localization techniques. Usually, a probabilistic localization method is characterized by an increased performance when compared with a deterministic one, since it provides not only a point estimate of the user's position but also gives a confidence interval for the quality of this estimate. This can be used to improve further the estimation accuracy with the goal of reducing the uncertainty. However, a first key observation is the highly reduced complexity of the MvG method [73] compared to the HMM-based approach. In particular, it is a one-iteration method, where in each iteration only the simple estimate of a mean vector, a covariance matrix,

and the computation of the Kullback-Leibler divergence between multivariate Gaussians (given in closed form) are required. On the other hand, the HMM-based localization technique requires several iterations to converge, while in each iteration several model parameters have to be estimated (approximately of the same dimensions as the parameters of our proposed method). However, the reduced computational complexity of the Gaussian-based technique comes at the cost of a potentially degraded location estimate under certain circumstances. For instance, in the case of corrupted measurements (e.g., due to an access point failure or the presence of an obstacle) the MvG method [73] is much more sensitive, since it is based on measurements collected instantaneously. In contrast, the HMM-based approach could provide a more accurate estimate via the prior knowledge of a transition-probability matrix, which is preserved and re-estimated in each iteration in conjunction with the refinement achieved by an Expectation Maximization algorithm. In conclusion, the major benefit of the MvG proposed algorithm [73], when compared with the HMM-based approach, is the significantly reduced computational complexity and implementation simplicity, as well as the high accuracy in several specific environments (obstacle-free, robust measurements) as it was revealed by the experimental evaluation. On the other hand, the HMM-based approach can be proved to be more robust in the case of system failures, but at the cost of requiring increased computational resources.

Liao et al. proposed a method to predict the user's orientation, which is then used for the next position estimate to improve the accuracy, from the previously computed location estimates [61]. A Viterbi-like algorithm, which is developed to enhance the RADAR system and is also implemented by [62], makes use of historical data based on the KNN method to determine the location estimates. Finally, a nonparametric information filter based on the kernel-based probabilistic method is proposed in [63]. This filter, whose computational complexity is lower than particle filter, is able to deal with tracking scenarios where Kalman filter is inapplicable.

2.6 Chapter Summary

This chapter gives a brief overview of different methods developed for the RSS-based WLAN indoor positioning systems. Five fingerprinting methods, KNN, confidence interval, percentiles, empirical distribution and spatial sparsity are described in details, as they are served as the performance benchmarks for the proposed positioning system. Moreover, several indoor tracking techniques that are able to improve the accuracy through the use of previous estimates and motion models are also discussed. Compressive Sensing theory is also introduced.

Chapter 3

Multivariate Gaussian Based Positioning System

Contents

3.1 Introduction

The vast majority of current fingerprint positioning methods does not take into account the interdependencies among the RSS measurements at a certain position from the various APs. These interdependencies provide important information about the geometry of the environment and can be quantified using the second-order spatial correlations among the measurements. Hence, the employment of multi-dimensional distributions is expected to provide a more accurate representation of the RSS profiles, leading to improved positioning performance. Of course, simple models should be used whose parameters (second-order statistics) could be accurately and easily estimated in a practical positioning scenario.
Based on this observation, a novel fingerprint approach has been designed and evaluated. Specifically, it makes two distinct contributions:

1. It proposes and evaluates a novel fingerprinting approach that exploits the spatial correlations of signal-strength measurements collected from various wireless APs based on a *multivariate Gaussian* model.

2. It performs a comparative performance analysis of various signal-strength fingerprinting methods and Ekahau in the premises of two research laboratories and an aquarium under different conditions, as it will be shown in Chapter 6.

3.2 Multivariate Gaussian Model and Statistical Similarity Measure

The multivariate Gaussian-based approach takes into consideration not only the signal strength measurements from each AP but also the interplay (covariance) of measurements collected from pairs of APs. As in other fingerprinting approaches, the physical space is discretized in a grid consisting of cells with known coordinates. During a training phase, signal-strength measurements are collected from the IEEE 802.11 APs at known positions of the physical space. The signature of each position is formed using a multivariate Gaussian distribution. At runtime, at the unknown position, the system creates a signature based on the signal-strength measurements collected by the wireless device on-the-fly, which is then compared with the signature of each cell, constructed using the training measurements.

The signature comparison and position estimation is based on the Kullback-Leibler divergence (KLD): the cell corresponding to the minimum KLD is reported as the estimated position. The method generalizes this approach by applying it iteratively in different spatial scales.

3.3 Positioning based on Statistical Signatures

Unlike other fingerprint positioning methods, this one focuses on the interdependencies among the RSS measurements in a cell from various APs. These interdependencies provide information about the geometry/topology of the environment and can be quantified using the second-order spatial correlations among the measurements. According to this proposed approach, in the training phase, a statistical signature is extracted for each cell of the grid by modelling the acquired signal-strength measurements using a multivariate Gaussian distribution. The density function of a multivariate Gaussian in $\mathbb{R}^K$, with a mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, is given by:

$$p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{K/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\Big(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\Big) \ , \tag{3.1}$$

where $|\boldsymbol{\Sigma}|$ is the determinant of $\boldsymbol{\Sigma}$.

Let P be the number of APs from which the mobile device receives the measurements, M be the number of measurements from each AP, and $\mathbf{S}_i = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_P]$ denote the $M \times P$ matrix for the i-th cell c_i, whose j-th column $\boldsymbol{y}_j \in \mathbb{R}^M$ contains the received signal-strength values from the j-th AP. The signal-strength measurements are modelled by a multivariate Gaussian distribution due to its simplicity and the closed-form expression of the associated similarity measure (KLD). More specifically, the signature $\mathcal{S}_i$ of the i-th cell is given by:

$$c_i \mapsto \mathcal{S}_i = \{\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\} \ , \tag{3.2}$$

where $\boldsymbol{\mu}_i = [\mu_{i,1}, \ldots, \mu_{i,P}]$, with $\mu_{i,j}$ being the mean of the j-th column of the measurement matrix $\mathbf{S}_i$, and $\boldsymbol{\Sigma}_i$ is the corresponding covariance matrix, with its mn-th element being equal to the covariance between the m-th and n-th columns of $\mathbf{S}_i$. Hence, the mn-th element of $\boldsymbol{\Sigma}_i$ corresponds to the *spatial correlation* between the RSS measurements of the i-th cell received from the $m-$th and n-th APs. Thus, if C is the number of cells in the grid representing the physical space, during the training phase, the following set of *training* signatures (T) is generated:

$$\{\mathcal{S}_{i,T}\}_{i=1}^{C} = \big\{\{\boldsymbol{\mu}_{i,T}, \boldsymbol{\Sigma}_{i,T}\}\big\}_{i=1}^{C} \ . \tag{3.3}$$

In addition, the i-th training cell, $c_{i,T}$, is also associated to a set of indices $I_{i,T}$ indicating its corresponding "active" APs, that is, the APs from which it acquires the measurements during the training phase.

During the *runtime* phase (R), we assume that the mobile user is placed at an unknown cell (c_R), whose location must be estimated. Following the approach used in the training phase, if $\mathbf{S}_R = [\boldsymbol{y}_{1,R}, \ldots, \boldsymbol{y}_{P',R}]$ is the $M' \times P'$ runtime signal-strength measurement matrix of c_R, a signature is generated as follows,

$$c_R \mapsto \mathcal{S}_R = \{\boldsymbol{\mu}_R, \boldsymbol{\Sigma}_R\} \ . \tag{3.4}$$

Notice here that in general the dimensions of the runtime measurement matrix are smaller than the dimensions of the corresponding training matrix ($M \times P$). This is due to the fact that during runtime it is more difficult to collect extensive measurements than during training. Furthermore, the set of APs operating during the training phase is *not* necessarily the same with the set of APs at runtime.

Let us denote as $I_R^{i,T}$ the set of APs from which signal-strength measurements were collected *both at runtime and training* at cell i. For the runtime (c_R) and the i-th training cell ($c_{i,T}$), we extract their corresponding mean sub-vectors $\boldsymbol{\mu}_R^s$, $\boldsymbol{\mu}_{i,T}^s$ and covariance sub-matrices $\boldsymbol{\Sigma}_R^s$, $\boldsymbol{\Sigma}_{i,T}^s$ according to the indices of $I_R^{i,T}$. Finally, if $p_R(\boldsymbol{x}|\boldsymbol{\mu}_R^s, \boldsymbol{\Sigma}_R^s)$ and $p_{i,T}(\boldsymbol{x}|\boldsymbol{\mu}_{i,T}^s, \boldsymbol{\Sigma}_{i,T}^s)$ denote the multivariate Gaussian densities of c_R and $c_{i,T}$, respectively, their KLD is given by the following closed-form expression:

$$\begin{aligned} D(p_R||p_{i,T}) \ = \ & \frac{1}{2}\Bigg((\boldsymbol{\mu}_{i,T}^s - \boldsymbol{\mu}_R^s)^T (\boldsymbol{\Sigma}_{i,T}^s)^{-1} (\boldsymbol{\mu}_{i,T}^s - \boldsymbol{\mu}_R^s) \\ & + \ tr\big(\boldsymbol{\Sigma}_R^s (\boldsymbol{\Sigma}_{i,T}^s)^{-1} - \mathbf{I}\big) - \ln |\boldsymbol{\Sigma}_R^s (\boldsymbol{\Sigma}_{i,T}^s)^{-1}| \Bigg) \ , \end{aligned} \tag{3.5}$$

where $tr(\cdot)$ denotes the trace of a matrix (sum of its diagonal elements) and $\mathbf{I}$ is the identity matrix. KLD is a (non-symmetric) measure of the difference between two probability distributions, well established and widely used in probability and information theory. The estimated location $[x_R^*, y_R^*]$ is given by the coordinates of the i^*-th cell, which minimizes (3.5), that is,

$$i^* = \arg \min_{i=1,\ldots,C} D(p_R||p_{i,T}) \ . \tag{3.6}$$

Algorithm 1 The multivariate Gaussian-based positioning method (spatial scale of a cell)

1. During training phase, collect RSS measurements from APs at each cell
   trainingAP(c): set of APs from which data are collected at cell c
2. During runtime, collect RSS measurements from each AP at the unknown position
   runtimeAPs: set of APs from data are collected
   $effectiveAP(c) : trainingAP(c) \cap runtimeAP$
3. During runtime, perform the following steps for each cell c:
   - Generate the signature for cell c using only training measurements collected from APs $\in$ effectiveAP(c) (*i.e., training signature(c)*)
   - Generate the runtime signature using only runtime measurements collected from APs *in* effectiveAP(c) (*i.e., runtime signature(c)*)
   - Estimate the KLD distance of the training and runtime signatures
4. Report as the estimated position the cell c^* with the smallest KLD distance

3.3.1 Multiscale spatial aggregation of fingerprints

To improve the accuracy, we propose a generalization of the approach presented in Algorithm 1: instead of applying the multivariate Gaussian *per cell*, we apply it in an iterative fashion in multiple spatial scales (*e.g., regions*). First, the physical space is divided into overlapping regions of equal size and the multivariate Gaussian algorithm is applied for each region separately. To generate the fingerprint of a region, we employ all the signal-strength measurements from all APs collected at positions within that region. This *spatial aggregation* reduces the likelihood of selecting a false region/cell (a region/cell that does not include/correspond to the actual position) over the correct one. Essentially, via this aggregation an incorrect region is eliminated (in the first iteration) while the “weight" of the correct region is enhanced by considering the signatures of the neighboring to the actual position cells. The region-based multivariate Gaussian algorithm proceeds iteratively: after it estimates the region at which the device is located, it repeats the process by dividing the selected region into sub-regions and applying the algorithm on them. There have been considered only two spatial scales: i) in the coarse granularity the area is divided into regions and ii) in the second (fine) level granularity the area reduces to a single cell.

The original area of interest is discretized in G regions, each of N cells. Let A_i be a $GK \times N$ matrix whose j-th column ($\in \mathbb{R}^{GK}$) contains the received signal-strength values from the j-th AP collected at cells of region i during training.

Let us denote with $A_{i,T}(\boldsymbol{x}|\boldsymbol{\mu}^s_{i,T}, \boldsymbol{\Sigma}^s_{i,T})$ the multivariate Gaussian density of region i and $p_R(\boldsymbol{x}|\boldsymbol{\mu}^s_R, \boldsymbol{\Sigma}^s_R)$ the multivariate Gaussian density of the unknown position (runtime signature). The KLD distance can be computed as in (3.5) and the region closest to the unknown position is given by

$$i^*_A = \arg \min_{i=1,\ldots,G} D(p_R||A_{i,T}) \ . \tag{3.7}$$

After the estimation of the correct region, the process is repeated (using Algorithm 1) to compute the cell in that region that corresponds to the unknown position (considering only the cells of that region). Finally, Table 3.1 gathers the symbols used in the MvG setup.

Table 3.1: MvG symbols and notation

N	Number of APs
S_i	$K \times N$ matrix for the i-th cell
c_i	i-th cell
μ_i	Mean values of the APs of the i-th cell
Σ_i	Covariance matrix of the i-th cell
L	Number of training cells
T	Set of training signanutres
$I_{i,T}$	Active APs of the i-th cell
R	Runtime phase
c_R	Cell to be estimated
S_R	$K' \times N'$ run time measurement matrix
$I_R^{i,T}$	Set of the common APs
$\vec{\mu}^s_R$	Mean sub-vector of c_R
$\vec{\mu}^s_{i,T}$	Mean sub-vector of $c_{i,T}$
$\boldsymbol{\Sigma}^s_R$	Covariance sub-matrix of c_R
$\boldsymbol{\Sigma}^s_{i,T}$	Covariance sub-matrix of $c_{i,T}$
$p_R(\vec{x}\|\vec{\mu}^s_R, \boldsymbol{\Sigma}^s_R)$	Multivariate Gaussian of c_R
$p_{i,T}(\vec{x}\|\vec{\mu}^s_{i,T}, \boldsymbol{\Sigma}^s_{i,T})$	Multivariate Gaussian of $c_{i,T}$
i^*	Estimated location

3.4 Chapter Summary

This chapter introduced a novel localization method that creates signal-strength fingerprints using multivariate Gaussian distributions. It estimates the position of the device by computing the region with the training fingerprint that has the closest KLD distance from the runtime fingerprint. Furthermore, in the case of the multivariate Gaussian-based algorithm we experimented with a multiple spatial scale iterative approach in which, we applied the algorithm on larger regions, to

select the correct one, and then within the selected region to estimate the correct cell.

Chapter 4

Compressive Sensing Based Positioning System

Contents

4.1 Introduction

Due to the unpredictable nature of the RSS distribution at indoor environment, most of the indoor RSS-based WLAN positioning systems use the fingerprinting approach to acquire the explicit RSS and position relationship, in order to compute a more accurate estimation of user's position.

4.2 Positioning and Inherent Spatial Sparsity

The compressive sensing based positioning system proposed in this chapter is also a fingerprinting method. Unlike traditional fingerprinting systems, the proposed system reformulates the localization problem into a sparse-natured problem and thus the compressive sensing concept can be applied to find the estimated positions.

4.3 CS-based Positioning System

In the following, let $\mathbf{x} \in \mathbb{R}^N$ denote the signal of interest, that is, a vector of RSS measurements. The efficiency of a compressive sensing (CS) method for signal approximation or reconstruction depends highly on the signal's sparsity structure in a suitable transform domain associated with an appropriate sparsifying basis $\mathbf{\Psi} \in \mathbb{R}^{N \times D}$. In terms of signal approximation it has been demonstrated [68, 69]

that if $\mathbf{x}$ is K-sparse in $\mathbf{\Psi}$ (meaning that the signal is exactly or approximately represented by K elements of this basis), then it can be reconstructed from $M = rK \ll N$ non-adaptive linear projections onto a second measurement basis, which is incoherent[1] with the sparsity basis, and where $r > 1$ is a small overmeasuring factor. For instance, in standard signal processing applications, several natural signals are often sparse in a discrete cosine transform (DCT) or in a Fourier basis.

The measurement model in the original space-domain is expressed as follows,

$$\mathbf{g} = \mathbf{\Phi x} \ , \tag{4.1}$$

where $\mathbf{g} \in \mathbb{R}^M$ is the measurement vector and $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$ denotes the measurement matrix. By noting that $\mathbf{x}$ can be expressed in terms of the basis $\mathbf{\Psi}$ as $\mathbf{x} = \mathbf{\Psi w}$, where $\mathbf{w} \in \mathbb{R}^D$ denotes the vector of transform coefficients, the measurement model has the following equivalent transform-domain representation

$$\mathbf{g} = \mathbf{\Phi\Psi w} \ . \tag{4.2}$$

Examples of measurement matrices $\mathbf{\Phi}$, which are incoherent with any fixed transform basis $\mathbf{\Psi}$ with high probability (universality property [69]), are random matrices with independent and identically distributed (i.i.d.) Gaussian or Bernoulli entries.

Since the original vectors of RSS measurements, $\mathbf{x}$, are not sparse in general, in the following study we focus on the more general case of reconstructing their equivalent sparse representations, $\mathbf{w}$, given a low-dimensional set of measurements $\mathbf{g}$ and the measurement matrix $\mathbf{\Phi}$.

The inherent sparsity in the problem of location estimation comes from the fact that the device to be localized can be placed in exactly one of the C non-overlapping cells. Let $\mathbf{w} = [0\ 0\ \cdots\ 0\ 1\ 0\ \cdots\ 0]^T \in \mathbb{R}^C$ be an indicator vector with its j-th component being equal to "1" if the device is located in the j-th cell. Thus, in the framework of CS, the problem of estimating the location of a mobile device is reduced to a problem of recovering the one-sparse vector $\mathbf{w}$. Of course in practice we do not expect an exact sparsity, thus, the estimated position corresponds simply to the largest-amplitude component of $\mathbf{w}$.

By employing the M compressive measurements and given the K-sparsity property in basis $\mathbf{\Psi}$, the sparse vector $\mathbf{w}$, and consequently the original signal $\mathbf{x}$, can be recovered perfectly with high probability by taking a number of different approaches. In the case of noiseless CS measurements the sparse vector $\mathbf{w}$ is

[1] Two bases $\mathbf{\Psi}_1$, $\mathbf{\Psi}_2$ are incoherent if the elements of the first are not represented sparsely by the elements of the second, and vice versa.

estimated by solving a constrained ℓ_0-norm optimization problem of the form,

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \|\mathbf{w}\|_0, \quad s.t. \;\; \mathbf{g} = \mathbf{\Phi}\mathbf{\Psi}\mathbf{w} \;, \tag{4.3}$$

where $\|\mathbf{w}\|_0$ denotes the ℓ_0 norm of the vector $\mathbf{w}$, which is defined as the number of its non-zero components. However, it has been proven that this is an NP-complete problem, and the optimization problem can be solved in practice by means of a relaxation process that replaces the ℓ_0 with the ℓ_1 norm,

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \|\mathbf{w}\|_1, \quad s.t. \;\; \mathbf{g} = \mathbf{\Phi}\mathbf{\Psi}\mathbf{w} \;. \tag{4.4}$$

In [68, 69] it was shown that these two problems are equivalent when certain conditions are satisfied by the two matrices $\mathbf{\Phi}$, $\mathbf{\Psi}$ (restricted isometry property (RIP)). In the later case, the sparse vector $\mathbf{w}$ can be recovered using $M \gtrsim K \cdot \log D$ CS measurements.

The objective function and the constraint in (4.4) can be combined into a single objective function, and several of the most commonly used CS reconstruction methods solve the following problem,

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \Big(\|\mathbf{w}\|_1 + \tau \|\mathbf{g} - \mathbf{\Phi}\mathbf{\Psi}\mathbf{w}\|_2 \Big) \;, \tag{4.5}$$

where τ is a regularization factor that controls the trade-off between the achieved sparsity (first term in (4.5)) and the reconstruction error (second term). Commonly used algorithms are based on linear programming [70], convex relaxation [68, 77], and greedy strategies (*e.g.,* Orthogonal Matching Pursuit (OMP) [78, 79]).

As it was mentioned before, a common characteristic of all RSS-based fingerprint methods is their implementation in two distinct phases, namely, a *training phase* (off-line), where the central server is mainly involved, and a *runtime phase* (on-line), which concerns the wireless device to be localized. In the following two subsections the several requirements of each phase are described in detail. Besides, for convenience the following notations are used in the subsequent derivations: i) y_T denotes any quantity y that is related with the training phase, ii) y_R denotes that y is associated with the runtime phase.

4.3.1 Training phase specifications

During the training phase, a set of RSS samples is collected at each cell from each AP. Let $\boldsymbol{x}_{j,T}^{i} \in \mathbb{R}^{n_{j,i}}$ denote the vector of training RSS measurements received at cell j from AP i. In general $n_{j,i} \neq n_{j',i'}$ for $j \neq j'$, $i \neq i'$. To compensate for the potentially different number of RSS measurements from cell to cell we set $N_i = \min_j\{n_{j,i}\}$, $i = 1, \ldots, P$, $j = 1, \ldots, C$. Then, these vectors are collected from all cells by a central server, which forms a single matrix $\mathbf{\Psi}_T^i \in \mathbb{R}^{N_i \times C}$ for the i-th AP by concatenating the corresponding C vectors. Then, this matrix

is used as the appropriate sparsifying dictionary for the i-th AP, since in the ideal case the vector of RSS measurements at a given cell j received from AP i should be closer to the corresponding vectors of its neighboring cells, and thus it could be expressed as a linear combination of a small subset of the columns of $\mathbf{\Psi}_T^i$.

Moreover, a measurement matrix $\mathbf{\Phi}_T^i \in \mathbb{R}^{M_i \times N_i}$ must be associated with each transform matrix $\mathbf{\Psi}_T^i$, where M_i is the number of CS measurements. In the proposed algorithm, a standard Gaussian measurement matrix is employed, with its columns being normalized to unit ℓ_2 norm.

4.3.2 Runtime phase specifications

A similar process is followed during the runtime phase. More specifically, let $\boldsymbol{x}_{c,R}^i \in \mathbb{R}^{n_{c,i}}$ be the RSS measurements received at the current unknown cell c from the i-th AP. Notice that, since the acquisition time interval during the runtime phase is smaller than that in the training phase, it holds that $n_{c,i} < n'_{c,i}$, where $n'_{c,i}$ denotes the length of the corresponding RSS vector generated at the same cell during the training phase. The runtime CS measurement model associated with the cell c and AP i is written as

$$\mathbf{g}_{c,i} = \mathbf{\Phi}_R^i \boldsymbol{x}_{c,R}^i \ , \tag{4.6}$$

where $\mathbf{\Phi}_R^i \in \mathbb{R}^{M_{c,i} \times n_{c,i}}$ denotes the corresponding measurement matrix during the runtime phase.

In order to overcome the problem of the difference in dimensionality between the training and runtime phase, while maintaining the robustness of the reconstruction procedure, we select $\mathbf{\Phi}_R^i$ to be a subset of $\mathbf{\Phi}_T^i$ with an appropriate number of rows such as to maintain equal measurement ratios, $\frac{M_i}{N_i} = \frac{M_{c,i}}{n_{c,i}}$. Then, the measurement vector $\mathbf{g}_{c,i}$ is formed for each AP i according to (4.6) and transmitted to the server, where the reconstruction takes place via the solution of (4.5), with the training matrix $\mathbf{\Psi}_T^i$ being used as the appropriate sparsifying dictionary. We emphasize at this point the significant conservation of the processing and bandwidth resources of the wireless device, by computing only low-dimensional matrix-vector products to form $\mathbf{g}_{c,i}$ $(i = 1, \ldots, P)$ and then transmitting a highly reduced amount of data ($M_{c,i} \ll n_{c,i}$). Then, the CS reconstruction is performed at the server for each AP independently and the final location estimate is the centroid of the estimated cells. The reason for carrying out the location estimation in a disjoint fashion among the APs is that, due to the network configuration, the RSS values are received independently.

The experimental evaluation presented Chapter 6 reveals an increased estimation accuracy of the proposed CS-based localization algorithm when compared with the statistical localization methods introduced in the previous sections. Finally, the overall CS-based localization method is summarized in Algorithm 2, while Table 4.1 gathers the symbols used in the CS setup.

Algorithm 2 The Compressive Sensing positioning method

1. Training phase: collect RSS measurements from all APs at each cell
   For each AP i generate $\mathbf{\Phi}_T^i$ and $\mathbf{\Psi}_T^i$
2. Runtime phase: collect RSS measurements from each AP at the unknown position
   For each AP i generate $\boldsymbol{x}_{c,R}^i$
3. At runtime perform the following steps:
   - Send the length of runtime RSS measurements, $n_{c,i}$, to the server
   - From each $\mathbf{\Phi}_T^i$ extract the columns until line $n_{c,i}$ and send it to the wireless device
   - Compute the measurements vector $\mathbf{g}_{c,i}$ and send it to the server
   - Perform CS reconstruction at the server by solving (4.5)
4. Report the estimated position, c^*, as the centroid of the individual estimates given by the CS reconstruction scheme per AP

The amount of transmitted data is further reduced in the proposed implementation by selecting to process the RSS readings of only the top P' strongest APs, that is, the APs with the highest mean RSS value of the corresponding vectors $\psi_{R,c}^i$. An additional advantage of this process is that in many cases we discard the potentially confusing information from APs from which either there was not any reception at all, or there was a link failure with the device during the runtime phase. The experimental evaluation presented in the next section reveals an increased estimation accuracy of the proposed CS localization algorithm. Finally, the overall system architecture is shown in Fig. 4.1(a).

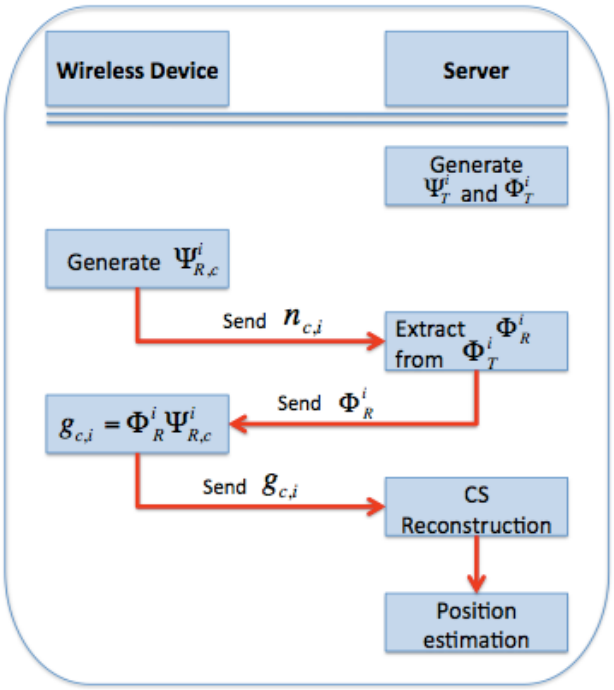
In a recent work [72], a CS-based indoor localization method was introduced based on RSS measurements. In particular, the location estimation algorithm is carried out on the mobile device by using the average RSS values in order to construct the transform basis. The sparsity-based CS localization algorithm proposed in this chapter differs from the work in [72] in several aspects. In contrast to [72], where the estimation is performed by the wireless device with the potentially limited resources, in our system the computational burden is put on the server, where increased storage and processing resources are available. Besides, in the proposed localization scheme the CS approach is applied directly on the raw RSS measurements and not on their average as in [72], thus exploiting their time-varying behavior. Then, the estimation of the unknown position is performed by solving a constraint optimization problem for reconstructing a sparse vector with its coordinates being "1" or "0" depending on whether the mobile device is placed or not at the corresponding cell.

Table 4.1: CS-WLAN symbols and notation

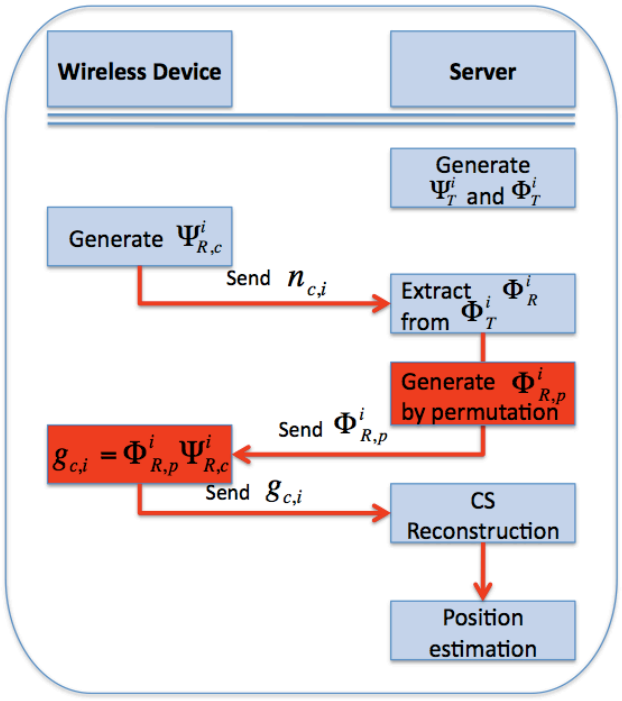
P	number of APs
C	number of cells
$\mathbf{x}^i_{j,T}$, $\mathbf{x}^i_{j,R}$	training/runtime RSS vector at cell j from AP i
$n_{j,i}$	number of RSS measurements at cell j from AP i
$N_i = \min_j\{n_{j,i}\}$	number of RSS measurements kept at all cells for AP i
M_i	number of CS measurements generated for AP i
$\mathbf{g}_{j,i}$	CS measurement vector at cell j for AP i
$\mathbf{\Phi}^i_T$, $\mathbf{\Phi}^i_R$	training/runtime measurement matrix for AP i
$\mathbf{\Psi}^i_T$	sparsifying matrix (dictionary) for AP i
$\mathbf{w}$	position indicator (sparse) vector

4.4 CS Weak Encryption Property and Secure Positioning

Due to their acquisition process, CS measurements can be viewed as "weakly encrypted" for an attacker without knowledge of the measurement matrices $\mathbf{\Phi}^i_T$. CS-based encryption provides both signal compression and encryption guarantees, without the additional computational cost of a separate encryption protocol and thus it could be useful in location estimation, where the implementation of an additional software layer for cryptography could be costly. The encryption property of a CS approach relies on the fact that the matrix $\mathbf{\Phi}^i_T$ is unknown to an unauthorized entity, since $\mathbf{\Phi}^i_T$ can be generated using a (time-varying) cryptographic key that only the device and the server share. An attack could be considered as the attempt to estimate the key by trying to find the special structure of the $\mathbf{\Phi}^i_T$ matrix [82]. In this proposed approach no cryptographic key is required, since it is based only on the matrices $\mathbf{\Phi}^i_T$ $(i = 1, \ldots, P)$. More specifically, the server extracts the sub-matrix $\mathbf{\Phi}^i_R$ from $\mathbf{\Phi}^i_T$ and then permutes its lines forming a new $\mathbf{\Phi}^i_{R,p}$, which is then sent to the wireless device, where the associated measurement vector $\mathbf{g}_{c,i} = \mathbf{\Phi}^i_{R,p}\psi^i_{R,c}$ is computed. A potential attacker has two options, either to try capturing $\mathbf{\Phi}^i_{R,p}$ by intercepting the server $\rightarrow$ device direction, or by acquiring $\mathbf{g}_{c,i}$ by intercepting the opposite direction. In the first case, modern network cryptographic protocols could guarantee that the decryption of $\mathbf{\Phi}^i_{R,p}$ is almost infeasible in practice due to the combinatorial nature of the inverse problem. In the second case, as it will be illustrated in Chapter 6, even the exact knowledge of $\mathbf{g}_{c,i}$ is insufficient, resulting in a significantly increased estimation error, when the attacker does not achieve the exact estimate of $\mathbf{\Phi}^i_{R,p}$. Finally, the overall security system architecture is shown in Fig. 4.1(b).



(a)



(b)

Figure 4.1: Flow diagram of the (a) CS-WLAN localization scheme (b) secure CS-WLAN localization scheme

4.5 Chapter Summary

This chapter introduced an indoor localization method based on CS RSS fingerprints. In the present work, the unknown location was estimated by performing separate reconstruction for each AP. The enhanced encryption capabilities of the proposed CS-WLAN architecture, without the additional computational cost of a separate encryption protocol, are also evaluated in Chapter 6.

Chapter 5

Indoor Tracking System

Contents

5.1 Introduction

Chapters 3 and 4 describe two positioning systems that can accurately estimate a stationary user's position. These positioning systems are combined in this chapter in order to track a dynamic mobile user. The proposed indoor tracking system uses the Kalman filter in conjunction with the power of compressive sensing to represent accurately sparse signals and a region-based multivariate Gaussian model. This chapter first describes briefly the general Bayesian tracking model and how it can be reduced to a Kalman filter. Then, the proposed indoor tracking system is presented.

5.2 General Bayesian Tracking Model

The problem can be modeled by a general Bayesian tracking model, with

$$d(t) = f_t(d(t-1), w(t)) \tag{5.1}$$
$$z(t) = h_t(d(t), w(t)) \tag{5.2}$$

where $d(t) = [x(t), y(t), v_x(t), v_y(t)]$. It is mentioned also that x, y are the cartesian coordinates of user's location and $v_x(t), v_y(t)$ are the velocities at the x, y directions. Assuming that the tracking is a Markov process of order one, the state evolves as a function of previous state and $w(t)$ which is a i.i.d. process noise vector. The measuremenent $z(t)$ depends on the current state and the i.i.d. measurement noise vector $w(t)$ through the function h_t.

The current location of the mobile user can be estimated recursively from the set of measurements up to time t,

$$z(i:t) = \{z(i), i = 1, 2, ..., t\}, \tag{5.3}$$

which in terms of pdf is denoted as

$$P(x(t)/z(1:t)). \tag{5.4}$$

5.3 Kalman Filter

If we assume that the process and measurement noises are Gaussian and the motion dynamic model is linear, then we can use the Kalman filter. The process and measurement equations of the Kalman tracking model are given by:

$$x(t) = \mathbf{F}\mathbf{x}(t-1) + \mathbf{w}(t) \tag{5.5}$$
$$z(t) = \mathbf{H}\mathbf{x}(t) + \mathbf{v}(t) \tag{5.6}$$

where $\mathbf{x}(t) = [x(t), y(t), v_x(t), v_y(t)]^T$ is the state vector, z(t) is the measurement vector and matrices $\mathbf{F}$ and $\mathbf{H}$ define the linear motion model. The process noise $\mathbf{w}(t) \sim N(\mathbf{0}, \mathbf{S})$ and the measurement noise $\mathbf{v}(t) \sim N(\mathbf{0}, \mathbf{U})$ are assumed to be independent with covariance matrices $\mathbf{S}$ and $\mathbf{U}$, respectively. The current location of the mobile user is assumed to be the previous location plus the distance travelled, which is computed as the time interval Δt times the current velocity.

The steps to obtain the final estimates of the state vector $\hat{x}(t)$ and the error covariance $\mathbf{P}(t)$ during the prediction and the update stage are computed by the following set of equations:

$$\hat{x}^-(t) = \mathbf{F}\hat{\mathbf{x}}(t-1) \tag{5.7}$$
$$\mathbf{P}^-(t) = \mathbf{F}\mathbf{P}(t-1)\mathbf{F}^{\mathbf{T}} + S \tag{5.8}$$
$$\mathbf{K}(t) = \mathbf{P}^-(t)\mathbf{H}^T(\mathbf{H}\mathbf{P}^-(t)\mathbf{H}^T + \mathbf{U})^{-1} \tag{5.9}$$
$$\hat{x}(t) = \hat{x}^-(t) + \mathbf{K}(t)(z(t) - \mathbf{H}\hat{\mathbf{x}}^-(t)) \tag{5.10}$$
$$\mathbf{P}(t) = (\mathbf{I} - \mathbf{K}(\mathbf{t})\mathbf{H})\mathbf{P}^-(\mathbf{t}) \tag{5.11}$$

5.4 CS-Kalman Filter-based Indoor Tracking

First, the path tracking model employs a region-based multivariate Gaussian (MvG) model and then, for each region, a CS approach is applied as a refinement step.

The device periodically collects the online RSS from each APs at a time interval Δt. Then, the indoor tracking system uses the RSS to estimate the user's location at time t, which is denoted as $\hat{p}(t) = [\hat{x}(t), \hat{y}(t)]^T$. For each time step t,

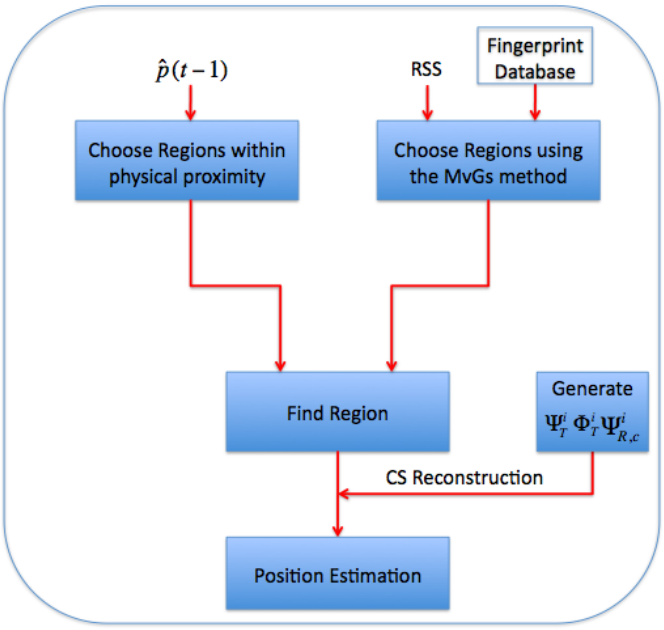


Figure 5.1: Flow diagram of the proposed indoor tracking system.

the measurement vector $z(t)$ is the current user's estimated location computed by the positioning system.

The Kalman filter can be applied on the CS-based positioning system described in Chapter 4 to improve the accuracy in estimating the dynamic user's route. The online measurement vector collected at time t, is first evaluated at the localization stage to reduce the area of interest by selecting the relevant training regions according to the MvG algorithm. The system uses not only the online RSS measurements, but also the previous user's location estimate to select the appropriate training regions, based on physical proximity. Fig. 5.1 shows the proposed indoor tracking system that is build on the top of the MvGs-CS-based positioning system.

5.5 Chapter Summary

In this chapter, a hybrid path tracking system is presented, which exploits the efficiency of a Kalman filter in conjunction with the power of compressive sensing to represent accurately sparse signals and a region-based multivariate Gaussian model. By using the user's previous estimated locations, the tracking system is able to refine the current estimate in two ways: 1) to select appropriate training regions and 2) to apply Kalman filter for better location estimate. This selection of the training regions ensures that the reduced region of interest are within the walking range of the user and provides a way to reject the invalid online RSS readings. The tracking

system also introduces the Kalman filter stage, which uses the temporal position estimation, $p(t)$, into the current final estimation.

Chapter 6

Performance analysis

Contents

6.1 Introduction

In this chapter, the localization performance of the algorithms decribed in the previous chapters is presented. The evaluation took place in three distinct environments with different area and topology. More specifically, the training signatures are generated by collecting signal-strength values at various predefined cells of the two grids corresponding to the two environments. In both cases, training measurements were collected at each cell of the grid, while runtime measurements were collected at 30 randomly selected cells. The trainer remained still for approximately 90 sec and 30 sec to collect beacons at each position during training and runtime, respectively, resulting in more than 100 and 200 RSS values per AP at each cell for the runtime and training phase, respectively. To capture signal-strength values, *iwlist*, which polls each channel and acquires the MAC address and RSS measurements from each AP (in dBm), and *tcpdump*, a passive scanner relying on *libpcap*, for the retrieval of each packet were employed. A Sony Vaio and a Toshiba laptop with the same wireless adapter (ipw2200) were used for the collection of both training and runtime signal-strength values. Besides, in the subsequent experiments we evaluate two distinct scenarios in both environments, differing in the number of people which were present during the collection of the RSS measurements. In the following, let Sc-A and Sc-B denote the scenario corresponding to the presence of a small and a large number of people, respectively.

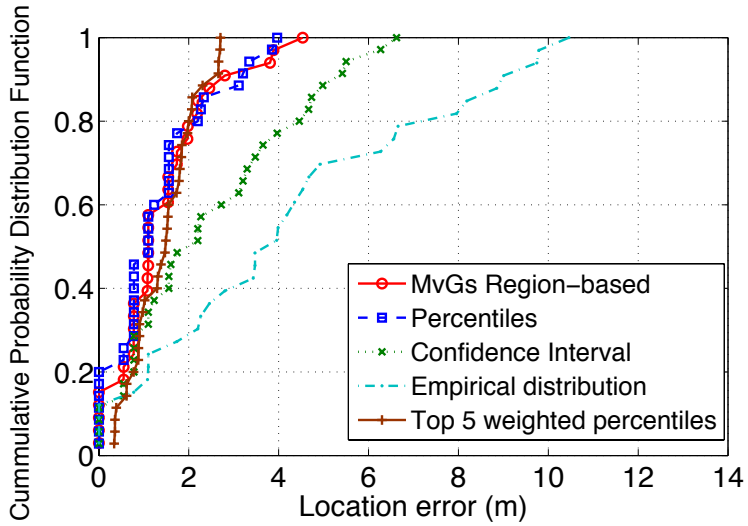
6.2 Evaluation at FORTH

The evaluation at FORTH took place in the Telecommunication and Networks Lab (TNL), an area of 7 m $\times$ 12 m, which was discretized in a grid structure with cells of 55 cm $\times$ 55 cm. During the runtime phase two datasets are collected corresponding to the two scenarios: (i) Sc-A: the dataset was collected on a Sunday, at around 2 a.m., with only one person present in the laboratory, except the one collecting the measurements, (ii) Sc-B: the dataset was collected on a typical weekday at around 5 p.m., during which there were from 10 to 15 people in the laboratory, and several others walking in the hallways outside. The training set was common for both Sc-A and Sc-B scenarios, and was collected on the same day as the Sc-A runtime dataset, including measurements from 84 different cells. The total number of APs covering the area was 10, while on average 5.4 APs were detected at a given cell.

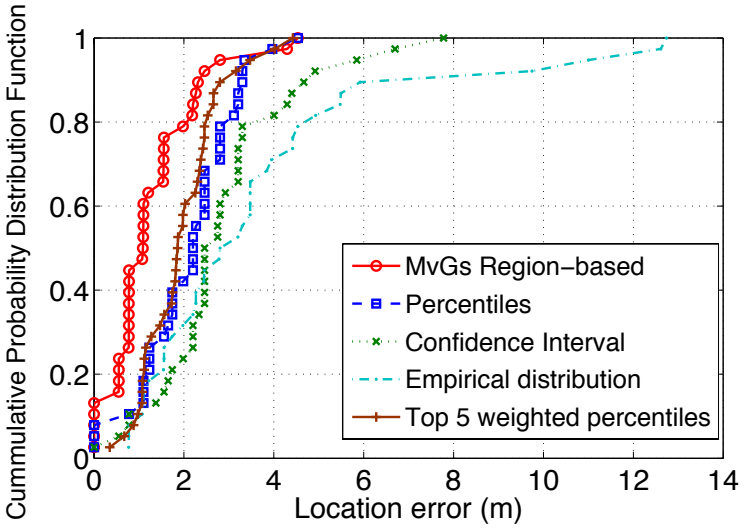
In order to evaluate the performance of the various fingerprinting methods we computed the *localization error*, measured as the Euclidean distance between the centers of the reported cell and the cell at which the mobile user was actually located at runtime.

Figs. 6.1(a)-6.1(b) present the localization error of the different signature-based approaches during the Sc-A and the Sc-B scenario, respectively. As it can be seen, the multivariate Gaussian model (MvG) outperforms the percentiles, the confidence interval (90%), and the empirical distribution approaches. More specifically, for the Sc-A dataset, the median error is equal to 2.19 m and 1.10 m for the confidence interval and percentiles, respectively, while the MvGs results in an error of 1.09 m. Regarding the Sc-B dataset, the median error of the MvGs is 1.10 m, while the confidence interval (90%) and the percentiles methods report an error of 2.60 m and 2.20 m, respectively. In general, it is expected that as the number of APs that participate in the signature generation increases, so does the accuracy of distinguishing the correct cell from other further-away cells, which may have similar training fingerprints with the runtime one due to transient phenomena or radio propagation characteristics in the given environment. To measure the impact of the number of APs on the localization accuracy, we associate each AP with a *popularity index* that indicates the number of cells at which measurements from the beacons of that AP were collected during both training and runtime phases. Let $|\{c|APi \in effectiveAP(c)\}|$ denote the popularity index of AP i. The APs were sorted in a descending order based on their popularity indices and the analysis was repeated using the top k most popular APs for the Sc-A and Sc-B datasets.

The number of selected APs is another factor that affects the localization performance. As we expected, the higher the number of APs, the lower the localization error as shown in Fig. 6.2. However, the effect of an increasing number of APs diminishes after a certain threshold. This stems from the fact that the



(a) Sc-A dataset



(b) Sc-B dataset

Figure 6.1: Localization performance of various fingerprint positioning methods at FORTH.

Sc-B dataset is subject to a larger number of transient phenomena than the Sc-A dataset, and thus affecting the estimation accuracy. This also explains the fact that the number of APs affects more the performance during the Sc-B than the Sc-A scenario.

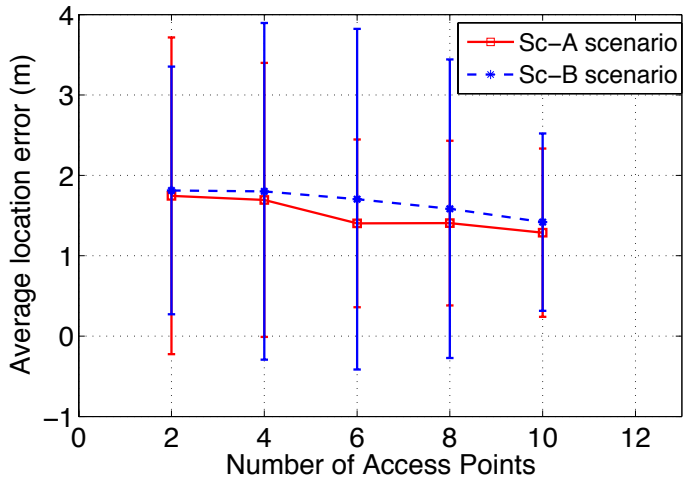


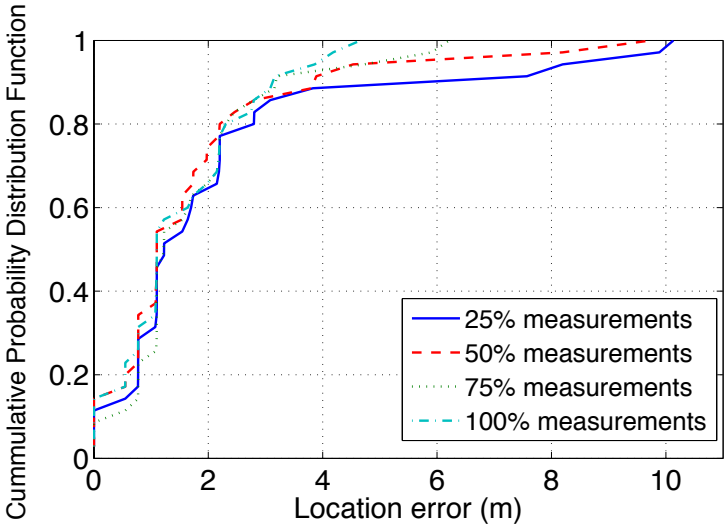
Figure 6.2: Impact of the number of APs on localization error. The x-axis indicates the number of the top x APs considered in both training and runtime datasets.

Figs. 6.3(a)-6.3(b) show the localization accuracy as a function of the number of RSS measurements for the MvG method. The % indicates the percentage of RSS measurements considered in both the training and runtime datasets. In general, the larger the measurement set, the more accurate the position estimation. Moreover, the increase in the number of RSS measurements affects more the performance in the Sc-B scenario, where the environment is more dynamic with the presence of an increased number of people.

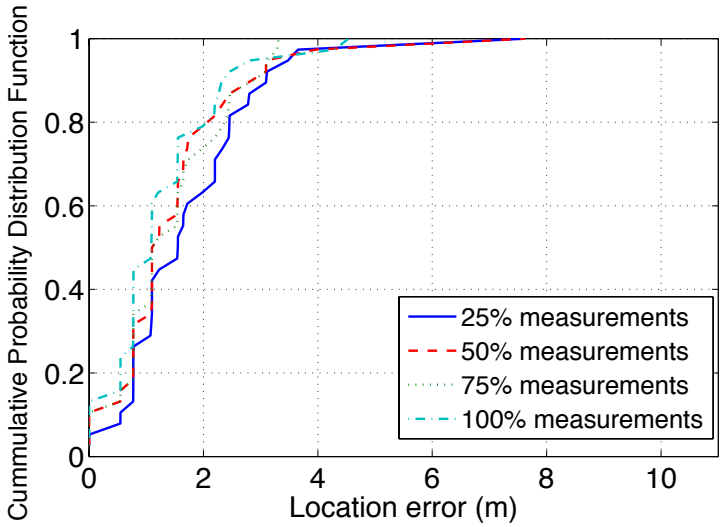
Finally, we evaluate and compare the performance of the proposed CS-based localization approach with the statistical fingerprint-based MvG method. For this purpose, the CS reconstruction problem (4.5) is solved using a primal-dual interior point method (L1) and the orthogonal matching pursuit (OMP) algorithm[1]. Both CS methods employ only the 25% of the total runtime RSS measurements.

Figs. 6.4(a)-6.4(b) show the corresponding cumulative probability distributions of the localization error for the three methods. In particular, the median error for the Sc-A scenario is equal to 1.09 m for the MvG, and 1.08 m for the L1 and OMP approaches. Similar results are obtained for the Sc-B scenario

[1]Matlab codes can be found in http://www.acm.caltech.edu/l1magic/,http://sparselab.stanford.edu/.



(a) Sc-A dataset

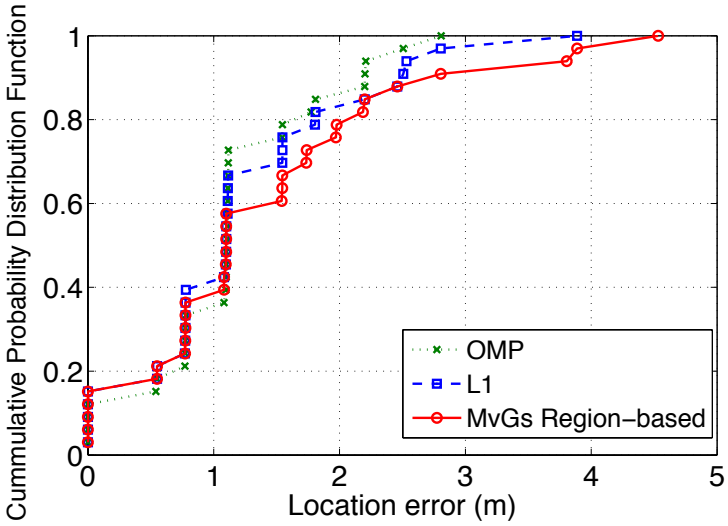


(b) Sc-B dataset

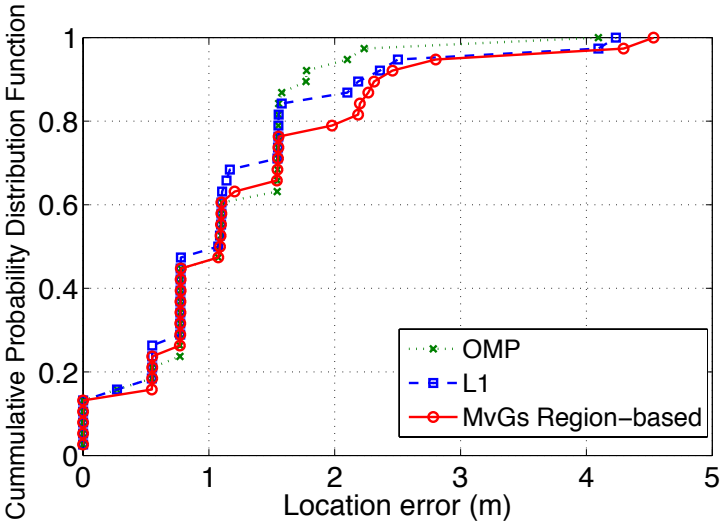
Figure 6.3: Localization error as a function of the number of RSS measurements for the MvG method.

with a median error of 1.09 m for the MvG and 1.08 m for the L1 and OMP methods. The similarity in performance in the case of TNL can be attributed to the simple topology of the indoor space and the relative position of the APs with respect to the people which were present in the lab, that did not affect significantly the variability of the RSS measurements between the two scenarios.

Besides, the difference in performance between the CS-based approach and the percentiles or the empirical distribution method, stems from the difference in the way they combine the individual estimations per AP into a single final estimate. More specifically, the CS-based approach for the location estimation is carried out for each AP separately, using the compressed RSS measurements, and the final estimate is given by the centroid of the individual estimated positions. On the other hand, the percentiles and the empirical distribution methods perform an averaging over all APs of the values of the corresponding distance function *before* the final location estimation. For instance, in the case of the empirical distribution method, each cell is assigned a weight which corresponds to the average empirical KLD of each AP (at that cell) from the runtime measurements collected at the unknown position from the same AP. As a result, two very distinct cells with similar KLD averages may be reported erroneously to be close to each other. This is not the case for a CS-based approach, where a potentially wrong estimate based on a single AP can be eliminated if the estimates based on the remaining APs are close to the true cell.



(a) Sc-A dataset



(b) Sc-B dataset

Figure 6.4: Performance evaluation of the CS-based methods at FORTH.

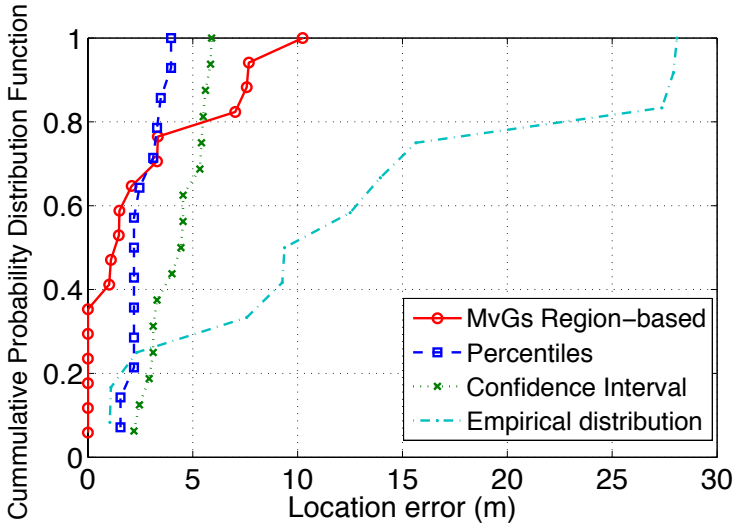
6.3 Evaluation at CretAquarium

Cretaquarium is the largest and most popular aquarium in Greece, covering an area of 1760 m^2 and consisting of more than 40 tanks. The physical space was represented as a grid with cells of 1 m $\times$ 1 m, with seven IEEE 802.11 APs covering the whole testbed, out of which 3.4 on average can be detected at a given cell. Training and runtime signal-strength measurements were collected in December, January and February of 2011 for the entire testbed during the two different scenarios, namely, the Sc-A and the Sc-B. During the Sc-A there were only two people (the ones collecting the signal-strength measurements) most of the time in the aquarium. On the other hand, in the Sc-B scenario there was a scheduled visit of a class of students with about 25 people near the trainers during the whole time period of the data collection. The training set was common for both scenarios and was collected during different days of December and January 2011.

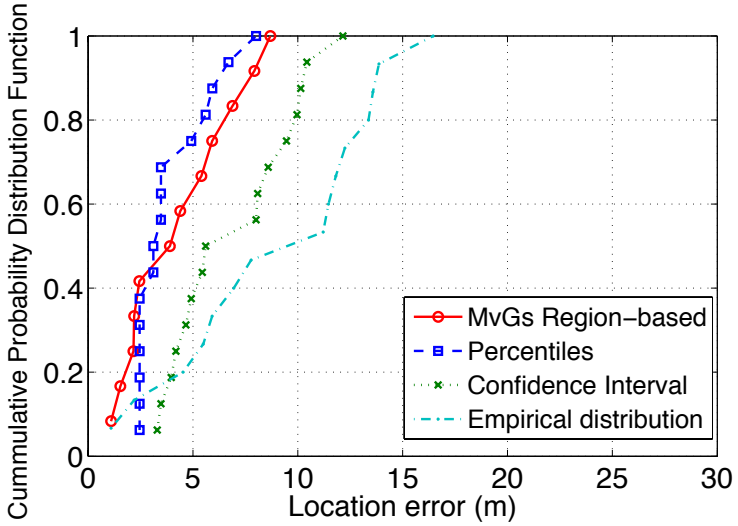
Under normal conditions (Sc-A), the median localization error using the MvG method is 1.48 m, while the percentiles resulted in a median localization error of 2.20 m, as shown in Fig. 6.5(a). During the Sc-B scenario, the median error for the percentiles method increases at 3.29 m, while the MvG gave an estimation error of 4.15 m (Fig. 6.5(b)). In both scenarios, the confidence intervals and the empirical distribution approach result in significantly higher errors. As it was also the case for the experimental evaluation in the premises of TNL, the largest the measurement set, the more accurate the position estimation is. Moreover, the relatively small size of the signal-strength measurement dataset has a noticeable effect on the performance of the multivariate Gaussian method in the Sc-B scenario, since the accuracy in estimating the signature (mean and covariance) for each cell decreases due to the high variability of the RSS measurements, whose statistics require an increased sample-size to be expressed accurately.

As the above results reveal, the presence of an increased number of people interfering with the two trainers, along with the complicated topology of the Cretaquarium facilities make the accurate location estimation a challenging task. As a final evaluation, we compare the performance of the proposed CS-based localization technique with the MvG region-based approach. In the previous subsection it was mentioned that one of the advantages of a CS method is its inherent ability to extract the salient signal's information content by suppressing potential noise-like features. The complicated topology of the Cretaquarium, as opposed to the simple topology of the TNL, along with the multi-path and/or fading phenomena characterizing the RSS measurements due to the tanks, enforce the presence of such noise-like features. Thus, we expect that CS-based methods will present an increased robustness resulting in a more accurate position estimate.

Indeed, as it is shown in Figs. 6.6(a)-6.6(b), for the Sc-A scenario the median error is equal to 1.48 m and 1.10 m for the MvG and the L1 method, respectively, while OMP results in a median error of 1.09 m. Besides, for the Sc-B scenario



(a) Sc-A dataset



(b) Sc-B dataset

Figure 6.5: Performance evaluation of various fingerprint positioning methods at Cretaquarium.

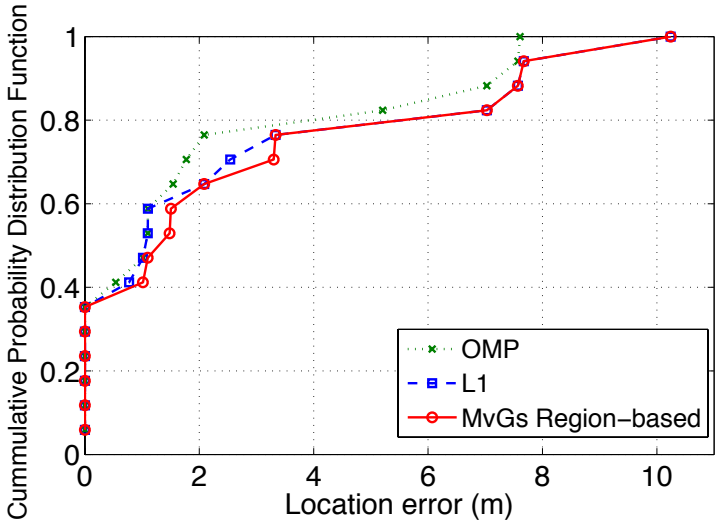
the corresponding median error is equal to 4.15 m and 4.14 m for the MvG and the L1 method, respectively, while the OMP approach results in a median error of 3.59 m. The increased localization performance associated with the Sc-A scenario is justified by the decreased variance of the RSS measurements, since the major factor that affects the behavior of the received signal is the reflection from the tanks and the potential fading when there is not a nearby AP to the mobile user. As before, only 25% of the total runtime RSS measurements were employed in the CS approach.

From the above, we notice that the proposed CS-based localization method improves the localization accuracy when compared with the MvG method in a more complicated environment, as the Cretaquarium is. A reason for this is that, in contrast to the MvG which employs the average RSS statistics (means and covariances), the CS-based scheme exploits directly the raw RSS measurements, via the low-dimensional measurement vectors, and thus it is able to account for the time-varying nature of the RSS readings. Moreover, a CS approach is characterized by the inherent property that it suppresses the noise-like features of the signal to be reconstructed, while extracting its prominent information content. This is not the case with the MvG method, where the statistical signatures (fingerprints), and consequently the computed KLD, could be affected more from potential noise-like fluctuations in the RSS measurements.

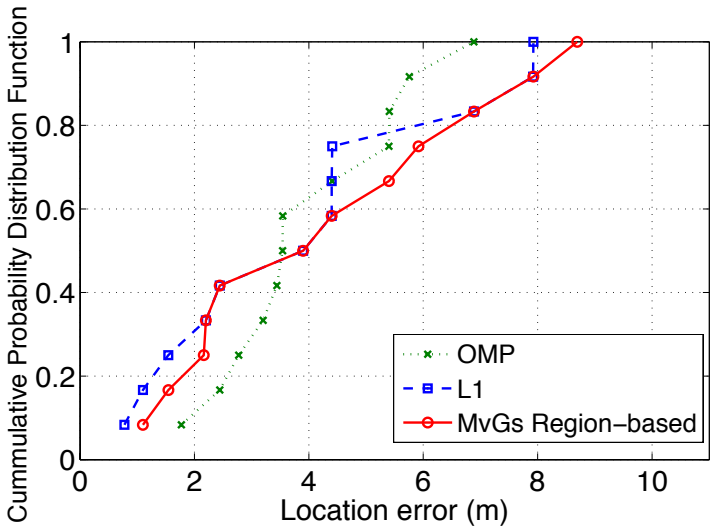
As it was mentioned in the previous subsection for the TNL environment, the increased accuracy of a CS-based approach when compared with the percentiles or the empirical distribution methods, is due to the different way the individual estimates per AP are averaged into a single final position estimate.

Moreover, the experimental evaluation using the collected data from two distinct environments shows that apart from a suitable selection of the sparsifying basis $\mathbf{\Psi}$ and the measurement matrix $\mathbf{\Phi}$, another factor that affects significantly the localization accuracy is the selection of the reconstruction algorithm. For the present experimental configurations in both the TNL and Cretaquarium, the OMP algorithm is an appropriate choice. However, one of the advantages of the CS framework is that the generation of CS measurements is fully decoupled with the process of reconstructing the corresponding sparse vector, that is, with the same set of CS measurements $\mathbf{g}$, the localization performance can be improved by developing a more efficient reconstruction technique.

A guiding application was designed for the aquarium to provide personalized information to visitors about the habitats in the tank in front of them. For this purpose, the physical space was divided into 17 zones according to the application requirement. The positioning system reported the zone in which the visitor was located. Experiments were contacted using the Ekahau, a commercial positioning system, which also employs RSS-based fingerprints. The tests took place at the same period and for the same runtime cells under an Sc-B scenario, that is, with a relatively large number of visitors moving close to the trainers. In each zone the



(a) Sc-A dataset



(b) Sc-B dataset

Figure 6.6: Performance evaluation of CS-based methods at Cretaquarium.

system was tested at three different positions. The correct zone was reported in 80% of the times resulting in a median error of 4.6 m.

6.4 Evaluation at INRIA

In this section, the performance of the proposed CS-WLAN localization method is evaluated and compared with previous fingerprint-based algorithms.

The dataset used in the present evaluation was acquired in the building 21 of INRIA, at Rocquencourt campus (Paris). The wireless coverage is achieved by employing an infrastructure consisting of five IEEE 802.11 APs. The area used in the formation of the RSS map is discretized in cells of equal dimensions 0.76 m $\times$ 0.76 m. The RSS map consists of measurements from different cells and for an average number of five APs per cell. The time intervals during the acquisition in the training and runtime phase were set to 90 sec and 30 sec, respectively.

The estimation accuracy of the methods tested hereafter is evaluated in terms of the localization error, which is defined as the Euclidean distance between the centers of the estimated cell and the true cell where the mobile user is located at runtime. Runtime measurements in 32 distinct cells are employed in the subsequent evaluation.

Fig. 6.7 presents the localization error of the three methods introduced briefly in Chapters 2 and 3 (region-based MvG, kNN ($k = 3$), and spatial sparsity-based). The median error is equal to 1.99 m and 2.34 m for the kNN, and the spatial sparsity-based methods, respectively, while the MvG approach results in a median error of 1.56 m. Fig. 6.8 shows the estimation error for the proposed CS-based

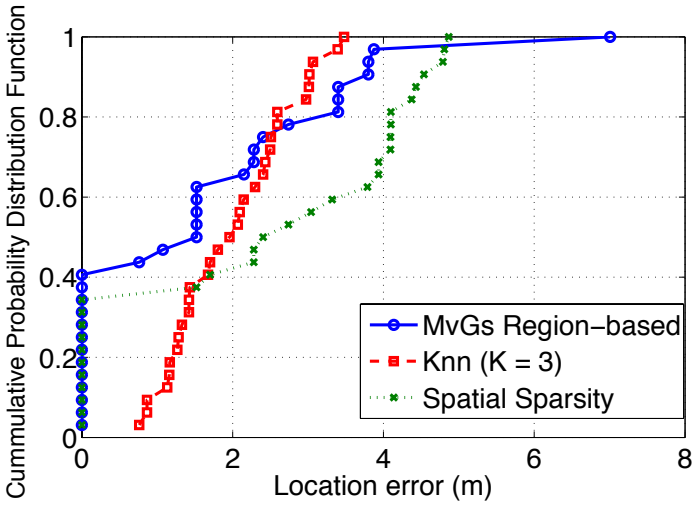


Figure 6.7: Performance comparison of previous fingerprint-based localization methods at INRIA.

method, averaged over 100 Monte-Carlo runs, where in each run a distinct measurement matrix is generated. The reconstruction performance is compared between several widely-used norm-based techniques and Bayesian CS algorithms. More specifically, the following methods are employed[2]: 1) ℓ_1-norm minimization using the primal-dual interior point method (L1EQ-PD), 2) Orthogonal Matching Pursuit (OMP), 3) Stagewise Orthogonal Matching Pursuit (StOMP), 4) LASSO, 5) BCS [80], and 6) BCS-GSM [81]. As it can be seen, the BCS and BCS-GSM methods outperform the others with a median error of 1.89 m and 1.78 m, respectively (ℓ_1 2 m, OMP 2.03 m, StOMP 2.01 m, and Lasso 2.30 m). In this experiment only 8% of the total runtime RSS measurements vector is employed. The effect of

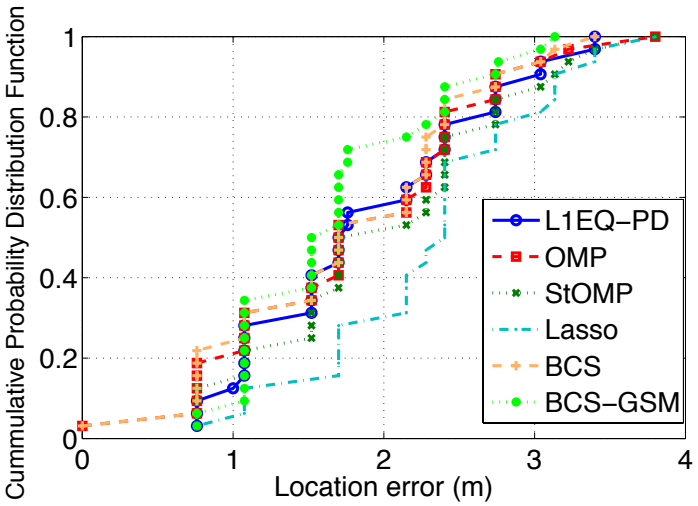


Figure 6.8: Performance evaluation of the CS-WLAN localization method for several reconstruction algorithms at INRIA.

the number of CS measurements in the estimation accuracy is further examined for the top two candidates, namely, the BCS and BCS-GSM. Fig. 6.9 shows the corresponding localization error as a function of the percentage of the number of RSS measurements ($M = rN$ with $r \in \{5\%, 10\%, 15\%, 20\%\}$). As we expected, the localization accuracy increases by increasing the number of CS measurements, and for 15% of the RSS values the proposed approach outperforms the MvG method, which was the best among the previous fingerprint-based techniques.

Fig. 6.10 compares the location error of the MvG and the BCS-GSM (with 25% of RSS measurements) methods, as a function of the input SNR. Each RSS vector is corrupted by additive white Gaussian noise with the SNR varying from 10 to 40 dB. As it can be seen, the proposed CS-based approach presents a clear superiority against MvG, especially for lower SNR values.

Finally, Fig. 6.11 illustrates the encryption capabilities of the proposed CS local-

[2]For the implementation of methods 1)-5) the MATLAB codes can be found in: http://sparselab.stanford.edu/, http://www.acm.caltech.edu/l1magic, http://people.ee.duke.edu/~lcarin/BCS.html

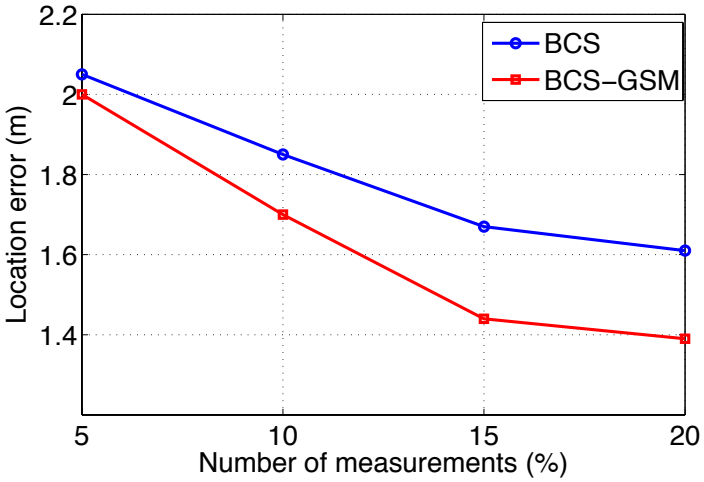


Figure 6.9: Localization accuracy of CS-WLAN using BCS and BCS-GSM, as a function of the number of CS measurements at INRIA.

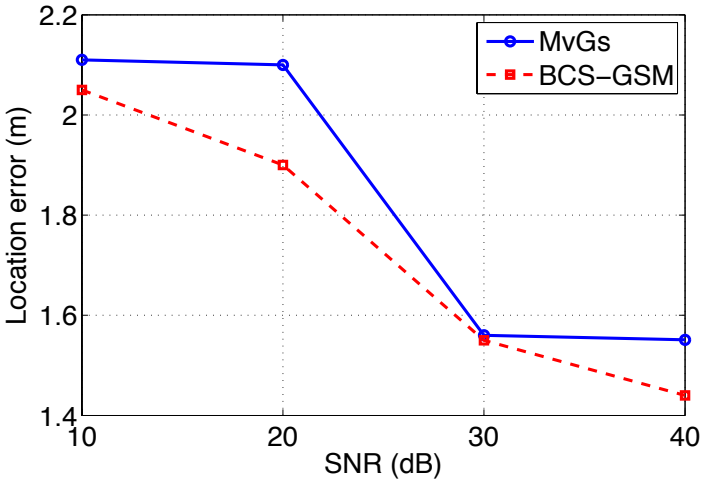


Figure 6.10: Localization accuracy of the MvG and BCS-GSM methods for varying input SNR at INRIA.

ization method for the BCS and BCS-GSM algorithms. In particular, the average localization error (over 100 Monte-Carlo runs) is shown as a function of the percentage of permuted lines ($\{0\% : 20\% : 100\%\}$) of the true matrices $\mathbf{\Phi}_R^i$, where the reconstruction is performed by considering exact knowledge of the measurement vectors $\mathbf{g}_{c,i}$. The results agree with our intuition that as the complexity of the permutation increases, the estimation accuracy decreases without an exact estimate of the true measurement matrix.

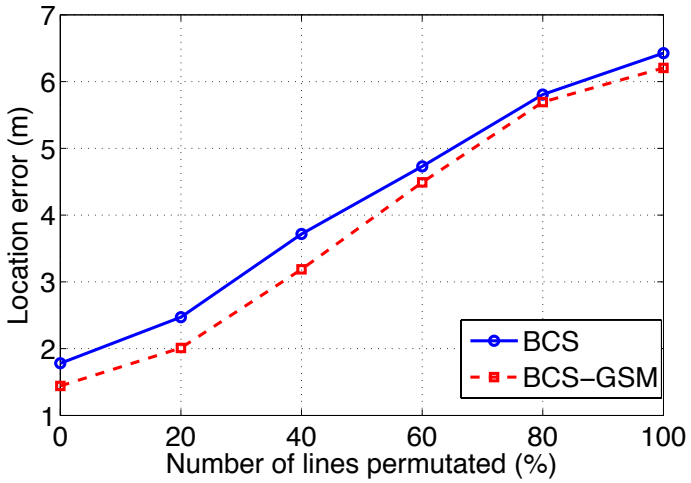


Figure 6.11: Evaluation of CS-WLAN encryption property using BCS and BCS-GSM, for a varying number of permuted lines of $\mathbf{\Phi}_R^i$ at INRIA.

6.5 Evaluation on simulated data

In this section, the performance of the proposed CS-WLAN localization method is evaluated and compared with previous fingerprint-based algorithms on a simulated space. It presents the experimental results of the proposed method by generating simulated RSS sequences using an appropriate propagation model in a more complicated indoor environment. The formulas used for the generation of the simulated data can be found in [83].

The estimation accuracy of the methods tested hereafter is evaluated in terms of the localization error, which is defined as the Euclidean distance between the centers of the estimated cell and the true cell where the mobile user is located at runtime. Runtime measurements in 32 distinct cells are employed in the subsequent evaluation.

Fig. 6.12 presents the localization error of the three methods introduced briefly in Chapters 2 and 3 (region-based MvG, kNN ($k = 3$), and spatial sparsity-based). The median error is equal to 2.76 m and 2.01 m for the kNN, and the spatial sparsity-based methods, respectively, while the MvG approach results in a median error of 2.15 m.

Fig. 6.13(a) shows the estimation error for the proposed CS-based method, averaged over 100 Monte-Carlo runs, where in each run a distinct measurement matrix is generated. The reconstruction performance is compared between several widely-used norm-based techniques and Bayesian CS algorithms. More specifically, the following methods are employed: 1) ℓ_1-norm minimization using the primal-dual interior point method (L1EQ-PD), 2) Orthogonal Matching Pursuit (OMP), 3) Stagewise Orthogonal Matching Pursuit (StOMP), 4) LASSO, 5) BCS, and 6) BCS-GSM. As it can be seen, the BCS and BCS-GSM methods outperform the others with a median error of 1.06 m and 1.04 m, respectively (ℓ_1 1.62 m, OMP 5 m, StOMP 4.1 m, and Lasso 1.24 m). In this experiment only 8% of the total runtime RSS measurements vector is employed.

The effect of the number of CS measurements in the estimation accuracy is further examined for the top two candidates, namely, the BCS and BCS-GSM. Fig. 6.13(b) shows the corresponding localization error as a function of the percentage of the number of RSS measurements ($M = rN$ with $r \in \{5\%, 10\%, 15\%, 20\%\}$). As we expected, the localization accuracy increases by increasing the number of CS measurements, and for 15% of the RSS values the proposed approach outperforms the MvG method, which was the best among the previous fingerprint-based techniques.

Fig. 6.14(a) compares the location error of the MvG and the BCS-GSM (with 25% of RSS measurements) methods, as a function of the input SNR. Each RSS

vector is corrupted by additive white Gaussian noise with the SNR varying from 10 to 40 dB. As it can be seen, the proposed CS-based approach presents a clear superiority against MvG, especially for lower SNR values.

Finally, Fig. 6.14(b) illustrates the encryption capabilities of the proposed CS localization method for the BCS and BCS-GSM algorithms. In particular, the average localization error (over 100 Monte-Carlo runs) is shown as a function of the percentage of permuted lines ($\{0\% : 20\% : 100\%\}$) of the true matrices $\mathbf{\Phi}_R^i$, where the reconstruction is performed by considering exact knowledge of the measurement vectors $\mathbf{g}_{c,i}$. The results agree with our intuition that as the complexity of the permutation increases, the estimation accuracy decreases without an exact estimate of the true measurement matrix.

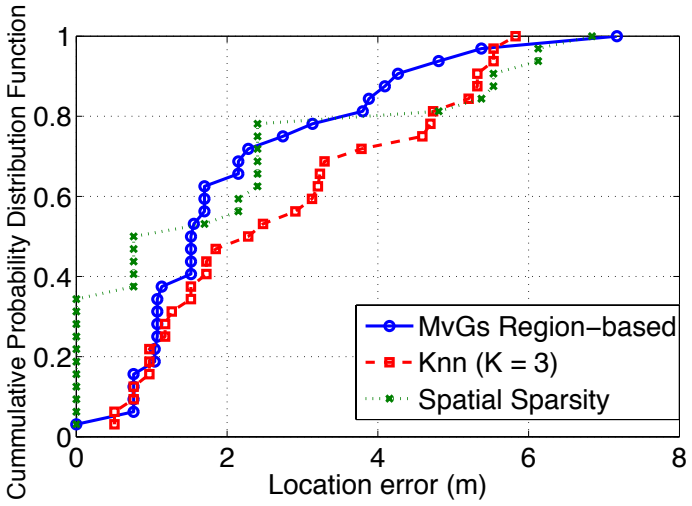
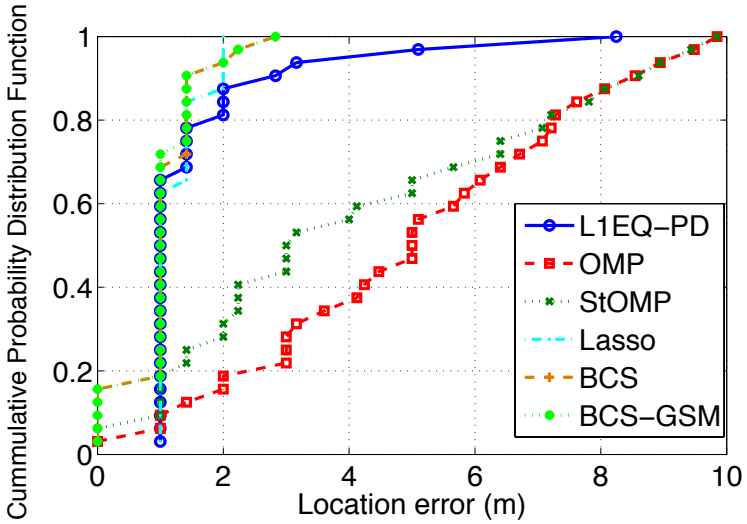
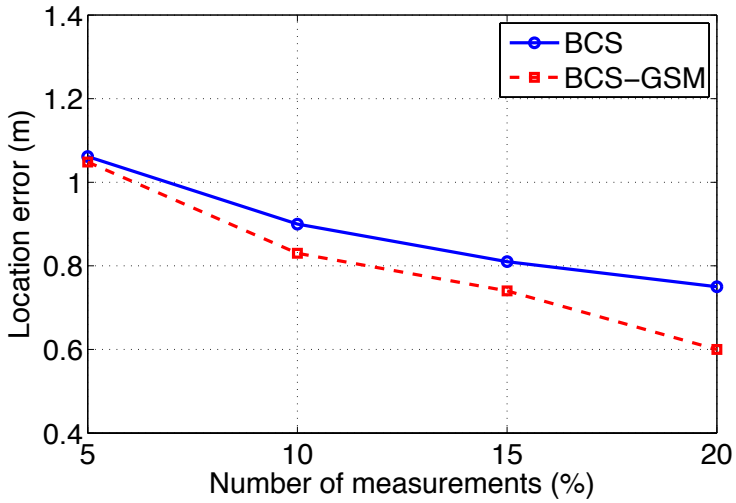


Figure 6.12: Performance comparison of previous fingerprint-based localization methods for the simulated dataset.

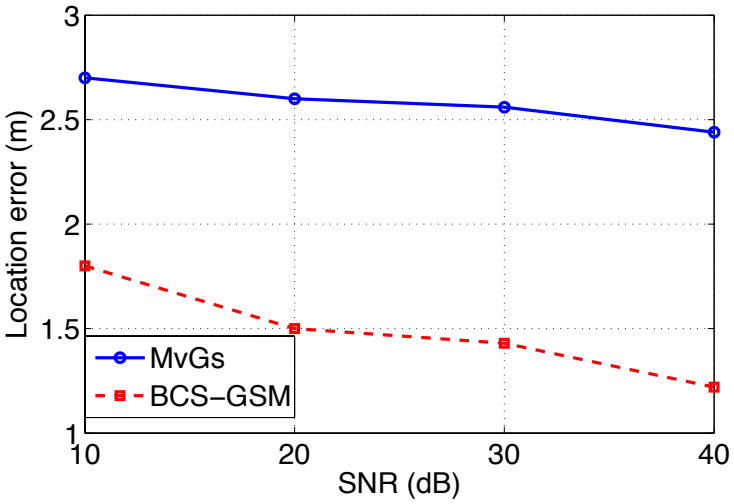


(a) Performance evaluation of the CS-WLAN localization method for several reconstruction algorithms.

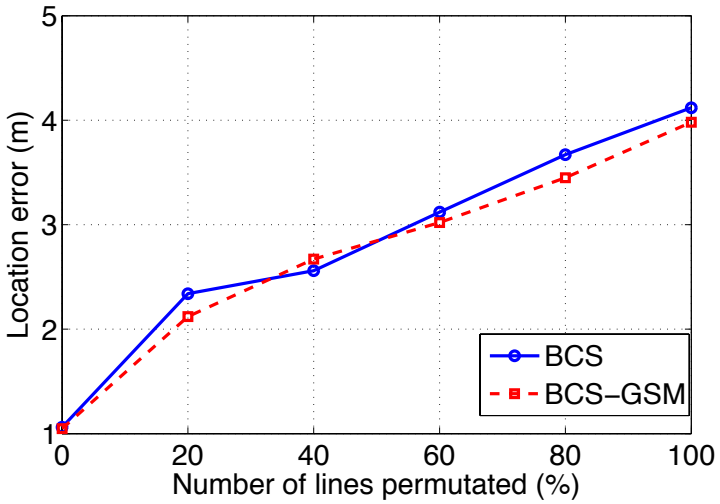


(b) Localization accuracy of CS-WLAN using BCS and BCS-GSM, as a function of the number of CS measurements on simulated data.

Figure 6.13: Performance evaluation of the CS-WLAN on simulated data



(a) Localization accuracy of the MvG and BCS-GSM methods for varying input SNR.



(b) Evaluation of CS-WLAN encryption property using BCS and BCS-GSM, for a varying number of permuted lines of $\mathbf{\Phi}_R^i$.

Figure 6.14: Localization accuracy for varying input SNR and performance evaluation of the CS-WLAN encryption on simulated data

6.6 Tracking Results in INRIA

Fig. 6.15 shows the original and estimated route of the mobile user during a test inside the building.

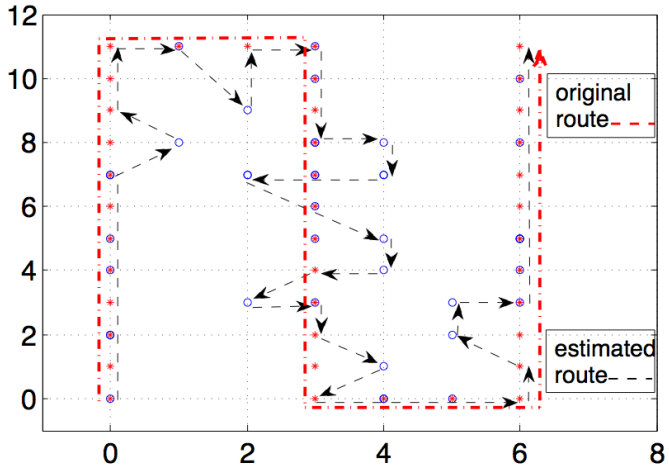


Figure 6.15: Original and estimated route of the mobile user using CS-KALMAN path tracking model at INRIA

Fig. 6.16 shows the localization error of the method introduced in Chapter 5, compared with the MvGs and a kNN-based approach (K = 3). The median error is equal to 1.9 m and 1.69 m for the kNN and the MvGs methods, respectively, while the BCS-GSM approach results in a median error of 1.36 m.

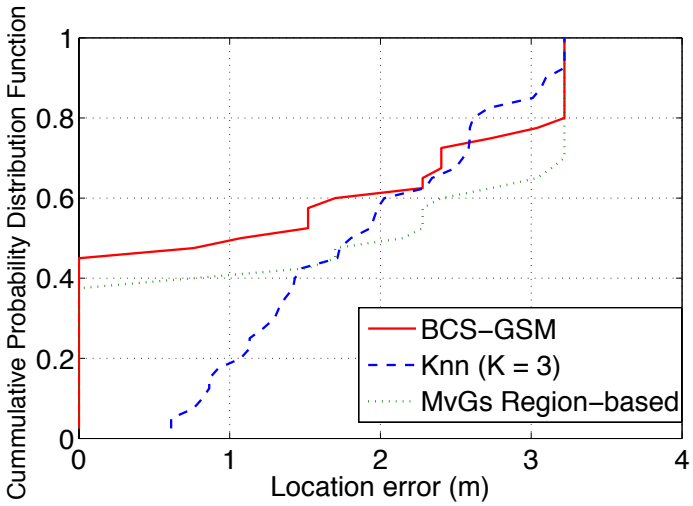


Figure 6.16: Performance evaluation of the CS-KALMAN path tracking model at INRIA

6.7 Chapter Summary

This chapter evaluates the performances of the proposed MvG, CS-WLAN positioning and indoor tracking system. There are three experimental sites. The experiments carried in FORTH-ICS, CretAquarium and INRIA-Rocquencourt campus.
The empirical experimental evaluation revealed that the multivariate Gaussian method usually outperforms previous statistical signal-strength fingerprint approaches, while the CS-based approach achieved a superior performance when compared to the multivariate Gaussian-based technique.
The reason for this is that, in contrast to the MvG which employs the average RSS statistics (means and covariances), the CS-based scheme exploits directly the raw RSS measurements, via the low-dimensional measurement vectors, and thus it is able to account for the time-varying nature of the RSS readings. Moreover, a CS approach is characterized by the inherent property that it suppresses the noise-like features of the signal to be reconstructed, while extracting its prominent information content. This is not the case with the MvG method, where the statistical signatures (fingerprints), and consequently the computed KLD, could be affected more from potential noise-like fluctuations in the RSS measurements.

Chapter 7

Conclusions and Future Work

This thesis introduced and compared two novel localization methods based on RSS measurements. In the first case, statistical signal-strength fingerprints are created using multivariate Gaussian distributions. Then the position of the device is estimated by computing the region with the training fingerprint that has the minimum KLD from the runtime fingerprint. In the second case, the localization problem was reduced in a sparse reconstruction problem in the framework of CS. The dimensionality of the original RSS measurements was reduced significantly via random linear projections on a suitable measurement basis, while maintaining an increased localization accuracy.

The empirical experimental evaluation revealed that the multivariate Gaussian method usually outperforms previous statistical signal-strength fingerprint approaches, while the CS-based approach achieved a superior performance when compared to the multivariate Gaussian-based technique. We performed an evaluation of various fingerprint methods in the premises of FORTH and an aquarium. The presence of people, as well as the density and placement of APs have a prominent impact on the positioning accuracy. Furthermore, in the case of the multivariate Gaussian-based algorithm we experimented with a spatial multiscale iterative approach in which we applied the algorithm on larger regions to select the correct one, and then within the selected region to estimate the correct cell. Something similar was performed in the case of percentiles by selecting the top 5 candidate cells. We showed that it improves the accuracy by eliminating the distant incorrect cells and taking also into consideration the effect of neighboring cells around the correct one. Other related works have also shown that the integration of user mobility models can further improve the accuracy. In the context of the aquarium, in which mobility patterns do exist, the integration of user mobility models could be helpful.

Finally, a hybrid path tracking system is presented, which exploits the efficiency of a Kalman filter in conjunction with the power of compressive sensing to represent accurately sparse signals and a region-based multivariate Gaussian model. The experimental evaluation reveals an increased localization performance, while maintaining a low computational complexity.

We have been also experimenting with other modalities, such as infrared,

cameras and QR codes to improve the location estimation. Specifically, in front of each landmark (*e.g.,* tank of the aquarium or office in TNL), a unique QR code can be placed along with three infrared sensors (*e.g.,* WII bar). The camera of the mobile device of a visitor may capture the QR code, recognize it, and thus identify the landmark, in front of which this visitor is standing. Similarly, when the camera captures the infrared light from at least two infrared sources, it can estimate its distance from the landmark by measuring the distance of the two infrared sources on the recorded image. We plan to extend our localization system by incorporating these multi-modalities measurements.

There is a growing interest in statistical methods that exploit various spatio-temporal statistical properties of the received signal to form robust fingerprints. In general, a channel exhibits very transient phenomena and is highly time-varying. At the same time, the collection of signal measurements is subject to inaccuracies due to various issues, such as hardware misconfigurations, limitations, time synchronization, fine-grain data sampling, incomplete information, and vendor-specific dependencies (often not publicly available). Thus, the general problem of building a theoretical framework to analyze these fingerprint techniques taking into consideration the above limitations opens up exciting research opportunities.

Regarding the part of CS, in the present work, the unknown location was estimated by performing separate reconstruction for each AP. A straightforward extension will be the use of the joint sparsity structure of the indicator vector $\mathbf{w}$ among the APs for the simultaneous location estimation. Moreover, the random nature of the measurement vectors associated with an RSS vector could be exploited in order to enhance the encryption capabilities of the proposed CS localization approach, without the additional computational cost of a separate encryption protocol. Besides, a more thorough study should be carried out for the robustness of the inherent encryption property in terms of the several network parameters. The choice of appropriate sparsifying and measurement bases is crucial for an increased localization accuracy. The design of new transform and measurement bases, $\mathbf{\Psi}$, $\mathbf{\Phi}$, respectively, being adaptive to the specific characteristics of the RSS data is also of significant importance.

Bibliography

[1] The bat ultrasonic location system.
http://www.cl.cam.ac.uk/research/dtg/attarchive/bat/. (Cited on page 17.)

[2] Ekahau v.3.1.
http://www.ekahau.com. (Cited on pages 1 and 3.)

[3] Future computing environments.
http://www.cc.gatech.edu/fce/smartfloor/. (Cited on page 17.)

[4] Precise Real-time Location.
http://www.ubisense.net/. (Cited on page 17.)

[5] Asthana, S., and Kalofonos, D. "The problem of bluetooth pollution and accelerating connectivity in bluetooth ad-hoc networks", in *IEEE PerCom* (New York, NY, USA, 2005). (Cited on page 1.)

[6] M. Azizyan, I. Constandache, and R. Choudhury, "SurroundSense: Mobile phone localization via ambience fingerprinting," in *Proc. ACM MOBICOM'09*, Sep. 2009. (Cited on pages 1 and 17.)

[7] P. Bahl and V. Radmanabhan, "An in-building RF-based user location and tracking system," in *Proc. IEEE INFOCOM 2000*, Mar. 2000. (Cited on pages 1, 8, 9 and 16.)

[8] U. Bandara *et al.*, "Design and implementation of a bluetooth signal strength based location sensing system, " in *Proc. IEEE RAWCON'04*, Atlanta, USA, Sep. 2004. (Cited on page 1.)

[9] Capkun, S., Hamdi, M., and Hubaux, J.-P. "GPS-Free Positioning in Mobile Ad-Hoc Networks", in *Proceedings of Hawaii International Conference On System Sciences* (Hawaii, Jan. 2001). (Cited on pages 2 and 17.)

[10] Chai, X., and Yang, Q. "Reducing the calibration effort for probabilistic indoor location estimation", *IEEE Transactions on Mobile Computing* (June 2007). (Cited on page 18.)

[11] Chintalapudi, K., Govindan, R., Sukhatme, G., and Dhariwal, A. "Ad-hoc localization using ranging and sectorin", in *IEEE InfoCom* (Hong Kong, March 2004). (Cited on page 2.)

[12] Fang, L., Du, W., and Ning, P. "A beacon-less location discovery scheme for wireless sensor networks", in *IEEE InfoCom* (Miami, Florida, March 2005), pp. 161-171. (Cited on page 2.)

[13] Feldmann, S., Kyamakya, K., Zapater, A., and Lue, Z. "An indoor Bluetooth-based positioning system: concept, implementation and experimental evaluation", in *International Conference on Wireless Networks* (Las Vegas, Nevada, USA, 2003), pp. 109-113. (Cited on page 1.)

[14] C. Fretzagias and M. Papadopouli, "Cooperative location sensing for wireless networks," in *Proc. IEEE PERCOM'04*, Orlando, Florida, Mar. 2004. (Cited on pages 2 and 17.)

[15] Gwon, Y., Jain, R., and Kawahara, T. "Robust indoor location estimation of stationary and mobile users", in *IEEE InfoCom* (Hong Kong, March 2004). (Cited on pages 1 and 2.)

[16] He, T., Huang, C., Blum, B. M., Stankovic, J. A., and Abdelzaher, T. "Range-free localization schemes for large-scale sensor networks", in *ACM MobiCom* (San Diego, CA, USA, Sep 2003). (Cited on page 17.)

[17] Hightower, J., and Borriello, G. "A Survey and Taxonomy of Location Sensing Systems for Ubiquitous Computing". Technical Report, University of Washington, Department of Computer Science and Engineering UW CSE 01-08-03, Seattle, WA, Aug. 2001. (Cited on page 2.)

[18] Ji, Y., Biaz, S., Pandey, S., and Agrawal, P. "Ariadne: A dynamic indoor signal map construction and localization system", in *ACM MobiSys* (June 19-22 2006). (Cited on page 18.)

[19] Kung, H. T., Lin, C.-K., Lin, T.-H., and Vlah, D. "Localization with snap-inducing shaped residuals (sisr): Coping with errors in measurement", in *ACM MobiCom* (September 20-25 2009). (Cited on page 16.)

[20] A. Ladd *et al.*, "Robotics-based location sensing using wireless ethernet, " in *Proc. ACM MOBICOM'02*, Atlanta, USA, Sep. 2002. (Cited on pages 1 and 16.)

[21] Niculescu, D., and Nath, B. , "Ad Hoc Positioning System (APS)", in *IEEE GlobeCom* (San Antonio, TX, Nov 2001). (Cited on pages 2 and 17.)

[22] Niculescu, D., and Nath, B. , "Ad Hoc Positioning System (APS) using AoA", in *IEEE InfoCom* (San Francisco, CA, Apr 2003). (Cited on page 17.)

[23] Paramvir Bahl, V. N. P., and Balachandran, A. "Enhancements to the radar user location and tracking system". Technical Report, Microsoft Research (February 2000). (Cited on page 16.)

[24] Patwari, N., and Kasera, S. K. "Robust location distinction using temporal link signatures", in *ACM MobiCom* (September 9-14 2007). (Cited on page 17.)

[25] N. Priyantha, A. Chakraborty, and H. Balakrishnan, "The cricket location-support system," in *Proc. ACM MOBICOM 2000*, Aug. 2000. (Cited on pages 1 and 2.)

[26] Priyantha, N. B., Miu, A. K. L., Balakrishnan, H., and Teller, S., "The Cricket compass for context-aware mobile applications", in *ACM MobiCom* (Rome, Italy, July 2001), pp. 1-14. (Cited on page 1.)

[27] Rodriguez, M., Pece, J. P., and Escudero, C. J. "In-building location using bluetooth", in *International Workshop on Wireless Ad-hoc Networks* (Coruna, Spain, May 2005). (Cited on page 1.)

[28] Roy, A., Misra, A., and Das, S. K. "An information theoretic framework for optimal location tracking in multi-system 4G wireless networks". In *IEEE InfoCom* (Hong Kong, March 2004). (Cited on page 1.)

[29] Savarese, C., Rabaey, J., and Langendoen, K. "Robust positioning algorithms for distributed ad-hoc wireless sensor networks", in *USENIX* (Monterey, CA, June 2002). (Cited on pages 2 and 17.)

[30] Tesoriero, M., Tebara, R., Gallud, J., Lozanoa, M., and Penichet, V. "Improving location awareness in indoor spaces using RFID technology", in *Expert System Application.* (Cited on page 18.)

[31] Vandikas, K., Kriara, L., Papakonstantinou, T., Katranidou, A., Baltzakis, H., and Papadopouli, M. "Empirical-based analysis of a cooperative location-sensing system", in *ACM First International Conference on Autonomic Computing and Communication Systems* (Autonomics) (Rome, Italy, Oct. 2007). (Cited on page 2.)

[32] Want, R., and Hopper, A. "Active badges and personal interactive computing objects". Technical Report ORL 92-2, Olivetti Research, also in *IEEE Transactions on Consumer Electronics* (February 1992). (Cited on page 17.)

[33] R. Want *et al.*, "The active badge location system", *ACM Trans. on Inf. Sys.*, Vol. 10, No. 1, pp. 91–102, Jan. 1992. (Cited on page 1.)

[34] Youssef, M., and Agrawala, A., "The horus WLAN location determination system", in *ACM MobiSys* (June 6-8 2005). (Cited on pages 1 and 16.)

[35] Youssef, M., Youssef, A., Rieger, C., Shankar, U., and Agrawala, A. "PinPoint: An asynchronous time-based location determination system", in *ACM MobiSys* (Uppsala, Sweden, June 2006), pp. 165-176. (Cited on pages 1 and 2.)

[36] Zhang, J., Firooz, M. H., Patwari, N., and Kasera, S. K. "Advancing wireless link signatures for location distinction", in *ACM MobiCom* (September 14-19 2008). (Cited on page 17.)

[37] K. Kaemarungsi and P. Krishnamurthy, "Properties of indoor received signal strength for wlan location fingerprinting", in *Mobile and Ubiquitous Systems: Networking and Services*, 2004. MOBIQUITOUS 2004. (Cited on pages 4, 9 and 15.)

[38] J. Hightower and G. Borriello, "Location systems for ubiquitous computing", Computer, vol. 34, no. 8, pp. 57-66, August 2001. (Cited on page 2.)

[39] A. Bensky, "Wireless Positioning Technologies and Applications". Artech House, Inc., 2008. (Cited on page 8.)

[40] R. Singh, L. Macchi, C. Regazzoni, and K. Plataniotis,"A statistical modelling based location determination method using fusion in WLAN", in *Proceedings of the International Workshop on Wireless Ad-hoc Networks*, 2005. (Cited on page 8.)

[41] N. K. Sharma, "A weighted center of mass based trilateration approach for locating wireless devices in indoor environment", in *Proceedings of the 4th ACM international workshop on Mobility management and wireless access*, 2006, pp. 112-115. (Cited on page 8.)

[42] A. Kushki, K. N. Plataniotis, and A. N. Venetsanopoulos,"Kernel-Based Positioning in Wireless Local Area Networks", *IEEE Transactions on Mobile Computing*, vol. 6, no. 6, pp. 689-705, June 2007. (Cited on page 8.)

[43] K. Kaemarungsi and P. Krishnamurthy, "Modeling of indoor positioning systems based on location fingerprinting", in *InfoCom 2004*. Twenty-third Annual-Joint Conference of the IEEE Computer and Communications Societies, vol. 2, March 2004, pp. 1012-1022 vol.2. (Cited on pages 8 and 10.)

[44] B. Li, Y. Wang, H. K. Lee, A. Dempster, and C. Rizos, "Method for yielding a database of location fingerprints in WLAN", in *IEEE Proceedings Communications*, vol. 152, no. 5, pp. 580-586, October 2005. (Cited on page 10.)

[45] J. Yin, Q. Yang, and L. M. Ni, "Learning Adaptive Temporal Radio Maps for Signal-Strength-Based Location Estimation", in *IEEE Transactions on Mobile Computing*, vol. 7, no. 7, pp. 869-883, July 2008. (Cited on page 10.)

[46] K. Kaemarungsi, "Distribution of WLAN received signal strength indication for indoor location determination", in *1st International Symposium on Wireless Pervasive Computing*, January 2006, p. 6 pp. (Cited on page 10.)

[47] A. K. M. Mahtab Hossain, H. N. Van, Y. Jin, and W.S. Soh, "Indoor Localization Using Multiple Wireless Technologies", in *IEEE Internatonal Conference on Mobile Adhoc and Sensor Systems*, October 2007, pp. 1-8. (Cited on page 11.)

[48] V. Honkavirta, T. Perala, S. Ali-Loytty, and R. Piche, "A comparative survey of WLAN location fingerprinting methods", in *6th Workshop on Positioning, Navigation and Communication*, March 2009, pp. 243-251. (Cited on page 11.)

[49] I. Guvenc, C. T. Abdallah, R. Jordan, and O. Dedeoglu, "Enhancements to RSS Based Indoor Tracking Systems Using Kalman Filters", in *Global Signal Processing Expo and International Signal Processing Conference*, 2003. (Cited on page 15.)

[50] J. A. Besada, A. M. Bernardos, P. Tarrio, and J. R. Casar, "Analysis of tracking methods for wireless indoor localization", in *2nd International Symposium on Wireless Pervasive Computing*, February 2007. (Cited on page 15.)

[51] A. Kushki, K. N. Plataniotis, and A. N. Venetsanopoulos, "Location Tracking in Wireless Local Area Networks with Adaptive Radio MAPS", in *IEEE International Conference in Acoustics, Speech and Signal Processing*, ICASSP '06, May 2006. (Cited on page 15.)

[52] F. Evennou and F. Marx, "Improving positioning capabilities for indoor environments with Wifi", in *IST Summit*, 2005. (Cited on page 15.)

[53] J. Yim, S. Jeong, J. Joo, and C. Park, "Utilizing Map Information for WLANBased Kalman Filter Indoor Tracking", in *2nd International Conference on Future Generation Communication and Networking Symposia*, vol. 5, December 2008, pp. 58-62. (Cited on page 15.)

[54] Y. Song and H. Yu, "A RSS Based Indoor Tracking Algorithm via Particle Filter and Probability Distribution", in *4th International Conference on Wireless Communications, Networking and Mobile Computing*, October 2008, pp. 1-4. (Cited on page 15.)

[55] M. S. Arlampalam, S. Maskell, N. Gordon, and T. Clapp, "A Tutorial on Particle Filters for Online Nonlinear/NonGaussian Bayesain Tracking", *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174-188, February 2002. (Cited on page 15.)

[56] H. Wang, H. Lenz, A. Szabo, J. Bamberger, and U. D. Hanebeck, "Enhancing the map usage for indoor location-aware systems", in *Proceedings of the 12th international conference on Human-computer interaction*, Berlin, Heidelberg: Springer-Verlag, 2007, pp. 151-160. (Cited on page 16.)

[57] C.-H. Chao, C.-Y. Chu, and A.-Y.Wu, "Location-Constrained Particle Filter human positioning and tracking system", in *IEEE Workshop on Signal Processing Systems*, SiPS 2008, October 2008, pp. 73-76. (Cited on page 16.)

[58] Widyawan, M. Klepal, and S. Beauregard, "A novel backtracking particle filter for pattern matching indoor localization", in *Proceedings of the first ACM international workshop on Mobile entity localization and tracking in GPS-less environments*, New York, NY, USA: ACM, 2008, pp. 79-84. (Cited on page 16.)

[59] H.Wang, H. Lenz, A. Szabo, J. Bamberger, and U. Hanebeck, "WLAN-Based Pedestrian Tracking Using Particle Filters and Low-Cost MEMS Sensors", in *4th Workshop on Positioning, Navigation and Communication*, March 2007, pp. 1-7. (Cited on page 16.)

[60] O. Woodman and R. Harle, "Pedestrian localisation for indoor environments", in *Proceedings of the 10th international conference on Ubiquitous computing*, New York, NY, USA: ACM, 2008, pp. 114-123. (Cited on page 16.)

[61] I.E. Liao and K.F. Kao, "Enhancing the accuracy of WLAN-based location determination systems using predicted orientation information", in *Inf. Sci.*, vol. 178, no. 4, pp. 1049-1068, 2008. (Cited on page 19.)

[62] R. Zhou, "Wireless Indoor Tracking System (WITS)", in *doIT Conference on Software Research*, 2006, pp. 163-177. (Cited on page 19.)

[63] A. Kushki, K. N. Plataniotis, and A. N. Venetsanopoulos, "Intelligent Dynamic Radio Tracking in Indoor Wireless Local Area Networks", *IEEE Transactions on Mobile Computing*, vol. 9, no. 3, pp. 405-419, March 2010. (Cited on page 19.)

[64] Rappaport, T. S., "Wireless Communications, principle and practices", 2002, Prentice-Hall (Cited on page 8.)

[65] Cox, D. C., Murray, R., and Norris, A., "800 Mhz Attenuation measured in and around Suburban Houses", AT&T Bell Laboratory Technical Journal, Vol. 673, No. 6, July - August 1984. (Cited on page 8.)

[66] Bernhardt, R. C., "Macroscopic diversity in frequency reuse systems", in *IEEE Journal on Selected Areas in Communications*, Vol-SAC 5, pp. 862-878, June 1987 (Cited on page 8.)

[67] R. Baraniuk, "Compressive sensing", *IEEE Signal Processing Magazine*, vol. 24, pp. 118–121, July 2007. (Cited on page 14.)

[68] E. Candés, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information", *IEEE Trans. on Inf. Th.*, Vol. 52, pp. 489–509, Feb. 2006. (Cited on pages 2, 15, 27 and 29.)

[69] D. Donoho, "Compressive sensing", *IEEE Trans. on Inf. Th.*, Vol. 52, No. 4, pp. 1289–1306, April 2006. (Cited on pages 2, 15, 27, 28 and 29.)

[70] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit", *SIAM J. on Sci. Comp.*, Vol. 20, No. 1, pp. 33–61, 1999. (Cited on page 29.)

[71] C. Feng, S. Valaee, and Z. Tan, "Multiple target localization using compressive sensing," in *Proc. IEEE GLOBECOM'09*, Hawaii, USA, 4 Nov.–5 Dec. 2009 . (Cited on page 2.)

[72] C. Feng, W. Au, S. Valaee, and Z. Tan,"Compressive sensing based positioning using RSS of WLAN access points," in *Proc. IEEE INFOCOM'10*, San Diego, CA, Mar. 2010. (Cited on page 31.)

[73] D. Milioris, L. Kriara, A. Papakonstantinou, G. Tzagkarakis, P. Tsakalides, and M. Papadopouli, "Empirical evaluation of signal-strength fingerprint positioning in wireless LANs," in *Proc. 13th ACM MSWIM'10*, Bodrum, Turkey, Oct. 2010. (Cited on pages 18 and 19.)

[74] B. Li, J. Salter, A. Dempster, and C. Rizos, "Indoor positioning techniques based on wireless LAN," in *Proc. AUSWIRELESS'06*, Sydney, Mar. 2006. (Cited on page 11.)

[75] S. Nikitaki and P. Tsakalides, "Localization in wireless networks via spatial sparsity," in *Proc. 44th ASILOMAR'10*, Pacific Grove, CA, Nov. 2010. (Cited on page 14.)

[76] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems", *IEEE Trans. on Sys., Man, and Cyber.*, Part C, Vol. 37, No. 6, pp. 1067–1080, Nov. 2007. (Cited on pages 2 and 8.)

[77] R. Tibshirani,"Regression shrinkage and selection via the lasso," *J. of the Royal Stat. Soc. Series B (Methodological)*, Vol. 58, No. 1, pp. 267–288, 1996. (Cited on page 29.)

[78] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. on Inf. Th.*, Vol. 53, pp. 4655–466, Dec. 2007. (Cited on page 29.)

[79] D. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit," *Tech. Report*, Stanford 2006 [on-line: `http://www-stat.stanford.edu/~donoho/Reports/2006/StOMP-20060403.pdf`] (Cited on page 29.)

[80] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Sig. Proc.*, Vol. 56, No. 6, pp. 2346–2356, June 2008. (Cited on page 51.)

[81] G. Tzagkarakis and P. Tsakalides, "Bayesian compressed sensing imaging using a Gaussian scale mixture," in *Proc. IEEE ICASSP'10*, Dallas, TX, Mar. 2010. (Cited on page 51.)

[82] A. Orsdemir, H. Altun, G. Sharma, and M. Bocko, "On the security and robustness of encryption via compressed sensing," in *Proc. MILCOM'08*, San Diego, CA, Nov. 2008. (Cited on page 32.)

[83] `http://www.wirelesscommunication.nl/reference/chaptr03/indoor.htm` (Cited on page 54.)